Perturbation theory for spectral subspaces

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften
am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

vorgelegt von

Albrecht Seelmann

geboren in Worms

Mainz, den 26. Mai 2014

1. Berichterstatter:
2. Berichterstatter:
3. Berichterstatter:

Datum der mündlichen Prüfung: 24.9.2014

Acknowledgements

For reasons of privacy protection, it is not allowed to state the names of others here. The acknowledgements are therefore left blank in this electronic version of the thesis.

Abstract

In the present thesis, the variation of closed subspaces of a Hilbert space associated with isolated components of the spectra of linear self-adjoint operators under a bounded additive perturbation is studied. Of particular interest is the least restrictive condition on the norm of the perturbation that guarantees that the difference of the corresponding orthogonal projections is a strict norm contraction. An overview on the results obtained so far is given.

Based on an iteration approach, a general bound on the variation of the subspaces is obtained for perturbations depending smoothly on a real parameter. The result is applied to the case of additive perturbations by introducing a coupling parameter on the perturbation. In this way, previously known results are strengthened.

In the case of additive perturbations, the bounds on the variation of the subspaces are sharpened further by an optimization procedure for the choice of the supporting points in the iteration approach. The corresponding results are the best ones obtained so far.

Zusammenfassung

In der vorliegenden Arbeit wird die Variation abgeschlossener Unterräume eines Hilbertraumes untersucht, die mit isolierten Komponenten der Spektren von selbstadjungierten Operatoren unter beschränkten additiven Störungen assoziiert sind. Von besonderem Interesse ist hierbei die am wenigsten restriktive Bedingung an die Norm der Störung, die sicherstellt, dass die Differenz der zugehörigen orthogonalen Projektionen eine strikte Normkontraktion darstellt. Es wird ein Überblick über die bisher erzielten Resultate gegeben.

Basierend auf einem Iterationsansatz wird eine allgemeine Schranke an die Variation der Unterräume für Störungen erzielt, die glatt von einem reellen Parameter abhängen. Durch Einführung eines Kopplungsparameters wird das Ergebnis auf den Fall additiver Störungen angewendet. Auf diese Weise werden zuvor bekannte Ergebnisse verbessert.

Im Falle von additiven Störungen werden die Schranken an die Variation der Unterräume durch ein Optimierungsverfahren für die Stützstellen im Interationsansatz weiter verschärft. Die zugehörigen Ergebnisse sind die besten, die bis zum jetzigen Zeitpunkt erzielt wurden.

Contents

Introduction

One of the fundamental problems in operator perturbation theory is the *subspace perturbation problem*, in which the variation of invariant subspaces for a self-adjoint or normal operator under a bounded additive perturbation is studied, see, e.g., [11, 12, 21] and the references therein.

The simplest particular case in this context is the study of one-dimensional eigenspaces: Let A be a self-adjoint operator on a Hilbert space with an isolated simple eigenvalue λ. It is well known that if V is a bounded self-adjoint operator with sufficiently small operator norm $\|V\|$, then the perturbed operator $A+V$ also has an isolated simple eigenvalue μ in a small neighbourhood of λ, and it is natural to ask how the respective eigenspaces for A and $A+V$ differ. Since eigenvectors are determined only up to phase factors, it is more suitable to study the variation of the eigenspaces in terms of the corresponding eigenprojections P and Q for A and $A+V$, respectively, rather than in terms of the eigenvectors, cf. [20].

It is well known that the operator norm of the difference $P-Q$ cannot exceed 1, and it turns out that it equals 1 if and only if the corresponding eigenvectors for A and $A+V$ are orthogonal to each other. However, if these eigenvectors are not orthogonal to each other, then Q does not vanish on the eigenspace for A. In this case, the norm of the difference $P-Q$ can be expressed as (see [21])

$$\|P-Q\| = \|x - Qx\| = \sin\theta < 1\,, \tag{1}$$

where x is a normalized eigenvector for A associated with λ and θ is the angle between x and Qx, that is,

$$\cos\theta = \left\langle x, \frac{Qx}{\|Qx\|} \right\rangle = \|Qx\| > 0\,.$$

In this regard, an eigenvector for $A + V$ can be obtained by rotating the eigenvector x through the angle $\theta < \pi/2$.

Another advantage of the use of projections rather than vectors is that the relation (1) can be studied in a much more general setting, such as eigenvalues of higher multiplicity or clusters of different eigenvalues, whereas the consideration of vectors here would have further complications if the eigenvalues are closely bunched, cf. [21]. Also more general invariant subspaces can be considered in terms of orthogonal projections. In these more general situations, instead of the angle θ in (1), an operator-valued analogue enters the considerations, the so-called *operator angle* Θ. This is a self-adjoint operator which is associated with the corresponding subspaces for the unperturbed and perturbed operators, respectively, and whose spectrum lies in the interval $\bigl[0, \frac{\pi}{2}\bigr]$. Suitable norms of it, or of trigonometric functions thereof, serve as a measure for the difference between the subspaces, and the main objective is to obtain efficient estimates on these norms in terms of the strength of the perturbation.

Usually, estimates of the mentioned sort require that associated parts of the spectra of the corresponding operators are separated from each other, and distances between these spectral parts typically enter the estimates. The four angle theorems by Davis and Kahan [21], namely $\sin\Theta$, $\sin 2\Theta$, $\tan\Theta$, and $\tan 2\Theta$, represent the pioneering work in this direction. Each of these four theorems is suited for a different situation with certain assumptions on the spectra and/or on the perturbation. Extensions and generalizations of the Davis-Kahan angle theorems have been considered in several recent works such as [7, 8, 28, 30, 31, 40].

In this thesis, besides providing a generalization of the Davis-Kahan $\sin 2\Theta$ theorem, we focus on the following more specific problem:

Let A be a possibly unbounded self-adjoint operator on a Hilbert space $\mathcal{H}$ such that the spectrum of A contains an isolated component σ, that is,

$$d := \operatorname{dist}\bigl(\sigma, \operatorname{spec}(A) \setminus \sigma\bigr) > 0\,;$$

one may think of σ as a cluster of isolated eigenvalues such as in the case of matrices or the quantum harmonic oscillator (see, e.g., [8, Section 6]), or as a cluster of bands in the spectrum such as in the case of Schrödinger operators with periodic potentials, see, e.g., [44, Section XIII.16]. Let V be a bounded self-adjoint operator on $\mathcal{H}$. We then ask for the least restrictive condition on the norm of V, independent of A and V, which guarantees that

$$\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)\| < 1\,. \tag{2}$$

Here, E_A and E_{A+V} denote the spectral measures for the self-adjoint operators A and $A+V$, respectively, and $\mathcal{O}_{d/2}(\sigma)$ stands for the open $d/2$-neighbourhood of σ. This problem has initially been discussed by Kostrykin, Makarov, and Motovilov in [26], but earlier works by Langer and Tretter [32], Adamjan, Langer, and Tretter [4], and Albeverio, Makarov, and Motovilov [5] are closely related. In the framework of the present thesis, we refer to the problem of establishing (2) also as the *subspace perturbation problem*.

It is well known that the norm of the difference $\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ agrees with the norm of the operator $\sin\Theta$, where Θ is the operator angle associated with the subspaces $\operatorname{Ran}\mathsf{E}_A(\sigma)$ and $\operatorname{Ran}\mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$, see [21]. In this sense, inequality (2) is a more or less direct extension of (1). Here, the strict inequality in (2) ensures that the spectral projections $\mathsf{E}_A(\sigma)$ and $\mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ are unitarily equivalent, see [25, Theorem I.6.32]. The spectral subspace $\operatorname{Ran}\mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ for the perturbed operator $A+V$ can then be understood as a rotation of the unperturbed subspace $\operatorname{Ran}\mathsf{E}_A(\sigma)$, and the associated operator angle Θ plays the role of a rotation angle. The norm of the difference of the projections $\mathsf{E}_A(\sigma)$ and $\mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ serves as a measure for this rotation, so that one is interested not only in establishing the inequality (2) but also in obtaining sharp estimates on the left-hand side of (2). Equivalently, one searches for estimates on the norm $\|\Theta\| < \pi/2$.

Clearly, the condition (2) implies that the operator $A+V$ has spectrum in the neighbourhood $\mathcal{O}_{d/2}(\sigma)$ of σ. Since (2) is supposed to hold for all choices of A and V simultaneously, this, in turn, requires that $\|V\| < d/2$, cf. [25, Theorem V.4.10]; in this case, the intersection $\operatorname{spec}(A+V)\cap\mathcal{O}_{d/2}(\sigma)$ even is an isolated component of $\operatorname{spec}(A+V)$. The main question that arises now is whether the bound $\|V\| < d/2$ is sufficient for inequality (2) to hold or if one has to impose a stronger condition on $\|V\|$ in order to ensure (2). Under certain additional assumptions on the spectrum of A such as that the convex hull of σ is disjoint from the remainder of the spectrum, that is, $\operatorname{conv}(\sigma)\cap\bigl(\operatorname{spec}(A)\setminus\sigma\bigr) = \varnothing$, the answer to this question is known to be positive; this is a consequence of the Davis-Kahan $\sin 2\Theta$ theorem in

[21]. It has been conjectured that the answer is positive also if no additional assumptions on the spectrum of A are imposed (see [8]; cf. also [26] and [31]), but no proof for this is available so far.

The principal result in this thesis is that (2) holds whenever

$$(3) \qquad \|V\| < c_{\mathrm{crit}} \cdot d \quad \text{with} \quad c_{\mathrm{crit}} = \frac{1}{2} - \frac{1}{2}\Bigl(1 - \frac{\sqrt{3}}{\pi}\Bigr)^3 = 0.4548\ldots\,,$$

see Theorem 8.9 below. Together with a corresponding estimate on the norm of the operator angle, this result is the best one obtained so far.

The problem of establishing (2) has also been discussed under the additional assumption that the perturbation V is off-diagonal with respect to the decomposition of the Hilbert space $\mathcal{H}$ induced by the orthogonal projection $\mathsf{E}_A(\sigma)$, see [31] and also [5, Remark 3.11 and Theorem 7.6]. This particular structure of the perturbation allows to obtain results substantially stronger than (3). The present thesis also contains contributions to this case, see Theorem 6.15 (b) and Section 8.3 below, and the corresponding results are the best ones obtained so far for this situation.

Another class of perturbations that lead to results stronger than (3) are semidefinite ones, that is, perturbations V with $V \geq 0$ or $V \leq 0$. Although such kind of perturbations are rather prominent in general perturbation theory, it seems that they have not explicitly been studied in the context of inequality (2) before. In the present thesis, this situation is discussed briefly in the form of an outlook for future research, see Section 2.4 below.

The key idea in the approach of the present thesis to the problem of establishing (2) is to iterate the bound on the rotation of the corresponding subspaces. To this end, a coupling parameter on the perturbation is introduced, namely

$$B_t := A + tV\,, \quad \mathrm{Dom}(B_t) := \mathrm{Dom}(A)\,, \quad t \in [0,1]\,,$$

and this parameter is increased in small steps according to a suitably chosen partition of the interval $[0,1]$. Of particular importance here is that the norm of the associated operator angle satisfies a triangle inequality with respect to the subspaces (see [16, Corollary 4]), and this triangle inequality is stronger than the one for the usual operator norm for the difference of the projections.

The approach of iterating the rotation bound leads to the study of smooth variations of spectral subspaces, where partitions of the interval $[0,1]$ with arbitrarily small mesh size are considered. The main result in this context is the estimate

$$\|\Theta\| = \arcsin\bigl(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)\|\bigr) \le \frac{\pi}{2}\int_0^1 \frac{\|\dot{B}_\tau\|}{\operatorname{dist}(\omega_\tau, \Omega_\tau)}\,\mathrm{d}\tau\,,$$

where $\dot{B}_\tau = \frac{\mathrm{d}}{\mathrm{d}\tau}B_\tau$ and ω_τ and Ω_τ are suitably chosen spectral components of the perturbed operator B_τ. The corresponding considerations in Chapter 6 below deal with the more general situation of smooth paths of arbitrary self-adjoint operators B_t with appropriately separated spectra. This represents one of core parts of the present thesis.

However, for the particular problem of establishing inequality (2), it turns out that partitions with small mesh size do not give the best results. Albeverio and Motovilov observed in [8] that in the case of general perturbations one can obtain a stronger result with a particular finite partition. This requires an estimate on the norm of the associated operator angle that is more accurate for perturbations with small norm than the previously known bounds. Albeverio and Motovilov provided such a bound in form of the *generic* $\sin 2\theta$ *estimate*, which resembles the bound from the Davis-Kahan $\sin 2\Theta$ theorem in [21]. The present author has noted that there is a better choice for the finite partition and has formulated an optimization problem to obtain the best possible choice. This optimization problem is solved explicitly in Chapter 8 below, and the solution yields the result (3). Similar considerations for off-diagonal perturbations lead to an optimization problem that is more difficult to deal with and that is not solved explicitly yet. Nevertheless, numerical evaluations yield a result stronger than the previously known ones, see Corollary 8.26 below.

The thesis is organized as follows:

In Chapter 1, we fix the standard notations used throughout this thesis. We also recall and discuss some basic notions such as the operator angle, graph subpaces, and reducing subspaces.

Chapter 2 provides an overview on the subspace perturbation problem for self-adjoint operators. Here, we discuss which particular cases are already solved and what kind of results have been achieved for the general problem so far. An outlook on the case of semidefinite perturbations for future research is also provided here.

Chapter 3 is devoted to so-called operator Sylvester equations of the form $XA_0 - A_1X = K$, which are a main tool in this thesis. Here, it is explained how Sylvester equations are related to the subspace perturbation problem, the central existence and uniqueness result is recalled, and some consequences of this result are discussed, including a variant of the Davis-Kahan symmetric $\sin\Theta$ theorem.

In Chapter 4, we revisit the block diagonalization of self-adjoint 2×2 block operator matrices with respect to reducing graph subspaces. This has previously been discussed in [5, Section 5], and the material here fills in a gap in reasoning in the proof of [5, Lemma 5.3]. This chapter is based on the joint work [38] with K. A. Makarov and S. Schmitz.

Chapter 5 provides an alternative proof for the fact that the norm of the operator angle defines a metric on the set of orthogonal projections, which is essential for the considerations in the following Chapters 6 and 8. This alternative proof is based on parts of the joint work [36] with K. A. Makarov.

Chapter 6 forms the main part of this work. Here, we discuss smooth variations of spectral subspaces for self-adjoint operators with separated spectra. The corresponding result is applied to the problem of establishing inequality (2). This chapter is based on the joint work [37] with K. A. Makarov published in *Journal für die reine und angewandte Mathematik* and also extends the considerations there to unbounded operators.

In Chapter 7, an analogue of the Davis-Kahan $\sin 2\Theta$ theorem under a general spectral separation condition is established. This extends the generic $\sin 2\theta$ estimate recently shown by Albeverio and Motovilov in [8]. The corresponding material is taken with only small changes from the author's article [50] published in *Integral Equations and Operator Theory.*

Based on the $\sin 2\theta$ estimate, in Chapter 8 we formulate an optimization problem, whose solution yields the result (3). The corresponding material is taken from the author's preprint [51]. An analogous optimization problem for off-diagonal perturbations is also discussed here.

Finally, Appendix A is devoted to some elementary inequalities used in Chapter 8. Except for minor changes, it agrees with the appendix in the author's preprint [51].

Chapter 1

Preliminaries

In this first chapter, we introduce the basic notations used in the present thesis and recall some fundamental notions and concepts.

1.1 Basic notations and general assumptions

Notations. Throughout this thesis, $\mathbb{N}$ denotes the set of positive integers and $\mathbb{N}_0$ the one of non-negative integers. Moreover, $\mathbb{R}$ and $\mathbb{C}$ stand for the sets of real and complex numbers, respectively. The Euler number is denoted by e, and i stands for the complex unit.

Given a subset $\Delta \subset \mathbb{R}$, the open r-neighbourhood of Δ with $r \geq 0$ is denoted by $\mathcal{O}_r(\Delta)$, that is, $\mathcal{O}_r(\Delta) := \{\lambda \mid \operatorname{dist}(\lambda, \Delta) < r\}$. We write $\operatorname{dist}(\Lambda, \Delta)$ for the distance between two subsets Λ and Δ of $\mathbb{R}$, which is understood as the infimum of the distances between points from the respective sets.

Given a Hilbert space $\mathcal{H}$, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\| \cdot \|_{\mathcal{H}}$ stand for the corresponding inner product and norm, respectively, where the subscript $\mathcal{H}$ is usually omitted. The space of bounded linear operators from a Hilbert space $\mathcal{H}$ to a Hilbert space $\mathcal{K}$ is denoted by $\mathcal{L}(\mathcal{H}, \mathcal{K})$, and $\| \cdot \|$ stands for the usual operator norm on $\mathcal{L}(\mathcal{H}, \mathcal{K})$. If $\mathcal{H} = \mathcal{K}$, we simply write $\mathcal{L}(\mathcal{H}) := \mathcal{L}(\mathcal{H}, \mathcal{H})$. The identity operator on $\mathcal{H}$ is denoted by $I_{\mathcal{H}}$. Multiples $\lambda I_{\mathcal{H}}$ of this operator are usually abbreviated by λ.

Unless stated otherwise, every operator in this thesis is allowed to be unbounded. The domain of a linear operator A is denoted by $\operatorname{Dom}(A)$, and its range by $\operatorname{Ran}(A)$. The restriction of A to a given subspace $\mathcal{U}$ is written as $A|_{\mathcal{U}}$ with $\operatorname{Dom}(A|_{\mathcal{U}}) := \operatorname{Dom}(A) \cap \mathcal{U}$. Given another linear operator B, we write the extension relation $A \subset B$ (or $B \supset A$) if B extends A, that is, if one has $\operatorname{Dom}(A) \subset \operatorname{Dom}(B)$ and $Ax = Bx$ for $x \in \operatorname{Dom}(A)$. The operator equality $A = B$ means that $A \subset B$ and $A \supset B$. Note that sums and products of operators are always understood on their natural domains.

If A is a closed densely defined operator on a Hilbert space, its adjoint operator is denoted by A^*, its spectrum by $\operatorname{spec}(A)$, and its resolvent set by $\rho(A)$. If A is self-adjoint, then E_A stands for its spectral measure.

For a self-adjoint operator A and $\lambda \in \mathbb{R}$ we write $A \geq \lambda$ (or $\lambda \leq A$) if $\langle x, Ax\rangle \geq \lambda \|x\|^2$ for all $x \in \operatorname{Dom}(A)$. For simplicity, we write $A \leq \lambda$ (or $\lambda \geq A$) instead of $-A \geq -\lambda$.

Finally, if P is an orthogonal projection in the Hilbert space $\mathcal{H}$, that is, $P \in \mathcal{L}(\mathcal{H})$ with $P^2 = P = P^*$, then we write $P^\perp := I_\mathcal{H} - P$ for the orthogonal projection onto the orthogonal complement $(\operatorname{Ran} P)^\perp$ of $\operatorname{Ran} P$. The orthogonal projection onto a given closed subspace $\mathcal{U} \subset \mathcal{H}$ is denoted by $P_\mathcal{U}$.

General assumptions. For convenience, every Hilbert space in this thesis is tacitly assumed to be complex. However, except for Theorem 3.2 and Corollary 3.5 below, the statements of all results presented here make perfect sense also if the underlying Hilbert space is real, and it is straightforward to extend them to this case, either directly or by complexification (see, e.g., [57, Abschnitt 4.4] and [56, Exercises 5.32 and 7.25]).

Every Hilbert space may also be assumed to be separable. This is done in many of the cited works. However, the results obtained in the present thesis do not need this assumption, so that we do not impose it explicitly here.

1.2 Invariant and reducing subspaces

For the concepts of invariant and reducing subspaces for a linear operator, we mainly rely on [49, Section 1.4], [56, Exercise 5.39 and Theorem 7.28], and [57, Satz 2.60].

Let A be a linear operator on the Hilbert space $\mathcal{H}$. A closed subspace $\mathcal{U} \subset \mathcal{H}$ is called *invariant* for A if A maps the intersection $\operatorname{Dom}(A) \cap \mathcal{U}$ into $\mathcal{U}$. The subspace $\mathcal{U}$ is called *reducing* for A if both $\mathcal{U}$ and its orthogonal complement $\mathcal{U}^\perp$ are invariant for A and the domain $\operatorname{Dom}(A)$ splits as

$$\operatorname{Dom}(A) = \bigl(\operatorname{Dom}(A) \cap \mathcal{U}\bigr) + \bigl(\operatorname{Dom}(A) \cap \mathcal{U}^\perp\bigr)\,. \tag{1.1}$$

Clearly, the subspace $\mathcal{U}$ is reducing for A if and only if $\mathcal{U}^\perp$ is. In this case, the operator A can be represented as the direct sum $A = A_0 \oplus A_1$ with respect to the orthogonal decomposition $\mathcal{H} = \mathcal{U} \oplus \mathcal{U}^\perp$, where A_0 and A_1 are the restrictions of A to $\mathcal{U}$ and $\mathcal{U}^\perp$, respectively, that is, $A_0 = A|_{\mathcal{U}}$ and $A_1 = A|_{\mathcal{U}^\perp}$. In particular, one has $\operatorname{Dom}(A) = \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)$. Equivalently, A can be written as the diagonal 2×2 block operator matrix

$$A = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}$$

with respect to $\mathcal{H} = \mathcal{U} \oplus \mathcal{U}^\perp$. The operators A_0 and A_1 are called the *parts of* A associated with $\mathcal{U}$ and $\mathcal{U}^\perp$, respectively.

If, in addition, A is a closed operator, then the parts A_0 and A_1 of A are closed as well and the spectrum of A decomposes as

$$\operatorname{spec}(A) = \operatorname{spec}(A_0) \cup \operatorname{spec}(A_1)\,,$$

see [57, Satz 5.11].

If A is a bounded self-adjoint operator, then every invariant subspace for A is automatically reducing. If A is unbounded, then this is in general not the case, see [49, Example 1.8] for a counterexample. In this respect, the splitting property (1.1) is not self-evident in the case of unbounded operators.

The property of a closed subspace to be reducing for a linear operator A can also be characterized in terms of the corresponding orthogonal projection. Namely, a closed subspace $\mathcal{U} \subset \mathcal{H}$ is reducing for A if and only if the orthogonal projection $P = P_{\mathcal{U}}$ onto $\mathcal{U}$ commutes with A, that is, if

$$PA \subset AP\,. \tag{1.2}$$

This means that one has $Px \in \operatorname{Dom}(A)$ and $PAx = APx$ for $x \in \operatorname{Dom}(A)$. In this regard, important examples of reducing subspaces for a self-adjoint operator A are provided in terms of its spectral measure E_A.

Example 1.1 (cf. [56, Theorem 7.28]). Let A be a self-adjoint operator. Then, for every Borel set $\Delta \subset \mathbb{R}$ the subspace $\operatorname{Ran} \mathsf{E}_A(\Delta)$ is reducing for A, and the part A_0 of A associated with $\operatorname{Ran} \mathsf{E}_A(\Delta)$ is self-adjoint with spectrum

$$\operatorname{spec}(A_0) = \overline{\operatorname{spec}(A) \cap \Delta}\,.$$

In view of the preceding example, the orthogonal projection $\mathsf{E}_A(\Delta)$ with $\Delta \subset \mathbb{R}$ a Borel set is called a *spectral projection* for A, and $\operatorname{Ran} \mathsf{E}_A(\Delta)$ is called a *spectral subspace* for A.

The characterization (1.2) of reducing subspaces combined with the functional calculus for self-adjoint operators also yields the following well-known result.

Lemma 1.2 (see [57, Satz 8.23])**.** *Let A be a self-adjoint operator, and let P be an orthogonal projection onto a reducing subspace for A. Then, for every Borel-measurable function $g\colon \mathbb{R} \to \mathbb{C}$, the subspace* $\operatorname{Ran} P$ *is reducing for the operator $g(A)$.*

Remark 1.3. In the situation of Lemma 1.2, it is easy to verify that if A_0 is the part of A associated with $\operatorname{Ran} P$, then $g(A_0)$ is the part of $g(A)$ associated with $\operatorname{Ran} P$.

1.3 Graph subspaces

A closed subspace $\mathcal{G}$ of the Hilbert space $\mathcal{H}$ is said to be a *graph subspace* associated with a closed subspace $\mathcal{N} \subset \mathcal{H}$ and a bounded operator X from $\mathcal{N}$ to its orthogonal complement $\mathcal{N}^\perp$ if

$$\mathcal{G} = \mathcal{G}(\mathcal{N}, X) := \{x \in \mathcal{H} \mid P_{\mathcal{N}^\perp} x = X P_{\mathcal{N}} x\}\,.$$

Here, X is identified with its trivial continuation to the whole Hilbert space $\mathcal{H}$. An equivalent representation for the graph subspace $\mathcal{G}(\mathcal{N}, X)$ is given by

$$\mathcal{G}(\mathcal{N}, X) = \{g \oplus Xg \mid g \in \mathcal{N}\}\,.$$

The operator X is called the associated *angular operator*.

In the context of the present thesis, we are interested only in graph subspaces that are associated with bounded operators X. A discussion of a more general concept of graph subspaces where the angular operator is allowed to be unbounded or even non-closable, especially in the context of operator Riccati equations (see Section 1.4 and Chapter 4 below), can be found in [27] and [29].

One can easily check that

$$\mathcal{G}(\mathcal{N},X)^{\perp} = \mathcal{G}(\mathcal{N}^{\perp},-X^{*})\,.$$

Moreover, the orthogonal graph subspaces $\mathcal{G}(\mathcal{N},X)$ and $\mathcal{G}(\mathcal{N}^{\perp},-X^{*})$ can be represented as

$$\mathcal{G}(\mathcal{N},X) = \operatorname{Ran}(T|_{\mathcal{N}}) \quad \text{and} \quad \mathcal{G}(\mathcal{N}^{\perp},-X^{*}) = \operatorname{Ran}(T|_{\mathcal{N}^{\perp}})\,, \tag{1.3}$$

where the operator $T \in \mathcal{L}(\mathcal{H})$ is given by the 2×2 block operator matrix

$$T = \begin{pmatrix} I_{\mathcal{N}} & -X^{*} \\ X & I_{\mathcal{N}^{\perp}} \end{pmatrix}$$

with respect to the decomposition $\mathcal{H} = \mathcal{N} \oplus \mathcal{N}^{\perp}$. In particular, one has

$$QT = TP\,,$$

where $P := P_{\mathcal{N}}$ and Q denotes the orthogonal projection onto $\mathcal{G}(\mathcal{N},X)$.

The operator T is normal, more precisely

$$T^{*}T = TT^{*} = \begin{pmatrix} I_{\mathcal{N}} + X^{*}X & 0 \\ 0 & I_{\mathcal{N}^{\perp}} + XX^{*} \end{pmatrix}\,. \tag{1.4}$$

It is also easy to see that the operators T and T^{*} each have a bounded inverse. Indeed, the spectrum of the skew-symmetric operator

$$Y := \begin{pmatrix} 0 & -X^{*} \\ X & 0 \end{pmatrix}$$

is a subset of the imaginary axis, so that zero belongs to the resolvent sets of $T = I_{\mathcal{H}} + Y$ and $T^{*} = I_{\mathcal{H}} - Y$, cf. [5, Theorem 5.5 (i)]. Hence, the partial isometry U from the polar decomposition $T = U|T|$ is unitary and can be represented as

$$(1.5)\qquad U = \begin{pmatrix} (I_{\mathcal{N}} + X^*X)^{-1/2} & -X^*(I_{\mathcal{N}^\perp} + XX^*)^{-1/2} \\ X(I_{\mathcal{N}} + X^*X)^{-1/2} & (I_{\mathcal{N}^\perp} + XX^*)^{-1/2} \end{pmatrix} .$$

In particular, U takes $\mathcal{N}$ to $\mathcal{G}(\mathcal{N}, X)$ and $\mathcal{N}^\perp$ to $\mathcal{G}(\mathcal{N}^\perp, -X^*)$, respectively, and the orthogonal projection Q onto $\mathcal{G}(\mathcal{N}, X)$ can be represented as

$$Q = UPU^* = \begin{pmatrix} (I_{\mathcal{N}} + X^*X)^{-1} & X^*(I_{\mathcal{N}^\perp} + XX^*)^{-1} \\ X(I_{\mathcal{N}} + X^*X)^{-1} & XX^*(I_{\mathcal{N}^\perp} + XX^*)^{-1} \end{pmatrix} ,$$

cf. [27, Remark 3.6] and also [49, Exercise 3.5.1].

A well-known characterization of the pairs of orthogonal projections P and Q in $\mathcal{H}$ for which $\operatorname{Ran} Q = \mathcal{G}(\operatorname{Ran} P, X)$ for some $X \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$ is given in Proposition 1.13 below.

1.4 Operator Riccati equations

There exist various approaches to studying operator Riccati equations, see, e.g., [6, Section 5] and references therein. In the framework of the present thesis, operator Riccati equations appear when considering graph subspaces which are reducing for a self-adjoint operator, see, e.g., [5, Section 5]. The corresponding results have valuable applications in perturbation theory for subspaces in general and throughout this thesis in particular. These results are revisited in Chapter 4 below.

In this section, we briefly recall the concept of strong solutions to operator Riccati equations.

Definition 1.4. Let A_0 and A_1 be closed densely defined operators on Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. A bounded operator $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ is called a *strong solution* to the operator Riccati equation

$$(1.6)\quad XA_0 - A_1X + XDX - E = 0\,, \quad D \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_0)\,, \quad E \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)\,,$$

if

$$\operatorname{Ran}\bigl(X|_{\operatorname{Dom}(A_0)}\bigr) \subset \operatorname{Dom}(A_1)$$

and

$$XA_0g - A_1Xg + XDXg - Eg = 0 \quad \text{for} \quad g \in \operatorname{Dom}(A_0)\,.$$

Along with (1.6), we also introduce the dual equation

$$YA_1^* - A_0^*Y + YD^*Y - E^* = 0\,, \tag{1.7}$$

for which the notion of strong solutions is analogous to that in Definition 1.4.

We have the following relationship between the Riccati equation (1.6) and the dual equation (1.7).

Lemma 1.5 ([6, Lemma 5.3])**.** *Let A_0 and A_1 be as in Definition 1.4. A bounded operator $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ is a strong solution to the Riccati equation* (1.6) *if and only if the operator $Y = -X^*$ is a strong solution to the dual Riccati equation* (1.7)*.*

1.5 Separation of two closed subspaces

In this section, we recall the notions of the operator angle and a direct rotation associated with a pair of closed subspaces. A more detailed discussion on this material can be found in [8, 19, 21, 24, 27, 40] and the references therein.

1.5.1 The operator angle

This subsection agrees, in essence, with parts of Section 2 of the author's article [50].

Let P and Q be two orthogonal projections in the Hilbert space $\mathcal{H}$. Following [19], we introduce the *closeness operator*

$$C := C(P,Q) := PQP + P^\perp Q^\perp P^\perp$$

and the *separation operator*

$$S := S(P,Q) := PQ^\perp P + P^\perp QP^\perp\,.$$

Since P and Q are self-adjoint, C and S are self-adjoint as well. Moreover, one has

$$0 \le C \le 1\,, \quad 0 \le S \le 1\,, \quad \text{and} \quad C + S = I_{\mathcal{H}}\,. \tag{1.8}$$

The operator angle with respect to P and Q can now be introduced via the functional calculus as follows.

Definition 1.6. Let P and Q be two orthogonal projections in a Hilbert space $\mathcal{H}$. Then, the operator

$$\Theta := \Theta(P,Q) := \arccos\bigl(\sqrt{C(P,Q)}\,\bigr) \tag{1.9}$$

is called the *operator angle* associated with the subspaces $\operatorname{Ran} P$ and $\operatorname{Ran} Q$.

Clearly, the operator angle Θ is self-adjoint and its spectrum lies in the interval $\bigl[0,\frac{\pi}{2}\bigr]$. Furthermore, taking into account (1.8) and (1.9), the operators C and S can be represented as

$$C = \cos^2\Theta \quad\text{and}\quad S = \sin^2\Theta\,. \tag{1.10}$$

Note that one has $C(P,Q) = C(P^\perp,Q^\perp)$, so that $\Theta(P,Q) = \Theta(P^\perp,Q^\perp)$.

It should be mentioned that in many works such as [27] and [30] the operator angle is introduced in a slightly different way. There, instead of Θ in (1.9), its restriction to $\operatorname{Ran} P$, or even to the maximal subspace of $\operatorname{Ran} P$ where it has trivial kernel, is considered. The above definition follows the approach by Davis and Kahan (cf. [21, Eqs. (1.16) and (1.17)]; see also [21, p. 17]) and provides a generalization of their notion of the operator angle. In fact, the definition (1.9) is universal in the sense that it does not require that a unitary operator taking $\operatorname{Ran} P$ to $\operatorname{Ran} Q$ exists.

As in [2, Section 34], one has

$$P-Q = P(I_\mathcal{H}-Q)-(I_\mathcal{H}-P)Q = PQ^\perp - P^\perp Q = Q^\perp P - QP^\perp\,,$$

so that

$$\begin{aligned}(P-Q)^2 &= \bigl(PQ^\perp - P^\perp Q\bigr)\bigl(Q^\perp P - QP^\perp\bigr)\\ &= PQ^\perp P + P^\perp Q P^\perp = S = \sin^2\Theta\,,\end{aligned}$$

that is,

$$|P-Q| = \sin\Theta\,. \tag{1.11}$$

In particular,

$$\|P-Q\| = \|\sin\Theta\| = \sin\|\Theta\| \le 1\,. \tag{1.12}$$

Thus, suitable norms of the operator angle Θ or of trigonometric functions thereof can be used to measure the difference between the subspaces $\operatorname{Ran} P$ and $\operatorname{Ran} Q$.

The operator norm of the angle operator is of particular importance in the present thesis.

Definition 1.7. Let P and Q be as in Definition 1.6. The quantity

$$\theta(P,Q) := \|\Theta(P,Q)\| = \arcsin\bigl(\|P-Q\|\bigr)$$

is called the *maximal angle* between the subspaces $\operatorname{Ran} P$ and $\operatorname{Ran} Q$.

The concept of the maximal angle between two closed subspaces has a long history. A short survey of this topic can be found, for example, in [8, Section 2].

In the framework of this thesis, one of the most important properties of the maximal angle is that it satisfies a triangle inequality: If P, Q, and R are orthogonal projections in a Hilbert space, then

$$\theta(P,Q) \le \theta(P,R) + \theta(R,Q)\,, \tag{1.13}$$

see [16, Corollary 4] and also [8, Lemma 2.15]. As already observed in [16], this inequality is stronger than the triangle inequality for the operator norm since $\sin(\theta_1+\theta_2) < \sin(\theta_1) + \sin(\theta_2)$ unless θ_1 or θ_2 is 0.

As a consequence of (1.13), the maximal angle defines a metric on the set of orthogonal projections, the so-called *angular metric*. An alternative proof of the corresponding triangle inequality (1.13) based on the joint work [36] with K. A. Makarov is provided in Chapter 5 below.

1.5.2 Direct rotations

The concept of direct rotations from one closed subspace of a Hilbert space to another was suggested by Davis [19] and Kato [25, Sections I.4.6 and I.6.8], but can yet be traced back to Sz.-Nagy [45, §105]. We adopt the following definition.

Definition 1.8 (cf. [21, Proposition 3.3]; see also [8, Definition 2.9])**.** Let P and Q be two orthogonal projections in the Hilbert space $\mathcal{H}$. A unitary operator $U \in \mathcal{L}(\mathcal{H})$ is called a *direct rotation* from $\operatorname{Ran} P$ to $\operatorname{Ran} Q$ if

$$QU = UP\,, \quad U^2 = (Q - Q^{\perp})(P - P^{\perp})\,, \quad \text{and} \quad \operatorname{Re} U \geq 0\,,$$

where $\operatorname{Re} U = (U + U^*)/2$ denotes the real part of U.

Surely, a direct rotation exists only if $\dim \operatorname{Ran} P = \dim \operatorname{Ran} Q$ and $\dim \operatorname{Ran} P^{\perp} = \dim \operatorname{Ran} Q^{\perp}$, but this is not sufficient if $\operatorname{Ran} P$ and $\operatorname{Ran} P^{\perp}$ are both infinite-dimensional, see Proposition 1.10 below and the remark to Proposition 3.2 in [21]. We introduce the following notions.

Definition 1.9 ([19], [21, Definition 3.2], [8, Definition 2.5])**.** Let P and Q be two orthogonal projections in the Hilbert space $\mathcal{H}$. The subspaces $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are said to be *equivalently positioned* if

$$\dim\bigl(\operatorname{Ran} P \cap \operatorname{Ran} Q^{\perp}\bigr) = \dim\bigl(\operatorname{Ran} P^{\perp} \cap \operatorname{Ran} Q\bigr)\,,$$

and they are in the *acute case* if

$$\operatorname{Ran} P \cap \operatorname{Ran} Q^{\perp} = \operatorname{Ran} P^{\perp} \cap \operatorname{Ran} Q = \{0\}\,.$$

Finally, $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are said to be in the *acute-angle* case if the corresponding maximal angle satisfies $\theta(P,Q) < \pi/2$, that is, if

$$\|P - Q\| < 1\,.$$

Clearly, if $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are in the acute-angle case, then they are in the acute case, and if they are in the acute case, then they are equivalently positioned. It should also be mentioned that the relation of being equivalently positioned is not transitive if the underlying Hilbert space is infinite-dimensional, see the discussion at the end of Section 3 in [19]. Similarly, the other two notions in Definition 1.9 are not transitive as well.

We have the following result due to Davis and Kahan.

Proposition 1.10 ([21, Propositions 3.1 and 3.2]; cf. [40, Theorem 2.14])**.** *Let P and Q be two orthogonal projections in the Hilbert space $\mathcal{H}$. Then, a direct rotation from* $\operatorname{Ran} P$ *to* $\operatorname{Ran} Q$ *exists if and only if* $\operatorname{Ran} P$ *and* $\operatorname{Ran} Q$ *are equivalently positioned. The direct rotation is unique if and only if* $\operatorname{Ran} P$ *and* $\operatorname{Ran} Q$ *are in the acute case.*

Remark 1.11. If $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are in the acute-angle case, then the operator $C = \cos^2\Theta$ has a bounded inverse. In this case, the direct rotation U from $\operatorname{Ran} P$ to $\operatorname{Ran} Q$ is explicitly given by

$$U = C^{-1/2} \cdot \bigl(QP + Q^\perp P^\perp\bigr)\,,$$

cf. [25, Theorem I.6.32]. This representation extends to the acute case. It is also the core of Davis' construction of a direct rotation in the case where $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are equivalently positioned, see [19, Section 3].

From a geometric point of view, direct rotations are of great importance. For instance, of all unitaries W taking $\operatorname{Ran} P$ to $\operatorname{Ran} Q$, direct rotations differ least from the identity, that is, the quantity $\|I_{\mathcal{H}} - W\|$ is minimized if W is a direct rotation, see [19, Theorem 7.1]. Moreover, direct rotations allow one to interpret the operator angle $\Theta = \Theta(P,Q)$ as an operator-valued rotation angle: Let U be a direct rotation from $\operatorname{Ran} P$ to $\operatorname{Ran} Q$. Upon observing that

$$(Q - Q^\perp)(P - P^\perp) + (P - P^\perp)(Q - Q^\perp) = 2C(P,Q) - 2S(P,Q)\,,$$

it is straightforward to verify that

$$\operatorname{Re} U = \sqrt{C} = \cos\Theta\,. \tag{1.14}$$

It is also easy to see that the skew-symmetric operator $(U - U^*)/2$ has a polar decomposition

$$\frac{1}{2}(U - U^*) = J \sin\Theta\,,$$

where J is a skew-symmetric partial isometry such that J^*J is the orthogonal projection onto $\overline{\operatorname{Ran}\sin\Theta} = \overline{\operatorname{Ran}\Theta}$, cf. [25, Section VI.2.7]. Moreover, J is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$, that is,

$$PJP = P^\perp J P^\perp = 0\,.$$

In addition, J commutes with $\sin\Theta$ and therefore also with Θ. Altogether, one concludes that U can be represented as

$$U = \cos\Theta + J\sin\Theta = \exp(J\Theta)\,, \tag{1.15}$$

where $J\Theta$ is skew-symmetric, satisfies $|J\Theta| = \Theta$, and is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$, cf. [21, Eq. (1.18)]. In particular, the operator angle $\Theta = \Theta(P,Q)$ has indeed a natural interpretation as a rotation angle if $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are equivalently positioned, cf. [50, Remark 2.1].

The following example illustrates that (1.15) characterizes the form of a direct rotation.

Example 1.12. Let P be an orthogonal projection in a Hilbert space $\mathcal{H}$, and let $Y \in \mathcal{L}(\mathcal{H})$, $\|Y\| \le \pi/2$, be skew-symmetric and off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$, that is,

$$Y^* = -Y \quad \text{and} \quad PYP = P^\perp Y P^\perp = 0\,.$$

Then, the unitary operator $U := \exp(Y)$ is a direct rotation from $\operatorname{Ran} P$ to $\operatorname{Ran}(U|_{\operatorname{Ran} P})$, and the associated operator angle $\Theta(P, UPU^*)$ is given by

$$\Theta(P, UPU^*) = |Y|\,.$$

Proof. Denote the orthogonal projection onto $\operatorname{Ran}(U|_{\operatorname{Ran} P})$ by $Q := UPU^*$. By definition, one has $QU = UP$. Moreover, one observes that

$$2\operatorname{Re} U = U + U^* = \exp(Y) + \exp(-Y) = 2\cos|Y| \ge 0\,, \tag{1.16}$$

where we have taken into account that $Y^2 = -Y^*Y = -|Y|^2$. Using the identities $PY = YP^\perp$ and $P^\perp Y = YP$, a straightforward computation shows that

$$(P - P^\perp)U^* = U(P - P^\perp)\,,$$

so that

$$(Q - Q^\perp)(P - P^\perp) = U(P - P^\perp)U^*(P - P^\perp) = U^2\,.$$

Thus, U is a direct rotation from $\operatorname{Ran} P$ to $\operatorname{Ran} Q = \operatorname{Ran}(U|_{\operatorname{Ran} P})$.

For the associated operator angle $\Theta = \Theta(P,Q)$ one concludes from (1.14) and (1.16) that

$$\cos\Theta = \operatorname{Re} U = \cos|Y|\,.$$

Hence, $\Theta = |Y|$ since $\|Y\| \le \pi/2$. This completes the proof. □

Using the representations

$$\Theta = \begin{pmatrix} \Theta_0 & 0 \\ 0 & \Theta_1 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 & -J_0^* \\ J_0 & 0 \end{pmatrix}$$

with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$, the direct rotation (1.15) may be written as

$$U = \begin{pmatrix} \cos\Theta_0 & -J_0^* \sin\Theta_1 \\ J_0 \sin\Theta_0 & \cos\Theta_1 \end{pmatrix} \tag{1.17}$$

with $J_0^* \sin\Theta_1 = (\sin\Theta_0)J_0^*$, cf. [21, Section 3]. In particular, one has $\|\sin\Theta_0\| = \|\sin\Theta_1\|$ and, therefore, $\|\Theta_0\| = \|\Theta_1\| = \|\Theta\|$.

Taking into account representation (1.17), one clearly has

$$\operatorname{Ran} Q = \operatorname{Ran}\bigl(U|_{\operatorname{Ran} P}\bigr) = \{\cos\Theta_0 x \oplus J_0 \sin\Theta_0 x \mid x \in \operatorname{Ran} P\}\,.$$

Moreover, if the subspaces $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are in the acute-angle case, then $\|\Theta_0\| = \|\Theta\| < \pi/2$, so that the operator $\cos\Theta_0$ has a bounded inverse. In this case,

$$\operatorname{Ran} Q = \{x \oplus J_0 \tan\Theta_0 x \mid x \in \operatorname{Ran} P\}\,,$$

that is, $\operatorname{Ran} Q$ is the graph of the bounded operator

$$X := J_0 \tan\Theta_0 \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)\,. \tag{1.18}$$

In particular, one has

$$\|X\| = \tan\|\Theta_0\| = \tan\|\Theta\| = \|\tan\Theta\|\,.$$

Conversely, if $\operatorname{Ran} Q = \mathcal{G}(\operatorname{Ran} P, X)$ for some $X \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$, then one can show that $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are in the acute-angle case, see, e.g., [18, Theorem 1]. In view of (1.12), this leads to the following well-known result.

Proposition 1.13 ([27, Corollary 3.4])**.** *Let P and Q be two orthogonal projections in the Hilbert space $\mathcal{H}$. The subspaces* $\operatorname{Ran} P$ *and* $\operatorname{Ran} Q$ *are in the acute-angle case if and only if one has* $\operatorname{Ran} Q = \mathcal{G}(\operatorname{Ran} P, X)$ *for some*

$X \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$. *In this case,*

$$\|P - Q\| = \frac{\|X\|}{\sqrt{1 + \|X\|^2}}$$

and, equivalently,

$$\|X\| = \frac{\|P - Q\|}{\sqrt{1 - \|P - Q\|^2}}\,.$$

Remark 1.14. In view of (1.18) and representation (1.17), it is easy to verify that in the situation of Proposition 1.13 the unitary operator (1.5) agrees with the direct rotation (1.17) from $\operatorname{Ran} P$ to $\operatorname{Ran} Q$.

1.6 Smooth paths of operators

Given fixed Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ and some bounded or unbounded interval $I \subset \mathbb{R}$, we consider operator-valued functions

$$I \ni t \mapsto B_t\,,$$

where each B_t is a densely defined operator from $\mathcal{H}$ to $\mathcal{K}$ on the same domain, that is,

$$\operatorname{Dom}(B_t) = \operatorname{Dom}(B_s) \quad \text{for} \quad s, t \in I\,. \tag{1.19}$$

The condition (1.19) ensures that the identity $B_s = B_t + (B_s - B_t)$ holds for all $s, t \in I$ as an operator equality. This allows to introduce the standard notions of *continuous*, *uniformly continuous*, $\mathcal{C}^1$*-smooth*, and *piecewise* $\mathcal{C}^1$*-smooth* paths of operators with respect to the operator norm on the dense subspace $\operatorname{Dom}(B_t)$. Here, every piecewise $\mathcal{C}^1$-smooth path is supposed to be continuous, and every continuous path clearly is uniformly continuous on compact subintervals. In particular, for a continuous path $t \mapsto B_t$ the difference $B_t - B_s$ is always bounded. The derivative of a (piecewise) $\mathcal{C}^1$-smooth path $t \mapsto B_t$ at $t \in I$ is denoted by $\dot{B}_t$ with $\operatorname{Dom}(\dot{B}_t) := \operatorname{Dom}(B_t)$. Sometimes, we also write $\frac{\mathrm{d}}{\mathrm{d}t} B_t$ instead of $\dot{B}_t$. Note that $\dot{B}_t$ is bounded on $\operatorname{Dom}(B_t)$, so that its closure satisfies $\overline{\dot{B}_t} \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ with $\|\overline{\dot{B}_t}\| = \|\dot{B}_t\|$.

The following examples of $\mathcal{C}^1$-smooth paths play a distinguished role throughout this thesis. Another, yet more technical, example is discussed in Lemma 3.12 below.

Example 1.15. Let $\mathcal{H}$ be a Hilbert space and $I \subset \mathbb{R}$ an arbitrary interval.

(a) For every densely defined operator A on $\mathcal{H}$ and every $V \in \mathcal{L}(\mathcal{H})$ the path

$$I \ni t \mapsto A + tV$$

is $\mathcal{C}^1$-smooth with $\frac{\mathrm{d}}{\mathrm{d}t}(A + tV) = V|_{\mathrm{Dom}(A)}$.

(b) For every $Y \in \mathcal{L}(\mathcal{H})$ the path

$$I \ni t \mapsto \exp(tY) = \sum_{k=0}^{\infty} \frac{t^k}{k!} Y^k \in \mathcal{L}(\mathcal{H})$$

is $\mathcal{C}^1$-smooth with $\frac{\mathrm{d}}{\mathrm{d}t} \exp(tY) = Y \exp(tY) = \exp(tY)Y$.

We need the following standard estimate for $\mathcal{C}^1$-smooth paths. For the sake of completeness, a short proof is provided.

Lemma 1.16. *Let $I \ni t \mapsto B_t$ be a $\mathcal{C}^1$-smooth path of densely defined operators between Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$. Then*

$$\|B_t - B_s\| \le \int_s^t \|\dot{B}_\tau\| \, \mathrm{d}\tau \qquad \textit{whenever} \qquad s \le t \, .$$

Proof. For arbitrary $x \in \mathrm{Dom}(B_t)$ and $y \in \mathcal{K}$, the scalar function

$$I \ni \tau \mapsto \langle y, B_\tau x \rangle$$

is $\mathcal{C}^1$-smooth with $\frac{\mathrm{d}}{\mathrm{d}\tau}\langle y, B_\tau x \rangle = \langle y, \dot{B}_\tau x \rangle$. For $s \le t$ this implies that

$$\langle y, (B_t - B_s)x \rangle = \int_s^t \langle y, \dot{B}_\tau x \rangle \, \mathrm{d}\tau \, ,$$

so that

$$|\langle y, (B_t - B_s)x \rangle| \le \int_s^t |\langle y, \dot{B}_\tau x \rangle| \, \mathrm{d}\tau \le \|x\| \, \|y\| \int_s^t \|\dot{B}_\tau\| \, \mathrm{d}\tau \, .$$

This proves the claim. □

In the framework of the present thesis, smooth paths of orthogonal projections are of particular interest. For those paths, a considerably stronger estimate than the one in Lemma 1.16 is available, which is closely related to the fact that the maximal angle satisfies a triangle inequality (see equation (1.13)). This is discussed in detail in Chapter 5 below.

1.7 Perturbation of the spectrum

We close this chapter with a detailed discussion of the variation of the spectrum of a self-adjoint operator under a bounded additive perturbation. The following well-known lemma represents the main result in this context.

Lemma 1.17 (see [25, Theorem V.4.10])**.** *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$, and let $V \in \mathcal{L}(\mathcal{H})$. Then, the spectrum of the perturbed operator $A+V$ is contained in the closed $\|V\|$-neighbourhood of the spectrum of A, that is,*

$$\operatorname{spec}(A+V) \subset \overline{\mathcal{O}_{\|V\|}\bigl(\operatorname{spec}(A)\bigr)}\,.$$

The property of the spectrum described by Lemma 1.17 is called the *upper semicontinuity* of the spectrum, see [25, Section IV.3.1–IV.3.2]. It implies that the spectrum of A does not expand by much when A is subjected to a small bounded perturbation. But, as described in [25, Section IV.3.2], the spectrum is not *lower semicontinuous* in general, so that it may very well shrink suddenly. However, if, in addition to the hypotheses of Lemma 1.17, the perturbation V is assumed to be self-adjoint as well, then the roles of A and $A+V$ can be switched via the identity $A = (A+V)-V$, so that the spectrum does also not shrink by much under the perturbation. Hence, the spectrum changes continuously when A varies over self-adjoint operators, cf. [25, Remark V.4.9].

Isolated parts of the spectrum

In the situation of Lemma 1.17, suppose that the spectrum of A contains an isolated component σ that has distance $d > 0$ from the remainder of the spectrum. In this case, it is a natural question whether the spectrum of the perturbed operator $A + V$ also has an isolated component, provided that the norm of the perturbation is small enough. More specifically: Is the set $\operatorname{spec}(A+V)\cap\mathcal{O}_{d/2}(\sigma)$ nonempty if V satisfies $\|V\| < d/2$? In the case where the perturbation V is assumed to be self-adjoint as well, this question can be answered affirmatively. More precisely, by switching the roles of A and

$A+V$ via $A=(A+V)-V$, we have the following well-known corollary to Lemma 1.17.

Corollary 1.18. *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$ such that the spectrum has an isolated component σ that has distance $d>0$ from the remainder* $\operatorname{spec}(A)\setminus\sigma$ *of the spectrum. Moreover, let $V\in\mathcal{L}(\mathcal{H})$ be self-adjoint. If $\|V\|<d/2$, then*

$$\operatorname{spec}(A+V)\cap\mathcal{O}_{d/2}(\sigma)=\operatorname{spec}(A+V)\cap\overline{\mathcal{O}_{\|V\|}(\sigma)}$$

is a nonempty isolated component of the spectrum of $A+V$.

As a consequence of Corollary 1.18, the perturbation V does not close gaps in the spectrum of A that are larger than $2\|V\|$. Recall that by a *gap* of a closed set $\Delta\subset\mathbb{R}$ one means an open interval in $\mathbb{R}$ that does not intersect Δ but the endpoints of which belong to Δ. The gap is said to be *finite* if this interval is bounded.

Clearly, the gap non-closing condition $\|V\|<d/2$ is sharp in the following sense: If in the situation of Corollary 1.18 one has $\|V\|\geq d/2$ instead of $\|V\|<d/2$, then the set $\operatorname{spec}(A+V)\cap\mathcal{O}_{d/2}(\sigma)$ may be empty or may not be separated from the remainder of the spectrum of $A+V$. In fact, in this case, the spectrum of $A+V$ may in general even have no isolated components at all.

However, under certain additional assumptions on the perturbation V, the gap non-closing condition $\|V\|<d/2$ can be relaxed considerably. This is the case, for example, if V is *semidefinite*, that is, if $V\geq 0$ or $V\leq 0$, or if V is *off-diagonal* with respect to the decomposition $\mathcal{H}=\operatorname{Ran}\mathsf{E}_A(\sigma)\oplus\mathsf{E}_A(\Sigma)$, $\Sigma:=\operatorname{spec}(A)\setminus\sigma$, that is, if

$$\mathsf{E}_A(\sigma)V\mathsf{E}_A(\sigma)=0=\mathsf{E}_A(\Sigma)V\mathsf{E}_A(\Sigma)\,.$$

These particular cases are discussed in the remaining part of this section.

We begin with the following well-known, yet remarkable, result, which applies in the case where V is off-diagonal and the convex hulls of the spectral components σ and Σ are disjoint, that is, $\sup\sigma<\inf\Sigma$ or vice versa.

Proposition 1.19 ([1, Theorem 2.1]; see also [21, Theorem 8.1])**.** *Let A be a self-adjoint operator such that its resolvent set contains an interval (a,b), $a<b$. Moreover, let $V\in\mathcal{L}(\mathcal{H})$ be off-diagonal with respect to the orthogonal*

decomposition $\mathcal{H} = \operatorname{Ran} \mathsf{E}_A\bigl((-\infty, a]\bigr) \oplus \operatorname{Ran} \mathsf{E}_A\bigl([b, \infty)\bigr)$*. Then, the interval* (a, b) *also belongs to the resolvent set of the perturbed operator* $A + V$*.*

The preceding proposition is surely of interest on its own, but it also plays a crucial part in obtaining the following two results.

The first one deals with the case of semidefinite perturbations and is extracted from the more general statement [55, Theorem 3.2]; cf. also [13, Eq. (9.4.4)].

Proposition 1.20. *Let* A *be as in Proposition 1.19, and let* $V \in \mathcal{L}(\mathcal{H})$ *be positive (resp. negative) semidefinite. If* $\|V\| < b - a$*, then the interval* $(a + \|V\|, b)$ *(resp.* $(a, b - \|V\|)$*) belongs to the resolvent set of the perturbed operator* $A + V$*.*

Proof. For the sake of completeness, we reproduce the proof.

Let $\|V\| < b - a$ and assume that V is positive semidefinite. The case where V is negative semidefinite can be treated analogously.

Denote $\mathcal{H}_- := \operatorname{Ran} \mathsf{E}_A\bigl((-\infty, a]\bigr)$ and $\mathcal{H}_+ := \operatorname{Ran} \mathsf{E}_A\bigl([b, \infty)\bigr)$, and decompose $V = V_{\text{diag}} + V_{\text{off}}$ into the sum of a diagonal part $V_{\text{diag}} = V_- \oplus V_+$ and an off-diagonal part V_{off} with respect to $\mathcal{H}_- \oplus \mathcal{H}_+$. Let $A_\pm := A|_{\mathcal{H}_\pm}$ be the parts of A associated with $\mathcal{H}_\pm$.

Since V is positive semidefinite, the diagonal part V_{diag} is also positive semidefinite, so that $V_\pm \geq 0$. Thus,

$$A_- + V_- \leq a + \|V\| < b \leq A_+ + V_+ \,.$$

In particular, the subspaces $\mathcal{H}_-$ and $\mathcal{H}_+$ are spectral subspaces for $A + V_{\text{diag}}$ associated with the sets $\bigl(-\infty, a + \|V\|\bigr]$ and $[b, \infty)$, respectively. Applying Proposition 1.19, one concludes that the interval $(a + \|V\|, b)$ belongs to the resolvent set of $A + V = A + V_{\text{diag}} + V_{\text{off}}$. □

The second result treats the general case of off-diagonal perturbations without any additional assumptions on the disposition of the spectral components σ and Σ.

Proposition 1.21 ([54, Proposition 2.5.22]; see also [31, Theorem 1.3])**.** *Let* A*,* V*, and* σ *be as in Corollary 1.18. Suppose, in addition, that* V *is off-diagonal with respect to the decomposition* $\mathcal{H} = \operatorname{Ran} \mathsf{E}_A(\sigma) \oplus \operatorname{Ran} \mathsf{E}_A(\Sigma)$*,*

$\Sigma := \operatorname{spec}(A) \setminus \sigma$. *Denote*

$$\delta_V := \|V\| \tan\Bigl(\frac{1}{2}\arctan\frac{2\|V\|}{d}\Bigr), \quad d = \operatorname{dist}(\sigma,\Sigma) > 0\,.$$

Then, the spectrum of $A+V$ is contained in the closed δ_V-neighbourhood of the spectrum of A, that is,

$$\operatorname{spec}(A+V) \subset \overline{\mathcal{O}_{\delta_V}\bigl(\operatorname{spec}(A)\bigr)}\,.$$

Moreover, if $\|V\| < \sqrt{3}d/2$, that is, $\delta_V < d/2$, then

$$\operatorname{spec}(A+V) \cap \mathcal{O}_{d/2}(\sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\delta_V}(\sigma)}$$

is a nonempty isolated component of the spectrum of $A+V$.

The following example of 4 × 4 matrices illustrates the statement of Proposition 1.21 and shows that the gap non-closing condition $\|V\| < \sqrt{3}d/2$ for off-diagonal perturbations is sharp.

Example 1.22 (cf. [31, Example 1.5]). On $\mathcal{H} = \mathbb{C}^4$ consider the 4×4 matrices

$$A = \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{array}\right) \quad \text{and} \quad V = \left(\begin{array}{cc|cc} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \alpha \\ \hline \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \end{array}\right), \quad \alpha \in \mathbb{R}\,.$$

Set $\sigma := \{2,4\}$ and $\Sigma := \operatorname{spec}(A) \setminus \sigma = \{1,3\}$, so that $d := \operatorname{dist}(\sigma,\Sigma) = 1$. Taking into account the identities

$$\delta_V = \alpha \tan\Bigl(\frac{1}{2}\arctan(2\alpha)\Bigr) = \frac{1}{2}\sqrt{1+4\alpha^2} - \frac{1}{2}\,,$$

it is straightforward to verify that the eigenvalues of the matrix $A+V$ are given by $\operatorname{spec}(A+V) = \omega \cup \Omega$ with

$$\omega := \{2+\delta_V, 4+\delta_V\} \subset \overline{\mathcal{O}_{\delta_V}(\sigma)} \quad \text{and} \quad \Omega := \{1-\delta_V, 3-\delta_V\} \subset \overline{\mathcal{O}_{\delta_V}(\Sigma)}\,.$$

In particular, if $\alpha = \sqrt{3}/2$, that is, $\delta_V = 1/2$, then $\omega = \{5/2, 9/2\}$ and $\Omega = \{1/2, 5/2\}$. In this case, the intersection $\operatorname{spec}(A+V) \cap \mathcal{O}_{1/2}(\sigma)$ is empty, and one has $\operatorname{dist}(\omega,\Omega) = 0$. The latter can be interpreted as the fact that the original gap between the components σ and Σ has been closed by the perturbation V.

Remark 1.23. Suppose, in addition to the hypotheses of Proposition 1.21, that the convex hull of σ is disjoint from the remainder of the spectrum, that is,

$$\operatorname{conv}(\sigma) \cap \Sigma = \varnothing\,.$$

It this case, one has the following stronger result: If $\|V\| < \sqrt{2}d$, that is, $\delta_V < d$, then

$$\operatorname{spec}(A+V) \cap \mathcal{O}_d(\sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\delta_V}(\sigma)}$$

is a nonempty isolated component of the spectrum of $A+V$, see [54, Proposition 2.5.22 (iii)] and also [31, Theorem 1.3 (iii)]. This stronger result is sharp in the same sense as Proposition 1.21 above, which can be seen from a suitable example of 3×3 matrices, see [54, Example 1.3.8] and [31, Example 1.6].

Chapter 2

The subspace perturbation problem. An overview

In the present chapter, an overview on the subspace perturbation problem for self-adjoint operators previously discussed in [7, 8, 26, 30, 31, 37, 51] is given. The problem is described in detail, and the main cases that appear in this context are introduced. For each of these cases, the results obtained so far are presented and discussed briefly. In particular, it is explained what contributions are made in this thesis.

Throughout this chapter, let A be a possibly unbounded self-adjoint operator on a Hilbert space $\mathcal{H}$ such that its spectrum is separated into two disjoint components, that is,

$$\operatorname{spec}(A) = \sigma \cup \Sigma \quad \text{with} \quad d := \operatorname{dist}(\sigma, \Sigma) > 0\,. \tag{2.1}$$

Moreover, let $V \in \mathcal{L}(\mathcal{H})$ be self-adjoint.

Under suitable additional assumptions on the operator V (see below), it can be guaranteed that the spectrum of the perturbed operator $A + V$ is likewise separated into two disjoint components,

$$\operatorname{spec}(A+V) = \omega \cup \Omega \quad \text{with} \quad \operatorname{dist}(\omega, \Omega) > 0\,, \tag{2.2}$$

where ω and Ω are contained in certain disjoint neighbourhoods of σ and Σ, respectively. In this sense, ω and Ω can be understood as perturbations of the original unperturbed components of $\operatorname{spec}(A)$, and the corresponding spectral subspaces $\operatorname{Ran} \mathsf{E}_{A+V}(\omega)$ and $\operatorname{Ran} \mathsf{E}_{A+V}(\Omega)$ can likewise be considered as perturbations of the unperturbed spectral subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} \mathsf{E}_A(\Sigma)$, respectively.

Conditions on V guaranteeing (2.2) are well understood in principle. They usually relate the norm of V and the distance d between the unperturbed spectral components σ and Σ. These conditions may depend on the disposition of the sets σ and Σ as well as on certain additional assumptions on the form of the perturbation, see below.

In this chapter, we focus on the problem under what possibly stronger conditions on V it can be ensured that the spectral subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} \mathsf{E}_{A+V}(\omega)$ are in the acute-angle case, that is, $\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}(\omega)\| < 1$ or, equivalently,

$$\theta = \arcsin\bigl(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}(\omega)\|\bigr) < \frac{\pi}{2}\,, \tag{2.3}$$

where $\theta = \theta(\mathsf{E}_A(\sigma), \mathsf{E}_{A+V}(\omega))$ is the maximal angle between the subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} \mathsf{E}_{A+V}(\omega)$, cf. Definition 1.7. In this concrete form, this problem has initially been discussed by Kostrykin, Makarov, and Motovilov in [26], but earlier works such as [21] by Davis and Kahan, [32] by Langer and Tretter, [4] by Adamjan, Langer, and Tretter, and [5] by Albeverio, Makarov, and Motovilov are closely related to this matter.

If inequality (2.3) holds, then a unique direct rotation $U = \exp(J\Theta)$ from $\operatorname{Ran} \mathsf{E}_A(\sigma)$ to $\operatorname{Ran} \mathsf{E}_A(\omega)$ exists, see Proposition 1.10 and equation (1.15). In this case, the associated operator angle $\Theta = \Theta(\mathsf{E}_A(\sigma), \mathsf{E}_{A+V}(\omega))$ can be interpreted as an operator-valued rotation angle between the subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} \mathsf{E}_A(\omega)$, and the corresponding maximal angle $\theta = \|\Theta\|$ serves as a measure for this rotation. As a consequence, one is not only interested in establishing (2.3), but also in sharp bounds on the maximal angle. Bounds of this sort usually have the form

$$\theta \le f\Bigl(\frac{\|V\|}{d}\Bigr)$$

with some function f independent of A and V.

Another perspective on the problem to establish (2.3) is given by the fact that (2.3) holds if and only if the subspace $\operatorname{Ran} \mathsf{E}_{A+V}(\omega)$ is the graph of a bounded linear operator X from the unperturbed subspace $\operatorname{Ran} \mathsf{E}_A(\sigma)$

to its orthogonal complement $\operatorname{Ran}\mathsf{E}_A(\Sigma)$, that is,

$$\operatorname{Ran}\mathsf{E}_{A+V}(\omega) = \mathcal{G}(\operatorname{Ran}\mathsf{E}_A(\sigma), X)\,, \tag{2.4}$$

see Proposition 1.13. This operator X satisfies

$$\|X\| = \frac{\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}(\omega)\|}{\sqrt{1 - \|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}(\omega)\|^2}} = \tan\theta\,. \tag{2.5}$$

Let A_0 and A_1 be the parts of A associated with $\operatorname{Ran}\mathsf{E}_A(\sigma)$ and $\operatorname{Ran}\mathsf{E}_A(\Sigma)$, respectively, and let

$$V = \begin{pmatrix} V_0 & W \\ W^* & V_1 \end{pmatrix} \tag{2.6}$$

be the representation of V as a 2×2 block operator matrix with respect to the decomposition $\mathcal{H} = \operatorname{Ran}\mathsf{E}_A(\sigma) \oplus \operatorname{Ran}\mathsf{E}_A(\Sigma)$. Taking into account that $\operatorname{Dom}(A_0 + V_0) = \operatorname{Dom}(A_0)$, $\operatorname{Dom}(A_1 + V_1) = \operatorname{Dom}(A_1)$, and

$$A + V = \begin{pmatrix} A_0 + V_0 & 0 \\ 0 & A_1 + V_1 \end{pmatrix} + \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix}\,, \tag{2.7}$$

it follows from [5, Lemma 5.3] (see also Corollary 4.9 below) that the operator X is a strong solution to the operator Riccati equation

$$X(A_0 + V_0) - (A_1 + V_1)X + XWX - W^* = 0\,. \tag{2.8}$$

This immediately widens the range of available methods to establish inequality (2.3) such as fixed point methods for the Riccati equation, see, e.g., [5, Section 3]. This connection to the operator Riccati equation also yields an explicit block diagonalization for the operator $A + V$ with respect to the decomposition $\mathcal{H} = \operatorname{Ran}\mathsf{E}_A(\sigma) \oplus \operatorname{Ran}\mathsf{E}_A(\Sigma)$, see Chapter 4 below.

The identity (2.7) illustrates a very important technique in the present context. Based on the representation (2.6), the perturbation V can be decomposed into the sum of a diagonal part V_{diag} and an off-diagonal part V_{off}, namely

$$V = V_{\text{diag}} + V_{\text{off}} := \begin{pmatrix} V_0 & 0 \\ 0 & V_1 \end{pmatrix} + \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix}\,. \tag{2.9}$$

Clearly, the subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} \mathsf{E}_A(\Sigma)$ are invariant for V_{diag}, so that the diagonal part of the perturbation only perturbs the spectrum and does not affect the subspaces. The off-diagonal part V_{off}, however, does change the subspaces and may also perturb the spectrum. Thus, the decomposition (2.9) can be used to reduce the consideration of V to the treatment of the off-diagonal part V_{off} provided that one has sufficient control over the spectrum of $A+V_{\mathrm{diag}}$, see, e.g., Section 2.4 below; see also the proofs of Proposition 1.20 and Proposition 7.9 in Chapter 7 below. In this sense, off-diagonal perturbations, that is, perturbations V with $V_{\mathrm{diag}} = 0$, play a very distinguished role when studying the rotation of spectral subspaces.

In what follows, the general separation condition (2.1) for $\operatorname{spec}(A)$ without any additional assumptions is referred to as the *generic case* or the *case of generic disposition.* We also discuss particular cases where additional assumptions on the mutual disposition of the spectral components σ and Σ are imposed, namely (see Fig. 2.1):

(1) The two components σ and Σ are *subordinated* in the sense that their convex hulls are disjoint, that is, $\sup \sigma < \inf \Sigma$ or vice versa.

or

(2) The two components σ and Σ are *annular separated*, that is, one of the components lies in a finite gap of the other one.

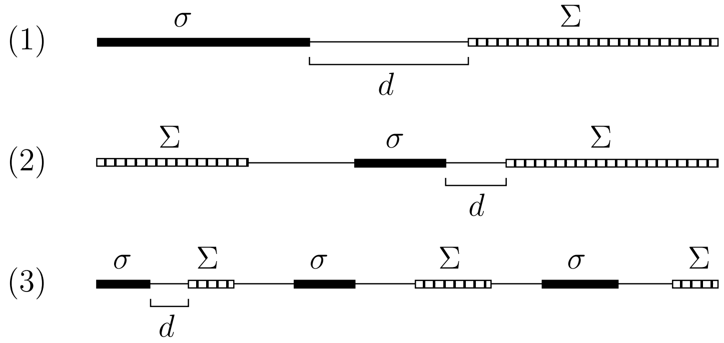

Fig. 2.1: Illustration of the three cases of spectral dispositions: (1) subordinated spectra with $\sup \sigma < \inf \Sigma$; (2) annular separated spectra where σ lies in a finite gap of Σ; (3) generic case.

The two particular dispositions (1) and (2) are the cases of *favourable* *geometry*, see [11, Section 3]. They play a very distinguished role in the context of this chapter since, in these cases, the problem of establishing inequality (2.3) has already been solved for a large class of perturbations, see Sections 2.1, 2.2, and 2.4 below. Note that in case of disposition (2) the assumption on the gap to be finite is needed to distinguish this case from the one of subordinated spectra. In fact, both dispositions (1) and (2) can be covered by the single condition that the convex hull of one of the components is disjoint from the other component, that is, $\operatorname{conv}(\sigma) \cap \Sigma = \varnothing$ or vice versa.

In the following sections we now discuss each of the three spectral dispositions in detail. Here, we distinguish between off-diagonal perturbations and general perturbations without any additional assumptions. Once the components ω and Ω of $\operatorname{spec}(A+V)$ have been chosen appropriately, θ and Θ always denote the maximal angle and the operator angle, respectively, associated with the subspaces $\mathsf{E}_A(\sigma)$ and $\mathsf{E}_{A+V}(\omega)$. Each section is closed with a concluding summary of the results. Finally, semidefinite perturbations are briefly discussed in the separate Section 2.4 as an outlook for future research.

2.1 Subordinated spectra

We begin with the case of subordinated spectra. For definiteness, assume that $\sup\sigma < \inf\Sigma$.

Off-diagonal perturbations

Suppose that the perturbation V is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran}\mathsf{E}_A(\sigma) \oplus \mathsf{E}_A(\Sigma)$. Then, regardless of the norm of V, the interval $(\sup\sigma, \inf\Sigma)$ belongs to the resolvent set of the perturbed operator $A+V$, see Proposition 1.19. In this case, the spectrum of $A+V$ is separated as in (2.2) with

$$\omega = \operatorname{spec}(A+V) \cap (-\infty, \sup\sigma] \quad \text{and} \quad \Omega = \operatorname{spec}(A+V) \cap [\inf\Sigma, \infty)\,,$$

and the Davis-Kahan $\tan 2\Theta$ theorem from [21] states that

$$\|\tan 2\Theta\| \le 2\,\frac{\|V\|}{d}\,.$$

This estimate is sharp (see [20, Theorem 5.1]) and can equivalently be rewritten as

$$\theta \leq \frac{1}{2}\arctan\Bigl(2\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4}\,, \tag{2.10}$$

see [21, Theorem 8.1]. In view of relation (2.5), the inequality $\theta < \pi/4$ in this situation also follows from the independent result [1, Theorem 2.3] by Adamjan and Langer, who proved that there is an operator X with $\|X\| < 1$ satisfying (2.4).

Of the four angle theorems by Davis and Kahan in [21], the $\tan 2\Theta$ theorem is probably the most studied one. Extensions to some unbounded off-diagonal perturbations V and even form perturbations have been considered in [40] and [23], respectively. The $\tan 2\Theta$ theorem has also been discussed under a relaxed condition on the subordinated spectral components allowing $\sup\sigma = \inf\Sigma$, see [28] for the case of bounded perturbations and [48] for the case of form perturbations; see also [4] and [39].

General perturbations

If no additional assumptions on the perturbation V are imposed, the optimal condition on $\|V\|$ that guarantees a spectral separation of the form (2.2) is $\|V\| < d/2$, see Corollary 1.18 and the discussion thereafter. In this case,

$$\omega = \operatorname{spec}(A+V) \cap \mathcal{O}_{d/2}(\sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\|V\|}(\sigma)} \tag{2.11}$$

and

$$\Omega = \operatorname{spec}(A+V) \cap \mathcal{O}_{d/2}(\Sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\|V\|}(\Sigma)}\,, \tag{2.12}$$

and it is a natural question whether the bound $\|V\| < d/2$ is sufficient to ensure (2.3). In the current situation, the answer to this question is affirmative. Indeed, the Davis-Kahan symmetric $\sin\Theta$ theorem [21, Proposition 6.1] states that

$$\|\sin\Theta\| \leq \frac{\|V\|}{\min\{\operatorname{dist}(\sigma,\Omega), \operatorname{dist}(\Sigma,\omega)\}}\,. \tag{2.13}$$

In view of the inequalities $\operatorname{dist}(\sigma,\Omega) \ge d - \|V\|$ and $\operatorname{dist}(\Sigma,\omega) \ge d - \|V\|$, this yields that

$$\theta \le \arcsin\Bigl(\frac{\|V\|}{d-\|V\|}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad \|V\| < \frac{d}{2}\,. \tag{2.14}$$

This bound was obtained in [26, Lemma 2.3 (i)] for the case where the operator A is additionally assumed to be bounded.

Nevertheless, there are stronger bounds on the maximal angle available. For instance, with the decomposition $V = V_{\mathrm{diag}} + V_{\mathrm{off}}$ as in equation (2.9), set

$$\widetilde{\omega} := \operatorname{spec}(A + V_{\mathrm{diag}}) \cap \mathcal{O}_{d/2}(\sigma) \quad \text{and} \quad \widetilde{\Omega} := \operatorname{spec}(A + V_{\mathrm{diag}}) \cap \mathcal{O}_{d/2}(\Sigma)\,.$$

Since $\|V_{\mathrm{diag}}\| \le \|V\| < d/2$, the sets $\widetilde{\omega}$ and $\widetilde{\Omega}$ are likewise subordinated with $\operatorname{dist}(\widetilde{\omega},\widetilde{\Omega}) \ge d - 2\|V\|$. Moreover, V_{diag} does not change the spectral subspaces $\operatorname{Ran}\mathsf{E}_A(\sigma)$ and $\operatorname{Ran}\mathsf{E}_A(\Sigma)$, that is, one has $\mathsf{E}_{A+V_{\mathrm{diag}}}(\widetilde{\omega}) = \mathsf{E}_A(\sigma)$ and $\mathsf{E}_{A+V_{\mathrm{diag}}}(\widetilde{\Omega}) = \mathsf{E}_A(\Sigma)$. The $\tan 2\Theta$ theorem therefore implies that

$$\theta \le \frac{1}{2}\arctan\biggl(\frac{2\|V_{\mathrm{off}}\|}{\operatorname{dist}\bigl(\widetilde{\omega},\widetilde{\Omega}\bigr)}\biggr) \le \frac{1}{2}\arctan\Bigl(\frac{2\|V\|}{d-2\|V\|}\Bigr) < \frac{\pi}{4} \tag{2.15}$$

for $\|V\| < d/2$, see [26, Lemma 2.3 (ii)]. Note that estimate (2.15) is considerably stronger than (2.14) if the quotient $\|V\|/d$ is not to small. If $\|V\|/d$ is small, then (2.14) gives slightly more accurate results.

However, an even stronger estimate on the maximal angle is provided by the Davis-Kahan $\sin 2\Theta$ theorem in [21], which states that

$$\|\sin 2\Theta\| \le 2\,\frac{\|V\|}{d}\,. \tag{2.16}$$

This estimate is sharp (see [20, Theorem 5.1] and also Remark 7.8 below) and can equivalently be rewritten as

$$\theta \le \frac{1}{2}\arcsin\Bigl(2\,\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4} \quad \text{for} \quad \|V\| < \frac{d}{2}\,, \tag{2.17}$$

see [21, Theorem 8.2]; cf. also Lemma 7.5 below. Note that the proof of the $\sin 2\Theta$ theorem in [21, Section 7] essentially uses the $\sin\Theta$ theorem, see also the proof of Theorem 7.1 in Chapter 7 below.

Conclusion

In the case of subordinated spectral components σ and Σ, the problem to find the least restrictive condition on the norm of V that establishes (2.2) and (2.3) is completely solved for both off-diagonal and general perturbations. It turns out that the condition on $\|V\|$ that guarantees (2.2) also implies (2.3). Moreover, sharp a priori bounds on the maximal angle are available, namely (2.10) for off-diagonal perturbations and (2.17) for general perturbations. In either case, the maximal angle is strictly less than $\pi/4$.

2.2 Annular separated spectra

In this section, we discuss the case of annular separated spectral components σ and Σ. For definiteness, assume that σ lies in a finite gap of Σ.

Off-diagonal perturbations

Suppose that V is off-diagonal with respect to $\mathcal{H} = \operatorname{Ran} \mathsf{E}_A(\sigma) \oplus \operatorname{Ran} \mathsf{E}_A(\Sigma)$. In contrast to the case of subordinated spectra, now a smallness assumption on $\|V\|$ is required in order to ensure a spectral separation of the form (2.2). By Remark 1.23, the optimal condition here is $\|V\| < \sqrt{2}d$. In this case, one can choose

$$\omega = \operatorname{spec}(A+V) \cap \mathcal{O}_d(\sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\delta_V}(\sigma)}$$

with

$$\delta_V = \|V\| \tan\Bigl(\frac{1}{2} \arctan \frac{2\|V\|}{d}\Bigr) < d \tag{2.18}$$

and

$$\Omega = \operatorname{spec}(A+V) \setminus \omega\,.$$

In this situation, it follows from the a posteriori $\tan\Theta$ theorem in [30], a generalization of the Davis-Kahan $\tan\Theta$ theorem, that

$$\|\tan\Theta\| \le \frac{\|V\|}{\operatorname{dist}(\omega,\Sigma)}\,, \tag{2.19}$$

see also [31, Lemma 2.3 and Theorem 2.4]); note that the quantity $\operatorname{dist}(\omega,\Sigma)$ here cannot be replaced by $\operatorname{dist}(\sigma,\Omega)$, see [30, Remark 4.1].

Using the inequality $\operatorname{dist}(\omega, \Sigma) \geq d - \delta_V$, one obtains from (2.19) that

$$\theta \leq \arctan\Bigl(\frac{\|V\|}{d-\delta_V}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad \|V\| < \sqrt{2}d\,,$$

see [31, Theorem 2.6]; cf. also [30, Theorem 5.1].

Recently, Albeverio and Motovilov have proved in [7] an a priori variant of the $\tan\Theta$ theorem. This variant yields the stronger sharp estimate

$$\theta \leq \arctan\Bigl(\frac{\|V\|}{d}\Bigr) < \arctan\sqrt{2} \quad \text{for} \quad \|V\| < \sqrt{2}d\,. \tag{2.20}$$

In particular, one has $\theta < \pi/4$ if $\|V\| < d$; in the case where A is additionally assumed to be bounded, the latter also follows from [30, Theorem 1 (ii)]. Note that for $\|V\| < d$, the bound (2.20) has already been shown in [40, Theorem 2].

General perturbations

For general perturbations V, the case of annular separated spectra is very similar to the case of subordinated spectra. Indeed, as long as $\|V\| < d/2$, the components ω and Ω of $\operatorname{spec}(A+V)$ can be chosen as in (2.11) and (2.12), respectively, and this condition on $\|V\|$ is optimal. Moreover, the $\sin\Theta$ and $\sin 2\Theta$ theorems remain valid in exactly the same form, so that one still has the bounds (2.14) and (2.17), and the latter is still sharp. Of course, the bound (2.15) is not available any more since the $\tan 2\Theta$ theorem does not apply for annular separated spectra. One can use the a posteriori $\tan\Theta$ theorem instead to obtain a bound on the maximal angle based on the decomposition $V = V_{\text{diag}} + V_{\text{off}}$, but the resulting estimate will be weaker than (2.17), so that we omit the details here. However, a similar reasoning is used for semidefinite perturbations in Section 2.4 below.

Conclusion

Also in the case of annular separated spectral components σ and Σ, the discussed problem is completely solved for both off-diagonal and general perturbations. Again, the condition on $\|V\|$ guaranteeing (2.2) also implies (2.3), and sharp a priori bounds on the maximal angle are available, namely (2.20) for off-diagonal perturbations and (2.17) for general perturbations. This time the inequality $\theta < \pi/4$ can be guaranteed only for general perturbations to the whole extent. However, for off-diagonal perturbations the maximal angle is known to be less than $\arctan\sqrt{2}$ and is thus still bounded away from $\pi/2$. The inequality $\theta < \pi/4$ here requires the stronger condition $\|V\| < d$.

2.3 The generic case

We now turn to the case where the unperturbed spectral components σ and Σ are in generic disposition, that is, no additional assumptions on σ and Σ other than (2.1) are imposed. This is the case the contributions in the present thesis deal with. Unlike the two preceding sections, we begin with general perturbations.

General perturbations

As before, for general perturbations the optimal condition on $\|V\|$ that guarantees a spectral separation of the form (2.2) is $\|V\| < d/2$ and, in this case, the components ω and Ω of $\operatorname{spec}(A + V)$ can be chosen as in (2.11) and (2.12), respectively.

One of the main differences between the generic case and the case of subordinated or annular separated spectra can be seen at the form of the symmetric $\sin\Theta$ theorem. In fact, the bound (2.13) is not available any more, but it does hold with an additional factor $\pi/2$, that is,

$$\|\sin\Theta\| \le \frac{\pi}{2}\,\frac{\|V\|}{\min\{\operatorname{dist}(\sigma,\Omega),\operatorname{dist}(\Sigma,\omega)\}}\,, \tag{2.21}$$

see the discussion in Section 3.2 below. In the same way as in (2.14), this yields that

$$\theta \le \arcsin\Bigl(\frac{\pi}{2}\,\frac{\|V\|}{d-\|V\|}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad 0 \le \|V\| < \frac{2d}{2+\pi}\,. \tag{2.22}$$

For the case where the operator A is additionally assumed to be bounded, the latter result was obtained in [26, Lemma 2.2].

Similarly, the bound from the $\sin 2\Theta$ theorem turns into

$$\|\sin 2\Theta\| \le \frac{\pi}{2}\cdot 2\,\frac{\|V\|}{d}\,, \tag{2.23}$$

see Theorem 7.1 below. A related estimate with the left-hand side of (2.23) replaced by $\sin 2\theta$ has previously been shown by Albeverio and Motovilov in [8, Corollary 4.3], see also Proposition 7.9 and the discussion in the introduction to Chapter 7 below.

In the current situation, the bound (2.23) can equivalently be rewritten as

$$\theta \le \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr) \le \frac{\pi}{4} \quad \text{for} \quad 0 \le \|V\| \le \frac{d}{\pi}\,, \tag{2.24}$$

see Corollary 7.2 below; cf. also [8, Remark 4.4].

In view of the bounds (2.22) and (2.24) and the inequalities $\frac{1}{\pi} < \frac{2}{2+\pi} < \frac{1}{2}$, it is unclear whether the condition $\|V\| < d/2$ this time is sufficient for the subspaces $\mathsf{E}_A(\sigma)$ and $\mathsf{E}_{A+V}(\omega)$ to be in the acute-angle case. Basically, the following problem arises:

What is the best possible constant $c_{\mathrm{opt}} \in \bigl(0, \frac{1}{2}\bigr]$ such that (2.3) holds whenever $\|V\| \le c_{\mathrm{opt}} \cdot d$?

This constant c_{opt} is supposed to be universal in the sense that it is independent of the operators A and V.

It has been conjectured that $c_{\mathrm{opt}} = 1/2$ (see [8]; cf. also [26] and [31]), but there is no proof available for this guess yet. So far, only lower bounds on c_{opt} can be given. For instance, it follows from (2.22) that

$$c_{\mathrm{opt}} \ge \frac{2}{2+\pi} = 0.3889845\ldots$$

In the joint work [37] with K. A. Makarov, a coupling parameter on the perturbation was introduced,

$$B_t := A + tV\,, \quad \mathrm{Dom}(B_t) := \mathrm{Dom}(A)\,, \quad t \in [0,1]\,,$$

with the idea to increase this parameter in small steps according to a suitably chosen partition of the interval $[0,1]$ and, thus, to iterate the estimate on the maximal angle by locally using the bound (2.22). Based on the triangle inequality for the maximal angle (equation (1.13); see also Chapter 5 below), the (more general) considerations in Chapter 6 yield the a posteriori bound

$$\theta \le \frac{\pi}{2}\|V\| \int_0^1 \frac{\mathrm{d}t}{\mathrm{dist}(\omega_t, \Omega_t)}\,, \tag{2.25}$$

where

$$\omega_t := \operatorname{spec}(B_t) \cap \mathcal{O}_{d/2}(\sigma) \quad \text{and} \quad \Omega_t := \operatorname{spec}(B_t) \cap \mathcal{O}_{d/2}(\Sigma)\,,$$

see equation (6.18) in Section 6.2 below; this result corresponds to the consideration of partitions of the interval $[0,1]$ with arbitrarily small mesh size, cf. Remark 8.2. The author's guess is that estimate (2.25) is optimal in general, but a rigorous proof for this guess is not available yet, see Conjecture 6.19 below and the corresponding discussion at the end of Chapter 6.

Taking into account the a priori type inequality $\operatorname{dist}(\omega_t, \Omega_t) \geq d - 2t\|V\|$, one obtains from (2.25) that

$$\theta \leq \frac{\pi}{4}\log\Bigl(\frac{d}{d-2\|V\|}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad 0 \leq \|V\| < \frac{\sinh(1)}{\mathrm{e}}\cdot d \tag{2.26}$$

and, therefore,

$$c_{\mathrm{opt}} \geq \frac{\sinh(1)}{\mathrm{e}} = 0.4323323\ldots\,,$$

see Theorem 6.15 (a) below and also [8, Theorem 3.5]; the case where A is additionally assumed to be bounded has previously been discussed in Theorem 3.2 of the joint work [37] with K. A. Makarov. Note that not only the lower bound on c_{opt} obtained from (2.26) is sharper than the one obtained from (2.22), but also estimate (2.26) on the maximal angle is stronger than (2.22), see Remark 6.16 below.

Although estimate (2.24) is valid only for $0 \leq \|V\| \leq \frac{d}{\pi} < \frac{2d}{2+\pi}$, the obtained bound on the maximal angle is substantially stronger than (2.22) and (2.26), that is, one has

$$\frac{1}{2}\arcsin\Bigl(\pi\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4}\log\Bigl(\frac{d}{d-2\|V\|}\Bigr) \quad \text{for} \quad 0 < \|V\| \leq \frac{d}{\pi}\,,$$

see Remark 7.7 below. In fact, for perturbations V satisfying $\|V\| \leq \frac{4d}{4+\pi^2}$, the bound (2.24) on the maximal angle is the strongest one available so far, cf. Remark 8.11 below and also [8, Remark 5.5].

At this point, Albeverio and Motovilov noticed in [8] that partitions of the interval $[0,1]$ with small mesh size do not give the best results. With a particular finite partition and a local use of the estimate (2.24), they obtained in [8, Theorem 5.4] that

$$\theta \le M_*\Big(\frac{\|V\|}{d}\Big) < \frac{\pi}{2} \quad \text{for} \quad 0 \le \|V\| < c_* \cdot d\,, \tag{2.27}$$

where

$$c_* = 16\,\frac{\pi^6 - 2\pi^4 + 32\pi^2 - 32}{(\pi^2+4)^4} = 0.4541692\ldots \tag{2.28}$$

and

$$M_*(x) = \begin{cases} \frac{1}{2}\arcsin(\pi x)\,, & 0 \le x \le \frac{4}{\pi^2+4}\,, \\ \frac{1}{2}\arcsin\big(\frac{4\pi}{\pi^2+4}\big) + \frac{1}{2}\arcsin\Big(\pi\,\frac{(\pi^2+4)x-4}{\pi^2-4}\Big)\,, & \frac{4}{\pi^2+4} < x \le \frac{8\pi^2}{(\pi^2+4)^2}\,, \\ \arcsin\big(\frac{4\pi}{\pi^2+4}\big) + \frac{1}{2}\arcsin\Big(\pi\,\frac{(\pi^2+4)^2x-8\pi^2}{(\pi^2-4)^2}\Big)\,, & \frac{8\pi^2}{(\pi^2+4)^2} < x \le c_*\,. \end{cases}$$

Albeverio and Motovilov also showed that estimate (2.27) is stronger than (2.26), see [8, Remark 5.5].

The present author noticed that there is a better choice for the finite partition of the interval $[0,1]$ and has formulated an optimization problem to obtain the best possible choice. The explicit solution to this optimization problem yields the bound

$$\theta \le N\Big(\frac{\|V\|}{d}\Big) < \frac{\pi}{2} \quad \text{for} \quad 0 \le \|V\| < c_{\mathrm{crit}} \cdot d\,, \tag{2.29}$$

where

$$c_{\mathrm{crit}} = \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{\sqrt{3}}{\pi}\Big)^3 = 0.4548399\ldots\,,$$

$N(x) = M_*(x) = \frac{1}{2}\arcsin(\pi x)$ for $0 \le x \le \frac{4}{\pi^2+4}$, and

$$N(x) = \begin{cases} \arcsin\Big(\sqrt{\frac{2\pi^2x-4}{\pi^2-4}}\Big) & \text{for} \quad \frac{4}{\pi^2+4} < x < 4\,\frac{\pi^2-2}{\pi^4}\,, \\ \arcsin\big(\frac{\pi}{2}(1-\sqrt{1-2x}\,)\big) & \text{for} \quad 4\,\frac{\pi^2-2}{\pi^4} \le x \le \kappa\,, \\ \frac{3}{2}\arcsin\big(\frac{\pi}{2}(1-\sqrt[3]{1-2x}\,)\big) & \text{for} \quad \kappa < x \le c_{\mathrm{crit}}\,, \end{cases}$$

see Theorem 8.9 below. Here, $\kappa \in \big(4\frac{\pi^2-2}{\pi^4}, 2\frac{\pi-1}{\pi^2}\big)$ is the unique solution to the equation

$$\arcsin\Big(\frac{\pi}{2}\big(1-\sqrt{1-2\kappa}\,\big)\Big) = \frac{3}{2}\arcsin\Big(\frac{\pi}{2}\big(1-\sqrt[3]{1-2\kappa}\,\big)\Big)$$

in the interval $\bigl(0, 2\frac{\pi-1}{\pi^2}\bigr]$. This choice of the constant κ ensures that the function N is continuous and as small as possible. In particular, one has

$$N(x) < M_*(x) \quad \text{for} \quad \frac{4}{\pi^2+4} < x \leq c_*$$

with c_* and M_* as in (2.27), see Remark 8.12 below. However, estimate (2.29) remains valid if κ is replaced by any other constant within the interval $\bigl(4\frac{\pi^2-2}{\pi^4}, 2\frac{\pi-1}{\pi^2}\bigr)$, see Remark 8.10 below. Numerical calculations yield that $\kappa = 0.4098623\ldots$

From (2.29) one immediately deduces that

$$c_{\mathrm{opt}} \geq c_{\mathrm{crit}} > c_* \,.$$

Together with the bound (2.29) on the maximal angle, this result is the strongest one obtained so far in the context of the generic spectral disposition (2.1) and general perturbations V. Since it corresponds to the solution of a suitable optimization problem, it is also best possible within the framework of iterating the estimate on the maximal angle with a local use of (2.24). As a consequence, one has to find an estimate substantially stronger than (2.24), at least for perturbations V with sufficiently small norm, in order to improve on the bound (2.29), see Remark 8.11 below.

Off-diagonal perturbations

One can refine the above considerations for general perturbations if more information on the variation of the spectrum under the perturbation is available. In particular, this is the case if the perturbation V is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} \mathsf{E}_A(\sigma) \oplus \operatorname{Ran} \mathsf{E}_A(\Sigma)$. For those perturbations V, the optimal condition on $\|V\|$ guaranteeing a spectral separation of the form (2.2) reads $\|V\| < \sqrt{3}d/2$, see Proposition 1.21 and Example 1.22. In this case, one can choose

$$\omega = \operatorname{spec}(A+V) \cap \mathcal{O}_{d/2}(\sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\delta_V}(\sigma)}$$

and

$$\Omega = \operatorname{spec}(A+V) \cap \mathcal{O}_{d/2}(\Sigma) = \operatorname{spec}(A+V) \cap \overline{\mathcal{O}_{\delta_V}(\Sigma)}$$

with δ_V as in (2.18). Recall that $\delta_V < d/2$ for $\|V\| < \sqrt{3}d/2$.

It is again a natural question whether the condition $\|V\| < \sqrt{3}d/2$ is sufficient for (2.3) to hold. Similar to the case of general perturbations, introduce the best possible constant $c_{\text{opt-off}} \in \bigl(0, \frac{\sqrt{3}}{2}\bigr]$ such that (2.3) holds for all off-diagonal perturbations V with $\|V\| < c_{\text{opt-off}} \cdot d$. It has been conjectured that

$$c_{\text{opt-off}} = \frac{\sqrt{3}}{2} = 0.8660254\ldots ,$$

see [31], but there is no proof available for this yet.

Based on fixed point theorems for the operator Riccati equation (2.8), it was shown in [5, Theorems 3.6 (i) and 7.6] that for $0 < \|V\| < d/\pi$ there is $X \in \mathcal{L}(\operatorname{Ran} \mathsf{E}_A(\sigma), \operatorname{Ran} \mathsf{E}_A(\Sigma))$ satisfying $\operatorname{Ran} \mathsf{E}_{A+V}(\omega) = \mathcal{G}(\operatorname{Ran} \mathsf{E}_A(\sigma), X)$ and

$$\|X\| \le \frac{d}{\|V\|}\Bigl(\frac{1}{\pi} - \sqrt{\frac{1}{\pi^2} - \frac{\|V\|^2}{d^2}}\Bigr) < 1 .$$

Taking into account (2.5), it is straightforward to verify that this estimate agrees with the bound (2.24) obtained from the $\sin 2\Theta$ theorem.

In contrast to the pure a priori result (2.24), the other results discussed for general perturbations can benefit from the additional knowledge on the perturbed spectral components ω and Ω. For instance, the symmetric $\sin\Theta$ theorem in the form (2.21) remains valid. In view of the inequalities $\operatorname{dist}(\sigma, \Omega) \ge d - \delta_V$ and $\operatorname{dist}(\Sigma, \omega) \ge d - \delta_V$, it yields that

$$\theta \le \arcsin\Bigl(\frac{\pi}{2}\,\frac{\|V\|}{d - \delta_V}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad 0 \le \|V\| < c_\pi d\,, \tag{2.30}$$

where

$$c_\pi = \frac{3\pi - \sqrt{\pi^2 + 32}}{\pi^2 - 4} = 0.5032886\ldots$$

This result was obtained in [31, Theorem 2.2] for the case where the operator A is additionally assumed to be bounded.

Similar to the case of general perturbations, the approach to iterate the bound on the maximal angle by introducing a coupling parameter on the perturbation can be used to get a result stronger than (2.30). This time, the spectral components ω_t and Ω_t in (2.25) admit the a priori type inequality $\operatorname{dist}(\omega_t, \Omega_t) \ge d - 2\delta_{tV}$, so that

$$\theta \le \frac{\pi}{2}\int_0^{\frac{\|V\|}{d}} \frac{\mathrm{d}\tau}{1 - 2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} < \frac{\pi}{2} \quad \text{for} \quad 0 \le \|V\| < c_{\text{off}}\, d\,, \tag{2.31}$$

where $c_{\mathrm{off}} = 0.6759893\ldots$ is given by

$$\int_0^{c_{\mathrm{off}}} \frac{\mathrm{d}\tau}{1 - 2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} = 1\,,$$

see Theorem 6.15 (b) below. The case where the operator A is additionally assumed to be bounded can also be found in Theorem 3.3 of the joint work [37] with K. A. Makarov.

Not only are the constants c_π and c_{off} above greater than $1/\pi$, this time also the bounds (2.30) and (2.31) are stronger than the a priori result (2.24) obtained from the $\sin 2\Theta$ theorem, cf. Lemma 8.23 (a) below. However, the bound (2.24) still plays an important role when considering a corresponding optimization problem for the choice of the partition of the interval $[0,1]$, see Section 8.3 below for details. For now, it suffices to note that this optimization problem is much harder to handle and is not solved explicitly yet. Nevertheless, based on numerical experiments, one can guarantee that

$$c_{\text{opt-off}} \geq 0.6940725\,, \tag{2.32}$$

see Corollary 8.26. A corresponding bound on the maximal angle is given in Example 8.25 below.

Conclusion

In contrast to the cases of subordinated and annular separated spectra discussed in Sections 2.1 and 2.2, respectively, the generic case is solved neither for off-diagonal perturbations nor for general ones. For the latter, the currently best known result is provided by (2.29). Since it corresponds to the solution of a suitable optimization problem, it is also best possible within the approach to iterate the bound on the maximal angle with a local use of (2.24).

For off-diagonal perturbations, the situation is more delicate. Estimate (2.31) provides a fairly reasonable bound on the maximal angle. A stronger but more technical and more involved result is available in form of (2.32) and the respective bound on the maximal angle discussed in Example 8.25 below. However, until the corresponding optimization problem can be solved explicitly, this should be considered only as an interim solution.

Note that, except for the bound (2.24), all results discussed in this section are derived from a posteriori type estimates using a priori knowledge on the perturbed spectrum. This is a potentially weak point in the whole approach of iterating the bound on the maximal angle, see the discussion at the end of Chapter 6. Nevertheless, the obtained results are the strongest ones obtained so far.

2.4 Semidefinite perturbations. An outlook

Suppose that V is positive semidefinite, that is, $V \geq 0$. The case where V is negative semidefinite can be treated analogously.

From Proposition 1.20 one concludes that the perturbation V moves the spectrum of A on the real axis only to the right and not to the left. As a consequence, the optimal condition on $\|V\|$ that guarantees a spectral separation of the form (2.2) reads $\|V\| < d$. In this case, the spectrum of $A+V$ is separated as

$$\operatorname{spec}(A+V) = \omega \cup \Omega \quad \text{with} \quad \operatorname{dist}(\omega, \Omega) \geq d - \|V\| > 0\,,$$

where ω and Ω are chosen as certain “right-side” neighbourhoods of σ and Σ, respectively. Namely,

$$\omega = \bigl\{\lambda \in \operatorname{spec}(A+V) \mid \sigma \cap [\lambda - \|V\|, \lambda] \neq \varnothing\bigr\}\,, \tag{2.33}$$

and analogously for Ω.

In the same way, with the decomposition $V = V_{\mathrm{diag}} + V_{\mathrm{off}}$ (cf. equation (2.9)) and $\|V_{\mathrm{diag}}\| \leq \|V\| < d$, the spectrum of $A + V_{\mathrm{diag}}$ is separated as

$$\operatorname{spec}(A+V_{\mathrm{diag}}) = \widetilde{\omega} \cup \widetilde{\Omega} \quad \text{with} \quad \operatorname{dist}(\widetilde{\omega}, \widetilde{\Omega}) \geq d - \|V\| > 0\,,$$

where $\widetilde{\omega}$ and $\widetilde{\Omega}$ are chosen analogously to ω and Ω above.

At first sight, the purely a priori type $\sin 2\Theta$ theorem does not benefit from the semidefiniteness of the perturbation. Also the a posteriori type $\sin\Theta$ theorem does not seem to gain anything from this additional knowledge on the perturbation. However, some of the results discussed above may be used to obtain a stronger bound on the maximal angle in this particular situation.

If the unperturbed spectral components σ and Σ are subordinated, then the components $\widetilde{\omega}$ and $\widetilde{\Omega}$ are also subordinated. Thus, analogously to (2.15), the $\tan 2\Theta$ theorem yields that

$$\theta \le \frac{1}{2}\arctan\Bigl(\frac{2\|V_{\mathrm{off}}\|}{\operatorname{dist}(\widetilde{\omega},\widetilde{\Omega})}\Bigr) \le \frac{1}{2}\arctan\Bigl(\frac{2\|V\|}{d-\|V\|}\Bigr) < \frac{\pi}{4} \quad \text{for} \quad \|V\| < d\,.$$

If the spectral components σ and Σ are annular separated, say, σ lies in a finite gap of Σ, then $\widetilde{\omega}$ and $\widetilde{\Omega}$ are annular separated as well. Moreover, the component ω lies in a finite gap of $\widetilde{\Omega}$ with $\operatorname{dist}(\omega,\widetilde{\Omega}) \ge d-\|V\|$. In this case, it follows from the a posteriori $\tan\Theta$ theorem [30, Theorem 2] (see equation (2.19)) that

$$\theta \le \arctan\Bigl(\frac{\|V_{\mathrm{off}}\|}{\operatorname{dist}(\omega,\widetilde{\Omega})}\Bigr) \le \arctan\Bigl(\frac{\|V\|}{d-\|V\|}\Bigr) < \frac{\pi}{2} \quad \text{for} \quad \|V\| < d\,.$$

Thus, for subordinated or annular separated spectral components the condition $\|V\| < d$ ensures that (2.2) and (2.3) hold. However, it deserves further studies to determine whether the corresponding bounds on the maximal angle are sharp and, if this is not the case, to obtain sharp bounds.

Matters change, again, in the generic case. Similar to the considerations in the preceding section, here it is yet unclear whether the condition $\|V\| < d$ is sufficient to ensure (2.3). Nevertheless, some bounds on the maximal angle stronger than (2.29) can be obtained with the same techniques as for general perturbations. For example, one can use (2.25) with appropriately chosen spectral components ω_t and Ω_t satisfying $\operatorname{dist}(\omega_t,\Omega_t) \ge d - t\|V\|$ to infer that

$$\theta \le \frac{\pi}{2}\|V\| \int_0^1 \frac{\mathrm{d}t}{d-t\|V\|} = \frac{\pi}{2}\log\Bigl(\frac{d}{d-\|V\|}\Bigr) < \frac{\pi}{2}$$

for $\|V\| < (1-\mathrm{e}^{-1})d$. An even stronger bound can be obtained with a suitable finite partition of the interval $[0,1]$. In the same way as for (2.29) one can define an optimization problem for the choice of this partition, and this optimization problem also seems to be explicitly solvable with basically the same technique. However, the corresponding considerations seem to be even more technical and more extensive than the ones for (2.29), see the separate discussion in Section 8.4 below. The explicit computation for this problem is therefore omitted and is left for future studies.

Chapter 3

Operator Sylvester equations and the $\sin\Theta$ theorem

Because of their great importance in various fields of mathematics, operator Sylvester equations have been studied extensively over the decades. In the context of this thesis, we point out the works [5, 6, 9, 11, 46]; see also the survey article [12] and the references therein.

In the present chapter, we collect the material on operator Sylvester equations that is needed throughout this thesis. Of particular importance here is the existence and uniqueness result in Theorem 3.2 below. A block variant of this result is formulated in Corollary 3.5, which proves useful, for instance, in Chapter 6. As the main application of these considerations, we have a variant of the Davis-Kahan symmetric $\sin\Theta$ theorem, see Section 3.2 below. Finally, in Section 3.3 we discuss a condition in terms of Sylvester equations which guarantees that a bounded operator on a Hilbert space is also bounded with respect to the graph norm topology of a given closed operator.

Most of the material presented in this chapter is essentially well known. However, to the author's best knowledge the results obtained in Section 3.3 below are new.

3.1 Strong solutions to Sylvester equations

We start with recalling the well-known concept of strong solutions to operator Sylvester equations. Here, we mainly follow [6, Section 4].

Definition 3.1. Let A_0 and A_1 be closed densely defined operators on Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. A bounded operator $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ is called a *strong solution to the operator Sylvester equation*

$$XA_0 - A_1X = K\,, \quad K \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)\,, \tag{3.1}$$

if

$$\operatorname{Ran}\bigl(X|_{\operatorname{Dom}(A_0)}\bigr) \subset \operatorname{Dom}(A_1)$$

and

$$XA_0g - A_1Xg = Kg \quad \text{for} \quad g \in \operatorname{Dom}(A_0)\,. \tag{3.2}$$

Operator Sylvester equations are closely related to operator Riccati equations discussed in Section 1.4. Indeed, the Sylvester equation (3.1) corresponds to the Riccati equation (1.6) with $D = 0$ and $E = K$. Conversely, the Riccati equation (1.6) can be rewritten as

$$XA_0 - A_1X = E - XDX \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)\,,$$

which can be interpreted as an implicit Sylvester equation in the sense that the right-hand side depends on the operator unknown.

In this regard, as a particular case of Lemma 1.5, we have that an operator $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ is a strong solution to the Sylvester equation (3.1) if and only if $Y = -X^*$ is a strong solution to the dual equation

$$YA_1^* - A_0^*Y = K^*\,, \tag{3.3}$$

see also [6, Lemma 4.3].

Existence of strong solutions

There are a number of existence results for strong solutions to Sylvester equations. In the case where the operators A_0 and A_1 are both assumed to be self-adjoint, a necessary and sufficient condition for the existence of a strong solution to (3.1) can be found in [47]. Cases of more general operators A_0 and A_1 in the context of strongly continuous semigroups have been discussed in [42].

Unlike these conditions in [42] and [47], more applicable sufficient conditions for the existence of strong solutions usually require that the spectra of the operators A_0 and A_1 are separated. Although this alone is not sufficient if the operators A_0 and A_1 are just closed and both are unbounded (see [42] for a counterexample), the separation of the spectra guarantees the existence of a strong solution under certain additional assumptions on the (unbounded) operators A_0 and A_1, see, e.g., [5, Lemma 2.18] and [6, Lemma 4.8].

In the particular case where both operators are assumed to be self-adjoint, we have the following well-known result essentially due to Bhatia, Davis, and McIntosh [11]. Basically all results mentioned in Section 2.3, in particular the main results obtained in Chapters 6–8 below, can be traced back to this theorem. It therefore plays a crucial role in our considerations.

Theorem 3.2. *Let A_0 and A_1 be two self-adjoint operators on Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, such that*

$$d := \operatorname{dist}\bigl(\operatorname{spec}(A_0), \operatorname{spec}(A_1)\bigr) > 0\,. \tag{3.4}$$

Then, the operator Sylvester equation (3.1) *has a unique strong solution X in $\mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$. This solution admits the representation*

$$X = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tA_1} K \mathrm{e}^{-\mathrm{i}tA_0} f_d(t)\,\mathrm{d}t\,, \tag{3.5}$$

where the integral is understood in the weak sense[1] *and f_d is any function in $L^1(\mathbb{R})$, continuous except at zero, such that*

$$\hat{f}_d(\lambda) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}t\lambda} f_d(t)\,\mathrm{d}t = \frac{1}{\lambda} \quad \textit{whenever} \quad |\lambda| \geq d\,.$$

In particular, X satisfies the norm bound

$$\|X\| \leq c\,\frac{\|K\|}{d}\,, \tag{3.6}$$

where

$$c = \inf\Bigl\{ \|f\|_{L^1(\mathbb{R})} \Bigm| f \in L^1(\mathbb{R})\,,\ \hat{f}(\lambda) = \frac{1}{\lambda}\,,\ |\lambda| \geq 1 \Bigr\} = \frac{\pi}{2}\,. \tag{3.7}$$

[1] This means that X is given by $\langle h, Xg\rangle = \int_{\mathbb{R}} \langle h, \mathrm{e}^{\mathrm{i}tA_1} K \mathrm{e}^{-\mathrm{i}tA_0} g\rangle f_d(t)\,\mathrm{d}t$ for $g \in \mathcal{H}_0$ and $h \in \mathcal{H}_1$.

Proof. This is obtained by combining [5, Theorem 2.7] and [6, Lemma 4.2]; cf. also [5, Remark 2.8], [8, Theorem 3.2], and [11, Theorem 4.1]. Note that the last equality in (3.7) goes back to Sz.-Nagy and Strausz [52], [53]. □

The constant $c = \pi/2$ in estimate (3.6) is optimal. This was shown by McEachin [33] by means of suitable finite-dimensional examples. These examples can be directly extended to the infinite-dimensional case, which leads to the following (stronger) sharpness result.

Remark 3.3. The estimate given by (3.6) and (3.7) is sharp in the sense that equality can be attained:

Let $\ell^2 = \ell^2(\mathbb{N})$ denote the Hilbert space of square-summable sequences, and let $\{e_k\}_{k\in\mathbb{N}} \subset \ell^2$ be the standard orthonormal basis of ℓ^2. On ℓ^2 define the (unbounded) self-adjoint operators A_0 and A_1 by

$$A_0 e_k := 2k e_k \quad \text{and} \quad A_1 e_k := (2k-1)e_k\,, \quad k \in \mathbb{N}\,,$$

with

$$\operatorname{Dom}(A_0) := \operatorname{Dom}(A_1) := \Big\{(x_k) \in \ell^2 \;\Big|\; \sum_{k=1}^{\infty} k^2 |x_k|^2 < \infty\Big\}\,.$$

Clearly, the spectra of A_0 and A_1 consist of the even and odd numbers in $\mathbb{N}$, respectively, so that

$$d := \operatorname{dist}\big(\operatorname{spec}(A_0), \operatorname{spec}(A_1)\big) = 1\,. \tag{3.8}$$

Define operators K and X from the finite sequences in ℓ^2 to ℓ^2 by

$$Ke_k := \sum_{j\in\mathbb{N}} \frac{e_j}{2(k-j)+1} \quad \text{and} \quad Xe_k := \sum_{j\in\mathbb{N}} \frac{e_j}{(2(k-j)+1)^2}$$

for $k \in \mathbb{N}$. One easily verifies that

$$\langle e_j, XA_0 e_k\rangle - \langle A_1 e_j, Xe_k\rangle = \langle e_j, Ke_k\rangle \quad \text{for} \quad j,k \in \mathbb{N}\,. \tag{3.9}$$

One can show that both K and X extend to bounded operators on ℓ^2, which we again denote by K and X, respectively. In fact, it follows from the considerations in the proofs of [33, Proposition 2 and 3] that

$$\|K\| = \frac{\pi}{2} \quad \text{and} \quad \|X\| = \frac{\pi^2}{4}\,. \tag{3.10}$$

Hence, the identity (3.9) extends to

$$\langle h, XA_0g\rangle - \langle A_1h, Xg\rangle = \langle h, Kg\rangle \quad \text{for} \quad g, h \in \operatorname{Dom}(A_0) = \operatorname{Dom}(A_1)\,,$$

that is, X is a so-called *weak solution* to the Sylvester equation (3.1), see, e.g., [6, Definition 4.1]. Applying [6, Lemma 4.2] yields that X is also a strong solution to (3.1), which, in view of (3.8) and (3.10), shows that the estimate given by (3.6) and (3.7) is sharp.

The finite-dimensional examples originally discussed by McEachin in [33] can be recovered from this example by truncating it to the finite-dimensional case.

Although the estimate given by (3.6) and (3.7) is sharp by Remark 3.3, in some cases a better constant than $\pi/2$ is available. If, for example, the convex hull of one of the sets $\operatorname{spec}(A_0)$ and $\operatorname{spec}(A_1)$ is additionally assumed to be disjoint from the other set, that is, $\operatorname{conv}\bigl(\operatorname{spec}(A_0)\bigr) \cap \operatorname{spec}(A_1) = \varnothing$ or vice versa, then the constant $c = \pi/2$ in the bound (3.6) can be replaced by 1, that is, in this case one has

$$\|X\| \le \frac{\|K\|}{d}\,. \tag{3.11}$$

This follows from corresponding alternative representation formulae for the solution, see [11, Theorem 3.3] and [12, Theorem 9.1]; cf. also [11, Theorem 3.1]. Other improvements on the constant may be available for small dimensions of the underlying Hilbert spaces, see [35].

Note that the estimate (3.11) is sharp even in the case of bounded operators A_0 and A_1, which can be seen from the case where both $\mathcal{H}_0$ and $\mathcal{H}_1$ are one-dimensional.

The bound (3.11) has originally been obtained directly for normal operators A_0 and A_1 under suitable assumptions on the disposition of their spectra. With slight modifications, a variant of Theorem 3.2 for normal operators is also valid:

Remark 3.4. A statement analogous to Theorem 3.2 holds if the operators A_0 and A_1 are assumed to be just normal and their spectra are separated as in (3.4). In this case, the solution to (3.1) admits a representation similar to (3.5), and the constant c in (3.6) has to be replaced by some constant less than 2.91, see [10] and [11, Theorem 4.2]. Note that the exact value of the optimal constant here is still unknown.

The following corollary to Theorem 3.2 is a very useful block variant of that result. It is extracted from a more specialized version in the proof of Proposition 4.1 in the author's article [50].

Corollary 3.5. *Let $A = A_0 \oplus A_1$ and $B = B_0 \oplus B_1$ be self-adjoint operators on Hilbert spaces $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ and $\mathcal{K} = \mathcal{K}_0 \oplus \mathcal{K}_1$, respectively. Suppose that there is $d > 0$ such that*

$$\operatorname{dist}\bigl(\operatorname{spec}(A_0), \operatorname{spec}(B_1)\bigr) \geq d \quad \text{and} \quad \operatorname{dist}\bigl(\operatorname{spec}(A_1), \operatorname{spec}(B_0)\bigr) \geq d\,,$$

and let $K \in \mathcal{L}(\mathcal{K}, \mathcal{H})$ have the off-diagonal representation

$$K = \begin{pmatrix} 0 & K_1 \\ K_0 & 0 \end{pmatrix}, \quad K_0 \in \mathcal{L}(\mathcal{K}_0, \mathcal{H}_1)\,, \quad K_1 \in \mathcal{L}(\mathcal{K}_1, \mathcal{H}_0)\,,$$

as an operator from $\mathcal{K} = \mathcal{K}_0 \oplus \mathcal{K}_1$ to $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$.

Then, the Sylvester equation

$$XB - AX = K \tag{3.12}$$

has exactly one strong solution $X \in \mathcal{L}(\mathcal{K}, \mathcal{H})$ of the form

$$X = \begin{pmatrix} 0 & X_1 \\ X_0 & 0 \end{pmatrix}, \quad X_0 \in \mathcal{L}(\mathcal{K}_0, \mathcal{H}_1)\,, \quad X_1 \in \mathcal{L}(\mathcal{K}_1, \mathcal{H}_0)\,.$$

This solution admits the representation

$$X = \int_{\mathbb{R}} e^{i\tau A} K e^{-i\tau B} f_d(\tau)\, d\tau\,, \tag{3.13}$$

where the integral is understood in the weak sense and $f_d \in L^1(\mathbb{R})$ is any function as in Theorem 3.2.

Proof. One easily verifies that the Sylvester equation (3.12) splits into the pair of Sylvester equations

$$X_0 B_0 - A_1 X_0 = K_0 \tag{3.14}$$

and

$$X_1 B_1 - A_0 X_1 = K_1\,. \tag{3.15}$$

Since, by hypothesis, the spectra of the parts A_1 and B_0, respectively A_0 and B_1, are separated with distance at least d, the claim now follows by applying Theorem 3.2 to each of these two Sylvester equations. In particular, taking into account that

$$\mathrm{e}^{\mathrm{i}\tau A} K \mathrm{e}^{-\mathrm{i}\tau B} = \begin{pmatrix} 0 & \mathrm{e}^{\mathrm{i}\tau A_0}\, K_1\, \mathrm{e}^{-\mathrm{i}\tau B_1} \\ \mathrm{e}^{\mathrm{i}\tau A_1}\, K_0\, \mathrm{e}^{-\mathrm{i}\tau B_0} & 0 \end{pmatrix},$$

representation (3.13) is obtained by combining the corresponding representations for X_0 and X_1. □

It should be emphasized that, in contrast to Theorem 3.2, the spectra of the operators A and B in Corollary 3.5 are not assumed to be separated. Moreover, the spectra of A_0 and A_1, respectively B_0 and B_1, are allowed to overlap. However, if the spectra of A_0 and A_1 are separated, then also the case $\mathcal{K} = \mathcal{H}$ and $B = A$ can be treated by Corollary 3.5.

In the situation of Corollary 3.5, representation (3.13) implies the estimate

$$\|X\| \le \frac{\pi}{2}\,\frac{\|K\|}{d}\,, \tag{3.16}$$

cf. Theorem 3.2. Taking into account the Sylvester equations (3.14) and (3.15) and that $\|X\| = \max\{\|X_0\|, \|X_1\|\}$ and $\|K\| = \max\{\|K_0\|, \|K_1\|\}$, this estimate also follows from the corresponding estimates for K_0 and K_1 obtained from Theorem 3.2. Nevertheless, representation (3.13) has its own right. It is used, for instance, in the proof of Lemma 6.7 in Chapter 6 below. It also allows to extend estimate (3.16) very easily to unitary invariant norms other than the usual operator norm, cf. [50, Section 4]. To the author's best knowledge representation (3.13) has not been stated anywhere before.

3.2 The symmetric $\sin\Theta$ theorem

In this section, we discuss a variant of the Davis-Kahan symmetric $\sin\Theta$ theorem from [21, Proposition 6.1]. The corresponding material is taken from the author's article [50].

We start with the following essentially well-known observation, see, e.g., [11, Section 2] and [34, Section 2]; see also the proof of [8, Proposition 3.4].

Lemma 3.6. *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$, and let $V \in \mathcal{L}(\mathcal{H})$ be self-adjoint. Moreover, suppose that P and Q are orthogonal projections onto reducing subspaces for A and $A+V$, respectively. Then*

$$X = P - Q = PQ^{\perp} - P^{\perp}Q \tag{3.17}$$

is a strong solution to the Sylvester equation

$$X(A+V) - AX = PVQ^{\perp} - P^{\perp}VQ\,. \tag{3.18}$$

Proof. Since $\operatorname{Ran} P$ is reducing for A, the projection P commutes with A, that is, one has $Px \in \operatorname{Dom}(A)$ and $PAx = APx$ for all $x \in \operatorname{Dom}(A)$, see Section 1.2. Analogously, $Q^{\perp}$ commutes with $A+V$. Hence, $PQ^{\perp}$ maps $\operatorname{Dom}(A) = \operatorname{Dom}(A+V)$ into $\operatorname{Dom}(A)$ and satisfies

$$PQ^{\perp}(A+V)x - APQ^{\perp}x = P(A+V)Q^{\perp}x - PAQ^{\perp}x = PVQ^{\perp}x \tag{3.19}$$

for $x \in \operatorname{Dom}(A)$. Analogously, $P^{\perp}Q$ maps $\operatorname{Dom}(A)$ into $\operatorname{Dom}(A)$ and satisfies

$$P^{\perp}Q(A+V)x - AP^{\perp}Qx = P^{\perp}VQx \quad \text{for} \quad x \in \operatorname{Dom}(A)\,. \tag{3.20}$$

Combining (3.19) and (3.20), one concludes that (3.17) is a strong solution to (3.18). □

Clearly, the operators $X = P - Q = PQ^{\perp} - P^{\perp}Q$ and $PVQ^{\perp} - P^{\perp}VQ$ in Lemma 3.6 can be represented as off-diagonal 2×2 block operator matrices from $\operatorname{Ran} Q \oplus \operatorname{Ran} Q^{\perp}$ to $\operatorname{Ran} P \oplus \operatorname{Ran} P^{\perp}$, so that Corollary 3.5 can be applied. Since

$$\|PVQ^{\perp} - P^{\perp}VQ\| \le \|V\|\,,$$

estimate (3.16) and the identity $|P-Q| = \sin\bigl(\Theta(P,Q)\bigr)$ (see equation (1.11)) then lead to the following variant of the Davis-Kahan symmetric $\sin\Theta$ theorem.

Proposition 3.7 (The symmetric $\sin\Theta$ theorem; see [50, Propositions 2.3 and 4.1])**.** *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$, let $V \in \mathcal{L}(\mathcal{H})$ be self-adjoint, and suppose that P and Q are orthogonal projections onto reducing subspaces for A and $A+V$, respectively. Let A_0 and A_1 denote the parts of A associated with* $\operatorname{Ran} P$ *and* $\operatorname{Ran} P^\perp$*, respectively, and let B_0 and B_1 likewise be the parts of $A+V$ associated with* $\operatorname{Ran} Q$ *and* $\operatorname{Ran} Q^\perp$*.*

Assume that there is $d > 0$ such that

$$\operatorname{dist}\bigl(\operatorname{spec}(A_0), \operatorname{spec}(B_1)\bigr) \geq d \quad \textit{and} \quad \operatorname{dist}\bigl(\operatorname{spec}(A_1), \operatorname{spec}(B_0)\bigr) \geq d\,.$$

Then, the operator angle $\Theta = \Theta(P,Q)$ associated with the subspaces $\operatorname{Ran} P$ *and* $\operatorname{Ran} Q$ *satisfies the bound*

$$\|\sin\Theta\| = \|P-Q\| \leq \frac{\pi}{2}\,\frac{\|V\|}{d}\,.$$

Note that in Proposition 3.7 information on the spectrum of both A and $A+V$ is required.

Remark 3.8. In the original version of the symmetric $\sin\Theta$ theorem [21, Proposition 6.1] by Davis and Kahan, it is additionally assumed that for each of the pairs $\bigl(\operatorname{spec}(A_0), \operatorname{spec}(B_1)\bigr)$ and $\bigl(\operatorname{spec}(A_1), \operatorname{spec}(B_0)\bigr)$ the convex hull of one of the sets is disjoint from the other set. As a consequence, instead of $\pi/2$, the constant 1 appears in the resulting estimate, cf. the discussion after Remark 3.3 above. Note that this original version of the symmetric $\sin\Theta$ theorem is formulated in [21] for arbitrary unitary-invariant norms. A corresponding extension of Proposition 3.7 is discussed in Proposition 4.1 of the author's article [50].

Although not stated in this explicit way, Proposition 3.7 is present in several recent works. For example, in the case where the operator A is assumed to be bounded, it is used to prove [26, Theorem 1] and [31, Theorem 1 (ii)]. In the unbounded setting, it appears, for instance, in the proof of [8, Theorem 3.5]. In each of these cases, the estimate on $\|P-Q\|$ is obtained by the identity

$$\|P-Q\| = \max\{\|PQ^\perp\|, \|P^\perp Q\|\}$$

and the corresponding estimates on $\|PQ^\perp\|$ and $\|P^\perp Q\|$ given by Theorem 3.2, cf. the discussion after Corollary 3.5.

3.3 Sylvester equations and graph norm topology

In this last section of the chapter, we consider the particular case of $\mathcal{H}_0 = \mathcal{H}_1$ and $A := A_0 = A_1$, and study solutions to the corresponding Sylvester equation in the graph norm topology of the operator A. Note that in the case in question the spectra of the coefficients coincide and a strong solution does therefore not exist in general. Nevertheless, relevant cases where a solution does exist arise in the context of Corollary 3.5.

To the author's best knowledge, the obtained results (Lemma 3.9 and Corollary 3.10 below) are new. They turn out to be quite useful, see Lemma 3.12 below and also the discussion at the end of Chapter 4.

Recall that for a closed operator A on a Hilbert space $\mathcal{H}$ its domain $\operatorname{Dom}(A)$ can be equipped with the inner product

$$\langle x,y\rangle_A := \langle x,y\rangle + \langle Ax,Ay\rangle\,,\quad x,y\in\operatorname{Dom}(A)\,, \tag{3.21}$$

which makes $(\operatorname{Dom}(A),\langle\,\cdot,\cdot\,\rangle_A)$ a Hilbert space. The corresponding norm is equivalent to the graph norm $|||\cdot|||_A$ on $\operatorname{Dom}(A)$ defined by

$$|||x|||_A := \|x\| + \|Ax\|\,,\quad x\in\operatorname{Dom}(A)\,.$$

It is a natural question when a bounded operator $Y\in\mathcal{L}(\mathcal{H})$ that maps $\operatorname{Dom}(A)$ into itself is also bounded with respect to the graph norm $|||\cdot|||_A$. The following lemma provides a sufficient condition for this property in terms of strong solutions to operator Sylvester equations.

Lemma 3.9. *Let A be a closed densely defined operator on a Hilbert space $\mathcal{H}$, and suppose that $Y\in\mathcal{L}(\mathcal{H})$ is a strong solution to the Sylvester equation*

$$YA - AY = K\,,\quad K\in\mathcal{L}(\mathcal{H})\,.$$

Then, Y is bounded on $\operatorname{Dom}(A)$ *with respect to the graph norm topology for A, and the corresponding spectral radius does not exceed $\|Y\|$.*

Proof. By hypothesis, one has $AYx = YAx - Kx$ for $x \in \operatorname{Dom}(A)$. Thus,

$$\|AYx\| \le \|Y\| \, \|Ax\| + \|K\| \, \|x\| \,, \quad x \in \operatorname{Dom}(A) \,,$$

and, therefore,

$$|\!|\!| Yx |\!|\!|_A = \|Yx\| + \|AYx\| \le \|Y\| \, |\!|\!| x |\!|\!|_A + \|K\| \, \|x\| \tag{3.22}$$

for $x \in \operatorname{Dom}(A)$. In particular, since $\|x\| \le |\!|\!| x |\!|\!|_A$ for $x \in \operatorname{Dom}(A)$, this yields that $Y|_{\operatorname{Dom}(A)}$ is bounded with respect to the graph norm $|\!|\!| \cdot |\!|\!|_A$, and the corresponding operator norm satisfies the estimate $|\!|\!| Y |\!|\!|_A \le \|Y\| + \|K\|$.

By induction, it easily follows from (3.22) that for all $n \in \mathbb{N}$ one has

$$|\!|\!| Y^n x |\!|\!|_A \le \|Y\|^n \, |\!|\!| x |\!|\!|_A + n\|K\| \, \|Y\|^{n-1} \|x\| \,, \quad x \in \operatorname{Dom}(A) \,.$$

This implies that $|\!|\!| Y^n |\!|\!|_A \le \|Y\|^n + n\|K\| \|Y\|^{n-1}$ for $n \in \mathbb{N}$. The corresponding spectral radius $r_A(Y) := \lim_{n\to\infty} |\!|\!| Y^n |\!|\!|_A^{1/n}$ can then be estimated as

$$r_A(Y) \le \lim_{n\to\infty} \bigl(\|Y\|^n + n\|K\| \, \|Y\|^{n-1} \bigr)^{1/n} = \|Y\| \,,$$

which completes the proof. □

Corollary 3.10. *Let A and Y as in Lemma 3.9. Then, for every scalar power series $h(z) = \sum_{n=0}^{\infty} a_n z^n$ with radius of convergence greater than $\|Y\|$, the operator $h(Y) = \sum_{n=0}^{\infty} a_n Y^n$ maps* Dom(A) *into itself. Here, the series is understood in the operator norm topology of $\mathcal{L}(\mathcal{H})$.*

Proof. Taking into account that the operator A is closed, it follows from Lemma 3.9 that the series $\sum_{n=0}^{\infty} a_n (Y|_{\operatorname{Dom}(A)})^n$ converges in the operator norm topology with respect to the graph norm $|\!|\!| \cdot |\!|\!|_A$. Moreover, since one has $\|x\| \le |\!|\!| x |\!|\!|_A$ for $x \in \operatorname{Dom}(A)$, the corresponding limit agrees with the restriction of $h(Y) \in \mathcal{L}(\mathcal{H})$ to $\operatorname{Dom}(A)$. This proves the claim. □

Remark 3.11. In the situation of Corollary 3.10, it is straightforward to verify that there is some $H \in \mathcal{L}(\mathcal{H})$ such that the operator $Z = h(Y)$ is a strong solution to the Sylvester equation

$$ZA - AZ = H \,.$$

We close this section with the following application of Lemma 3.9 and Corollary 3.10 in the context of $\mathcal{C}^1$-smooth paths of operators, cf. Section 1.6. The result is used in the forthcoming Chapter 6, see Lemma 6.11 below.

Lemma 3.12. *Let A be a closed densely defined operator on a Hilbert space $\mathcal{H}$, and let $Y \in \mathcal{L}(\mathcal{H})$ be a strong solution to the Sylvester equation*

$$YA - AY = K\,, \quad K \in \mathcal{L}(\mathcal{H})\,.$$

Then, for every interval $I \subset \mathbb{R}$, the path

$$I \ni t \mapsto \exp(tY)A\exp(-tY)$$

is $\mathcal{C}^1$-smooth on $\operatorname{Dom}(A)$ *with*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl(\exp(tY)A\exp(-tY)\bigr) = \exp(tY)K\exp(-tY)|_{\operatorname{Dom}(A)}\,.$$

Proof. For $t \in I$ define

$$B_t := \exp(tY)A\exp(-tY) \quad \text{on} \quad \operatorname{Dom}(B_t) := \operatorname{Ran}\bigl(\exp(tY)|_{\operatorname{Dom}(A)}\bigr)\,.$$

By Corollary 3.10, the operators $\exp(tY)$ and $\exp(tY)^{-1} = \exp(-tY)$ map $\operatorname{Dom}(A)$ into itself, so that $\exp(tY)$ maps $\operatorname{Dom}(A)$ onto itself. Hence, one has

$$\operatorname{Dom}(B_t) = \operatorname{Dom}(A)\,, \quad t \in I\,.$$

Moreover, the formal derivative of the path $t \mapsto B_t$ at $t \in I$ is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}B_t &= \exp(tY)\bigl(YA - AY\bigr)\exp(-tY)\\ &= \exp(tY)\,K\exp(-tY)|_{\operatorname{Dom}(A)}\,,\end{aligned} \tag{3.23}$$

cf. Example 1.15 (b). If A is bounded, this already completes the argument. However, if A is unbounded, then we have to justify that the path $I \ni t \mapsto B_t$ is indeed $\mathcal{C}^1$-smooth in norm with $\dot{B}_t = \exp(tY)K\exp(-tY)|_{\operatorname{Dom}(A)}$.

To do this, consider the Hilbert space $(\mathcal{H}_A, \langle\cdot,\cdot\rangle_A) := (\operatorname{Dom}(A), \langle\cdot,\cdot\rangle_A)$ with $\langle\cdot,\cdot\rangle_A$ as in (3.21). By Lemma 3.9, the restriction $Y|_{\operatorname{Dom}(A)}$ lies in $\mathcal{L}(\mathcal{H}_A)$, so that the path $I \ni t \mapsto \exp(-tY|_{\operatorname{Dom}(A)}) = \exp(-tY)|_{\operatorname{Dom}(A)}$ is $\mathcal{C}^1$-smooth in norm with respect to $\mathcal{H}_A$, see Example 1.15 (b). The corresponding derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp(-tY)|_{\mathrm{Dom}(A)} = -Y\exp(-tY)|_{\mathrm{Dom}(A)}\,, \quad t \in I\,.$$

With this, it is straightforward to verify that for arbitrary $x \in \mathrm{Dom}(A)$ and $y \in \mathcal{H}$ the scalar function

$$I \ni t \mapsto \langle y, B_t x\rangle$$

is $\mathcal{C}^1$-smooth with derivative $\frac{\mathrm{d}}{\mathrm{d}t}\langle y, B_t x\rangle = \langle y, K_t x\rangle$, where

$$K_t := \exp(tY)K\exp(-tY)\,,$$

cf. equation (3.23). Therefore,

$$\Big\langle y, \Big(\frac{B_t - B_s}{t-s} - K_s\Big)x\Big\rangle = \frac{1}{t-s}\int_s^t \langle y, (K_\tau - K_s)x\rangle\,\mathrm{d}\tau \quad \text{for} \quad s \neq t\,.$$

This implies that

$$\Big|\Big\langle y, \Big(\frac{B_t - B_s}{t-s} - K_s\Big)x\Big\rangle\Big| \le \|x\|\,\|y\| \sup_{\tau\in[s,t]}\|K_\tau - K_s\| \quad \text{for} \quad s < t, \tag{3.24}$$

and a similar inequality holds for $s > t$. Taking into account that the path $\tau \mapsto K_\tau$ is continuous, one concludes from (3.24) that the operator $(B_t - B_s)/(t-s)$ converges to K_s in norm as t approaches s. Hence, the path $\tau \mapsto B_t$ is $\mathcal{C}^1$-smooth in norm with $\dot{B}_t = K_t|_{\mathrm{Dom}(A)}$. □

Chapter 4

Reducing graph subspaces and block diagonalization

In this chapter, we address the problem of block diagonalization for possibly unbounded operator matrices in a Hilbert space $\mathcal{H}$ of the form

$$B = \begin{pmatrix} A_0 & W \\ W^* & A_1 \end{pmatrix} = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} + \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix} =: A + V$$

with respect to a given orthogonal decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. Here, the diagonal part A of B is a self-adjoint operator on

$$\operatorname{Dom}(A) = \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)\,,$$

the off-diagonal part V is assumed to be bounded on the whole Hilbert space $\mathcal{H}$, and B is understood as the operator sum

$$B = A + V \quad \text{on} \quad \operatorname{Dom}(B) = \operatorname{Dom}(A)\,.$$

In the particular case where one of the diagonal entries A_0 and A_1 is bounded, the block diagonalization of operators of this form has already been discussed by Adamjan, Langer, and Tretter in [4, Section 2]; cf. [54, Proposition 2.9.12]. Also the case where the spectra of the diagonal entries are additionally assumed to be subordinated has been studied in detail, see, e.g., [1, Theorem 2.3] and [32, Sections 3 and 4].

The situation described above without any additional assumptions on the self-adjoint diagonal entries A_0 and A_1 has been investigated by Albeverio, Makarov, and Motovilov in [5, Section 5]. In particular, it has been stated in [5, Lemma 5.3] that a graph subspace $\mathcal{G}(\mathcal{H}_0, X)$ with $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ is reducing for the operator $B = A + V$ if and only if the bounded skew-symmetric operator Y given by the 2×2 block operator matrix

$$Y = \begin{pmatrix} 0 & -X^* \\ X & 0 \end{pmatrix}$$

is a strong solution to the operator Riccati equation

$$YA - AY + YVY - V = 0\,. \tag{4.1}$$

In this case, it follows from [5, Theorem 5.5] that the operator $B = A + V$ admits the block diagonalization

$$(I_{\mathcal{H}}+Y)^{-1}(A+V)(I_{\mathcal{H}}+Y) = A+VY = \begin{pmatrix} A_0 + WX & 0 \\ 0 & A_1 - W^*X^* \end{pmatrix}. \tag{4.2}$$

In particular, one has

$$\operatorname{spec}(B_0) = \operatorname{spec}(A_0 + WX) \quad \text{and} \quad \operatorname{spec}(B_1) = \operatorname{spec}(A_1 - W^*X^*)\,,$$

where B_0 and B_1 denote the parts of B associated with $\mathcal{G}(\mathcal{H}_0, X)$ and $\mathcal{G}(\mathcal{H}_0, X)^{\perp}$, respectively. Moreover, the operator $A + V$ is unitarily equivalent to a block diagonal operator with respect to the orthogonal decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, namely

$$U^*(A + V)U = \begin{pmatrix} \Lambda_0 & 0 \\ 0 & \Lambda_1 \end{pmatrix},$$

where $U := (I_{\mathcal{H}} + Y)|I_{\mathcal{H}} + Y|^{-1}$ is the unitary operator from the polar decomposition for the isomorphism $I_{\mathcal{H}} + Y$ and the diagonal entries Λ_0 and Λ_1 are similar to $A_0 + WX$ and $A_1 - W^*X^*$, respectively, see [5, Theorem 5.5 (iii)].

However, there is a gap in reasoning in the proof of [5, Lemma 5.3]. More precisely, the inclusion $\operatorname{Ran}(Y|_{\operatorname{Dom}(A)}) \subset \operatorname{Dom}(A)$, which is required for Y to be a strong solution to (4.1), has been taken for granted. At the same time, the domain splitting

$$\operatorname{Dom}(A) = \big(\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_0, X)\big) + \big(\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_0, X)^{\perp}\big)\,,$$

which is needed for $\mathcal{G}(\mathcal{H}_0, X)$ to be not only invariant but also reducing for $A+V$, has not been discussed. Neither of these two properties is self-evident (cf. Section 1.2), and discussions with K. A. Makarov confirmed that they indeed need a careful justification.

The present author has found a way to close this gap in reasoning, and the corresponding argument is given in this chapter. For completeness of the presentation, the whole statement including the block diagonalization (4.2) is shown, see Theorem 4.7 below. In fact, the proof presented here does not allow to establish [5, Lemma 5.3] in its whole extent without discussing this block diagonalization, see Remark 4.8 below.

Corollary 4.9 provides a reformulation of [5, Lemma 5.3]. Here, the Riccati equation (4.1) for Y is replaced by a Riccati equation for the operator X alone.

Finally, at the end of the chapter, we discuss alternative arguments for the part of [5, Lemma 5.3] that states that Y is a strong solution to (4.1) if $\mathcal{G}(\mathcal{H}_0, X)$ is reducing for $A + V$.

The presented proof of Theorem 4.7 can be modified to extend [5, Lemma 5.3] to unbounded off-diagonal parts V with a sufficiently small relative bound with respect to A. This has been done in the joint work [38] with K. A. Makarov and S. Schmitz; see also Chapter 4 of the Ph.D. thesis [48] by the latter author. The presentation of the current chapter is based to a large extent on this joint work but is adapted to the easier case of bounded off-diagonal parts V as discussed in [5]. The main difference to the approach in [38] is briefly discussed in Remark 4.4 below.

Closing the gap in reasoning

For the rest of this chapter we make the following assumptions.

Hypothesis 4.1. *Let $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ be a Hilbert space decomposed into the sum of two orthogonal subspaces $\mathcal{H}_0$ and $\mathcal{H}_1$. Let A be a self-adjoint operator* *on $\mathcal{H}$ given by the representation*

$$A = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}, \quad \operatorname{Dom}(A) = \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1),$$

with respect to the decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. Moreover, let $V \in \mathcal{L}(\mathcal{H})$ be given by the 2×2 block operator matrix

$$V := \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix}, \quad W \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_0).$$

In the situation of Hypothesis 4.1, the first step towards a block diagonalization for $A + V$ with respect to a reducing graph subspace is the consideration of an invariant graph subspace $\mathcal{G}(\mathcal{H}_0, X)$ such that its orthogonal complement $\mathcal{G}(\mathcal{H}_0, X)^\perp = \mathcal{G}(\mathcal{H}_1, -X^*)$ is also invariant. In the setting of unbounded operators A, this requires to consider the intersections of the invariant subspaces with the operator domain. It is therefore convenient to fix the following additional assumptions.

Hypothesis 4.2. *Assume Hypothesis 4.1. Let $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$, and define $Y \in \mathcal{L}(\mathcal{H})$ by*

$$Y := \begin{pmatrix} 0 & -X^* \\ X & 0 \end{pmatrix}.$$

Finally, set

$$\mathcal{D} := \{x \in \operatorname{Dom}(A) \mid Yx \in \operatorname{Dom}(A)\} = \mathcal{D}_0 \oplus \mathcal{D}_1,$$

where

$$\mathcal{D}_0 := \{g \in \operatorname{Dom}(A_0) \mid Xg \in \operatorname{Dom}(A_1)\}$$

and

$$\mathcal{D}_1 := \{h \in \operatorname{Dom}(A_1) \mid X^*h \in \operatorname{Dom}(A_0)\}.$$

By definition, the set $\mathcal{D}$ in Hypothesis 4.2 is the maximal linear subset of $\operatorname{Dom}(A)$ that Y maps into $\operatorname{Dom}(A)$.

We have the following invariance criterion for graph subspaces, which is extracted from the proofs of [5, Lemma 5.3 and Theorem 5.5]. The corresponding reasoning is taken over and repeated briefly.

Lemma 4.3. *Assume Hypotheses 4.1 and 4.2. The following are equivalent:*

(i) The graph subspaces $\mathcal{G}(\mathcal{H}_0, X)$ and $\mathcal{G}(\mathcal{H}_1, -X^*)$ are invariant for the operator $A + V$.

(ii) The operator Y satisfies the Riccati equation

$$YAx - AYx + YVYx - Vx = 0 \quad \textit{for} \quad x \in \mathcal{D}\,. \tag{4.3}$$

(iii) One has

$$(A + V)(I_{\mathcal{H}} + Y)x = (I_{\mathcal{H}} + Y)(A + VY)x \quad \textit{for} \quad x \in \mathcal{D}\,. \tag{4.4}$$

Proof. Observing that

$$\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_0, X) = \{g \oplus Xg \mid g \in \mathcal{D}_0\} \tag{4.5}$$

and that

$$(A + V)\begin{pmatrix} g \\ Xg \end{pmatrix} = \begin{pmatrix} A_0 g + WXg \\ W^* g + A_1 Xg \end{pmatrix} \quad \text{for} \quad g \in \mathcal{D}_0\,,$$

one concludes that the graph subspace $\mathcal{G}(\mathcal{H}_0, X)$ is invariant for $A + V$ if and only if the equation $(W^* + A_1 X)g = X(A_0 + WX)g$ holds for all $g \in \mathcal{D}_0$. This, in turn, can be rewritten as

$$XA_0 g - A_1 Xg + XWXg - W^* g = 0 \quad \text{for} \quad g \in \mathcal{D}_0\,. \tag{4.6}$$

In view of

$$\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_1, -X^*) = \{-X^* h \oplus h \mid h \in \mathcal{D}_1\}\,, \tag{4.7}$$

the analogous reasoning yields that the graph subspace $\mathcal{G}(\mathcal{H}_1, -X^*)$ is invariant for the operator $A + V$ if and only if X^* satisfies

$$X^* A_1 h - A_0 X^* h - X^* W^* X^* h + Wh = 0 \quad \text{for} \quad h \in \mathcal{D}_1\,. \tag{4.8}$$

Thus, both subspaces $\mathcal{G}(\mathcal{H}_0, X)$ and $\mathcal{G}(\mathcal{H}_1, -X^*)$ are invariant for the operator $A + V$ if and only if the pair of Riccati equations (4.6) and (4.8) hold. This pair of equations can be rewritten as the single Riccati equation (4.3), which proves the stated equivalence of (i) and (ii).

Finally, equation (4.4) is just a reformulation of the Riccati equation (4.3), so that (iii) is equivalent to (ii). □

Remark 4.4. It is easy to verify that each of the statements (i)–(iii) in Lemma 4.3 is equivalent to

$$(I_{\mathcal{H}} - Y)(A+V)x = (A - YV)(I_{\mathcal{H}} - Y)x \quad \text{for} \quad x \in \mathcal{D}\,. \tag{4.9}$$

This latter equation has been used in the joint work [38] with K. A. Makarov and S. Schmitz to extend [5, Lemma 5.3] to certain unbounded off-diagonal parts V that are relatively bounded with respect to A. In fact, if V is allowed to be unbounded with $\operatorname{Dom}(A) \subset \operatorname{Dom}(V)$, then the natural domain of the operator $A+VY$ depends on the choice of Y, whereas the operator $A-YV$ still has natural domain $\operatorname{Dom}(A)$. Equation (4.9) is therefore better to handle than (4.4) in this situation. This is discussed in detail in [38, Remark 2.4 and Section 2.6].

However, with V being assumed to be bounded here, we follow [5, Section 5] and consider equation (4.4) instead of (4.9). In view of the representations for the graph subspaces given by (1.3), this equation appears to be more natural than (4.9) when dealing with invariant graph subspaces.

In the original proof of [5, Lemma 5.3], it has implicitly been assumed that the set $\mathcal{D}$ coincides with $\operatorname{Dom}(A)$. Furthermore, the condition for $\mathcal{G}(\mathcal{H}_0, X)$ to be reducing for $A+V$ has implicitly been replaced by the condition for $\mathcal{G}(\mathcal{H}_0, X)$ and $\mathcal{G}(\mathcal{H}_0, X)^{\perp} = \mathcal{G}(\mathcal{H}_1, -X^*)$ to be invariant for $A+V$. In this sense, Lemma 4.3 sums up what has essentially been proved in [5, Lemma 5.3 and Theorem 5.5 (ii)].

Note that the operator $A+VY$ is block diagonal, namely

$$A + VY = \begin{pmatrix} A_0 + WX & 0 \\ 0 & A_1 - W^*X^* \end{pmatrix},$$

and recall that the operator $I_{\mathcal{H}} + Y$ has a bounded inverse, see Section 1.3. It is therefore natural to use the similarity transformation given by $I_{\mathcal{H}} + Y$ to block diagonalize the operator $A+V$, cf. equation (4.4). However, unless $I_{\mathcal{H}} + Y$ maps $\operatorname{Dom}(A)$ onto itself, the operator $(I_{\mathcal{H}} + Y)^{-1}(A+V)(I_{\mathcal{H}} + Y)$

does not agree with $A+VY$ with respect to its natural domain. As it turns out in the proof of Theorem 4.7 below, this desired property of $I_{\mathcal{H}}+Y$ is closely related to the implicit assumptions in the proof of [5, Lemma 5.3].

Our approach to close this gap in reasoning and to obtain the actual statement of [5, Lemma 5.3] is based on the following elementary observation.

Lemma 4.5 ([49, Lemma 1.3])**.** *Let $\mathcal{T}$ and $\mathcal{S}$ be linear operators such that $\mathcal{S}\subset\mathcal{T}$. If $\mathcal{S}$ is surjective and $\mathcal{T}$ is injective, then $\mathcal{S}=\mathcal{T}$.*

Proof. For the sake of completeness, we reproduce the proof from [49].

Let $y\in\operatorname{Dom}(\mathcal{T})$ be arbitrary. Since $\mathcal{S}\subset\mathcal{T}$ and $\mathcal{S}$ is surjective, one can choose $x\in\operatorname{Dom}(\mathcal{S})\subset\operatorname{Dom}(\mathcal{T})$ such that $\mathcal{T}y=\mathcal{S}x=\mathcal{T}x$. The injectivity of $\mathcal{T}$ now implies that $y=x\in\operatorname{Dom}(\mathcal{S})$. Thus, $\operatorname{Dom}(\mathcal{T})=\operatorname{Dom}(\mathcal{S})$ and, hence, $\mathcal{S}=\mathcal{T}$. □

The preceding lemma is used in the following more specialized form.

Corollary 4.6. *Let K and L be closed linear operators on Hilbert spaces $\mathcal{K}$ and $\mathcal{L}$, respectively, and let $S\colon\mathcal{K}\to\mathcal{L}$ be an isomorphism. Suppose that*

$$SK\subset LS\,.$$

If the resolvent sets of K and L are not disjoint, that is, $\rho(K)\cap\rho(L)\neq\emptyset$, then

$$SK=LS$$

holds as an operator equality.

Proof. By a standard shift argument, one may assume that $0\in\rho(K)\cap\rho(L)$. In this case, since S is an isomorphism, the operators SK and LS are both bijective. The operator equality $SK=LS$ then follows from Lemma 4.5. □

We are now ready to prove [5, Lemma 5.3]. In fact, we combine the statements of [5, Lemma 5.3] and [5, Theorem 5.5 (ii)] to obtain a result analogous to Lemma 4.3. This is not just for reasons of convenience but is also essential part of our way to prove [5, Lemma 5.3], see Remark 4.8 below.

Theorem 4.7. *Assume Hypotheses 4.1 and 4.2. The following are equivalent:*

(i) The graph subspace $\mathcal{G}(\mathcal{H}_0, X)$ is reducing for the operator $A + V$.

(ii) The operator Y is a strong solution to the operator Riccati equation

$$YA - AY + YVY - V = 0\,. \tag{4.10}$$

(iii) The operator $A + V$ admits the block diagonalization

$$(I_{\mathcal{H}} + Y)^{-1}(A + V)(I_{\mathcal{H}} + Y) = A + VY = \begin{pmatrix} A_0 + WX & 0 \\ 0 & A_1 - W^*X^* \end{pmatrix}.$$

Proof. First, observe that the resolvent sets of the two operators $A+V$ and $A + VY$ are not disjoint, that is, $\rho(A + V) \cap \rho(A + VY) \neq \varnothing$. Indeed, the spectrum of $A + V$ is real since $A + V$ is self-adjoint, and by Lemma 1.17 the spectrum of $A + VY$ is contained in the closed $\|VY\|$-neighbourhood of $\operatorname{spec}(A)$.

Denote

$$T := I_{\mathcal{H}} + Y = \begin{pmatrix} I_{\mathcal{H}_0} & -X^* \\ X & I_{\mathcal{H}_1} \end{pmatrix}.$$

This operator $T \in \mathcal{L}(\mathcal{H})$ has a bounded everywhere defined inverse (see Section 1.3), and one has

$$\bigl(\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_0, X)\bigr) + \bigl(\operatorname{Dom}(A) \cap \mathcal{G}(\mathcal{H}_1, -X^*)\bigr) = \operatorname{Ran}(T|_{\mathcal{D}})\,, \tag{4.11}$$

cf. equations (4.5) and (4.7).

Suppose that (i) holds. In particular, the graph subspaces $\mathcal{G}(\mathcal{H}_0, X)$ and $\mathcal{G}(\mathcal{H}_1, -X^*)$ are invariant for $A + V$. Thus, Lemma 4.3 implies that

$$(A + V)Tx = T(A + VY)x \quad \text{for} \quad x \in \mathcal{D}\,. \tag{4.12}$$

Furthermore, it follows from equation (4.11) that $\operatorname{Dom}(A) = \operatorname{Ran}(T|_{\mathcal{D}})$, that is, $\operatorname{Ran}(T^{-1}|_{\operatorname{Dom}(A)}) = \mathcal{D} \subset \operatorname{Dom}(A)$. Equation (4.12) can then be rewritten as $T^{-1}(A + V)y = (A + VY)T^{-1}y$ for $y \in \operatorname{Dom}(A)$, so that

$$T^{-1}(A + V) \subset (A + VY)T^{-1}\,.$$

Since $\rho(A+V)\cap\rho(A+VY)\neq\varnothing$ as stated above, Corollary 4.6 implies that the identity

$$T^{-1}(A+V)=(A+VY)T^{-1}\,.$$

holds as an operator equality. This yields (iii).

Now, suppose that (ii) holds. In particular, one has $\mathcal{D}=\operatorname{Dom}(A)$ in this case, so that $\operatorname{Ran}(T|_{\operatorname{Dom}(A)})\subset\operatorname{Dom}(A)$. Moreover, Lemma 4.3 implies that $(A+V)Tx=T(A+VY)x$ for $x\in\operatorname{Dom}(A)=\operatorname{Dom}(A+VY)$. Hence,

$$T(A+VY)\subset(A+V)T\,.$$

Again, it follows from Corollary 4.6 that the identity

$$T(A+VY)=(A+V)T$$

holds as an operator equality, which yields (iii).

Finally, suppose that (iii) holds, that is,

$$(A+V)T=T(A+VY)\,.$$

In this case, one has

$$\operatorname{Dom}(A)=\operatorname{Dom}\bigl(T(A+VY)\bigr)=\operatorname{Dom}\bigl((A+V)T\bigr)=\operatorname{Ran}\bigl(T^{-1}|_{\operatorname{Dom}(A)}\bigr)\,,$$

which is equivalent to $\operatorname{Dom}(A)=\operatorname{Ran}(T|_{\operatorname{Dom}(A)})$. Since $Y=T-I_{\mathcal{H}}$, this implies that $\mathcal{D}=\operatorname{Dom}(A)$. The statement (ii) then follows by Lemma 4.3. Moreover, taking into account (4.11), Lemma 4.3 also yields that (i) holds. This completes the proof. □

Remark 4.8. Not only is the equivalence of (i) and (ii) in Theorem 4.7 established by the equivalence to (iii), the presented proof also requires equation (4.4), which is extracted from [5, Theorem 5.5 (ii)]. In other words, our proof of [5, Lemma 5.3] requires some elements of [5, Theorem 5.5]. In this sense, the two results [5, Lemma 5.3] and [5, Theorem 5.5 (ii)] should not be considered as separate statements, which is why Theorem 4.7 has been formulated as a combination of them.

The following reformulation of [5, Lemma 5.3] is an immediate consequence of Theorem 4.7 and Lemma 1.5. It replaces the Riccati equation for the operator Y by an Riccati equation for X and, thus, does not involve the adjoint operator X^*.

Corollary 4.9. *Assume Hypothesis 4.1, and let $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$. Then, the graph subspace $\mathcal{G}(\mathcal{H}_0, X)$ is reducing for the operator $A + V$ if and only if X is a strong solution to the operator Riccati equation*

$$XA_0 - A_1X + XWX - W^* = 0\,. \tag{4.13}$$

Proof. Let $Y \in \mathcal{L}(\mathcal{H})$ be given as in Hypothesis 4.2. Then, one has the inclusion $\operatorname{Ran}(Y|_{\operatorname{Dom}(A)}) \subset \operatorname{Dom}(A)$ if and only if $\operatorname{Ran}(X|_{\operatorname{Dom}(A_0)}) \subset \operatorname{Dom}(A_1)$ and $\operatorname{Ran}(X^*|_{\operatorname{Dom}(A_1)}) \subset \operatorname{Dom}(A_0)$, and it is easy to verify that the Riccati equation (4.10) for Y splits into the Riccati equation (4.13) for X and its dual equation

$$ZA_1 - A_0Z + ZW^*Z - W = 0$$

for $Z = -X^*$. Note that A_0 and A_1 are both self-adjoint. It now follows from Lemma 1.5 that Y is a strong solution to (4.10) if and only if X is a strong solution to (4.13). Applying Theorem 4.7 then proves the claim. □

Alternative arguments

The technique used above to prove Theorem 4.7 appears to be a reasonable way to establish the equivalence stated in [5, Lemma 5.3]. Yet, if one is interested only in the statement that Y is a strong solution to (4.10) if $\mathcal{G}(\mathcal{H}_0, X)$ is reducing for $A+V$, that is, the implication (i)⇒(ii) in Theorem 4.7, then one can also use other methods that are not based on Lemma 4.5. For the converse implication (ii)⇒(i), however, the author is not aware of any alternative way of reasoning.

The main issue in the proof of the implication (i)⇒(ii) is to show that $\mathcal{D} = \operatorname{Dom}(A)$. A direct way to prove this, that is, one that does not consider the block diagonalization (4.10), can be extracted from the proof of [32, Theorem 4.1]; cf. also [54, Theorem 2.7.21 (iii)]. In fact, the argument there works without any essential modifications in the current situation as well. It even allows to extend the implication (i)⇒(ii) to the same relatively bounded off-diagonal parts V discussed in [38, Section 4]. A similar reasoning can also be found in the proof of [58, Proposition 7.5].

We now present another alternative way to prove (i)⇒(ii). This one replaces the use of Corollary 4.6 by an argument based on Corollary 3.10. This proof, however, seems to work only for bounded perturbations V.

Alternative proof of the implication (i)⇒(ii) in Theorem 4.7.

Suppose that $\mathcal{G}(\mathcal{H}_0, X)$ is reducing for $A+V$. In this case, as in the proof of Theorem 4.7, one has

$$T^{-1}(A+V) \subset (A+VY)T^{-1} \quad \text{with} \quad T := I_{\mathcal{H}} + Y\,.$$

We show that the operator

$$Z := I_{\mathcal{H}} - T^{-1} \in \mathcal{L}(\mathcal{H})$$

is a strong solution to the operator Sylvester equation

$$Z(A+V) - (A+V)Z = VT^*T^{-1} \in \mathcal{L}(\mathcal{H})\,.$$

Indeed, one has

$$\operatorname{Ran}(Z|_{\operatorname{Dom}(A)}) \subset \operatorname{Dom}(A)$$

since $\operatorname{Ran}\bigl(T^{-1}|_{\operatorname{Dom}(A)}\bigr) \subset \operatorname{Dom}(A)$, and for $x \in \operatorname{Dom}(A)$ one computes

$$\begin{aligned} Z(A+V)x - (A+V)Zx &= (A+V)T^{-1}x - T^{-1}(A+V)x \\ &= (A+V)T^{-1}x - (A+VY)T^{-1}x \\ &= V(I_{\mathcal{H}} - Y)T^{-1}x = VT^*T^{-1}x\,. \end{aligned}$$

Clearly, one has $Z = I_{\mathcal{H}} - T^{-1} = T^{-1}(T - I_{\mathcal{H}}) = (I_{\mathcal{H}} + Y)^{-1}Y$. Since Y is skew-symmetric, in particular normal, by spectral mapping theorem this implies that $\|Z\| < 1$. It then follows from Corollary 3.10 that the operator

$$\sum_{n=0}^{\infty} Z^n = (I_{\mathcal{H}} - Z)^{-1} = T$$

maps $\operatorname{Dom}(A) = \operatorname{Dom}(A+V)$ into itself. Thus, $Y = T - I_{\mathcal{H}}$ maps $\operatorname{Dom}(A)$ into itself, so that $\mathcal{D} = \operatorname{Dom}(A)$. By Lemma 4.3, one concludes that Y is a strong solution to (4.10). □

Chapter 5

The angular metric on the set of orthogonal projections

In this chapter, we show that the maximal angle between closed subspaces of a Hilbert space indeed defines a metric on the set of orthogonal projections, the so-called *angular metric*, cf. Definition 1.7 in Chapter 1 and the discussion thereafter. An elementary proof for the corresponding triangle inequality has already been provided by Brown [16]. Here, we use a different technique that relies on tools from geometric perturbation theory. Although the corresponding material can essentially also be found in the joint work [36] with K. A. Makarov, the reasoning here is substantially simpler than the one in [36].

Let $t \mapsto P_t$ be a piecewise $\mathcal{C}^1$-smooth path of orthogonal projections. The main object of our studies in the present chapter is the following inequality, which we refer to as the *arcsine law*:

$$\arcsin\bigl(\|P_t - P_s\|\bigr) \le \int_s^t \|\dot{P}_\tau\| \,\mathrm{d}\tau \quad \text{whenever} \quad s \le t\,. \tag{5.1}$$

Note that (5.1) is stronger than the standard estimate from Lemma 1.16 since $x \le \arcsin(x)$ for $0 \le x \le 1$. Moreover, inequality (5.1) is sharp in the sense that equality can be attained, see [16, Proposition 5]; see also [36, Lemma 3.5] and Lemma 5.5 below.

The arcsine law is closely related to the triangle inequality for the maximal angle. Indeed, if this triangle inequality is taken for granted, then inequality (5.1) can easily be shown by using the standard estimate from

Lemma 1.16 in combination with partitions of the interval $[s,t]$ with arbitrarily small mesh size, see Remark 5.9 below.

However, in this chapter we proceed the other way around: An alternative proof for inequality (5.1) is presented that uses perturbation results for graph subspaces and that does not rely on the triangle inequality for the maximal angle, see Proposition 5.4 below. In turn, this inequality and the corresponding sharpness result are used to show that the maximal angle indeed defines a metric on the set of orthogonal projections, see Proposition 5.8 below.

5.1 Variation of graph subspaces

In this section we study how angular operators (cf. Section 1.3) vary with their corresponding graphs. Although these considerations aim for the proof of the arcsine law in Section 5.2, they may also be of interest on their own, see, e.g., Remark 5.2 below.

We begin with the following purely algebraic observation.

Lemma 5.1. *Let P, Q_1, and Q_2 be orthogonal projections in a Hilbert space $\mathcal{H}$ such that*

$$\operatorname{Ran} Q_j = \mathcal{G}(\operatorname{Ran} P, X_j)\,, \quad j = 1, 2\,,$$

for some $X_j \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$. Then

$$X_2 - X_1 = P^\perp T_1^*(Q_2 - Q_1)T_2 P|_{\operatorname{Ran} P}\,, \tag{5.2}$$

where $T_j \in \mathcal{L}(\mathcal{H})$ is given with respect to $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$ by

$$T_j = \begin{pmatrix} I_{\operatorname{Ran} P} & -X_j^* \\ X_j & I_{\operatorname{Ran} P^\perp} \end{pmatrix}, \quad j = 1, 2\,, \tag{5.3}$$

and the right-hand side of (5.2) *is understood as an operator from* $\operatorname{Ran} P$ *to* $\operatorname{Ran} P^\perp$*.*

Proof. Clearly, the identity $Q_1 T_1 = T_1 P$ holds (cf. Section 1.3), so that $T_1^* Q_1 = P T_1^*$. Hence, taking into account that $\operatorname{Ran}(T_2 P) = \operatorname{Ran} Q_2$, one concludes that

$$P^\perp T_1^*(Q_2 - Q_1)T_2 P = P^\perp T_1^* Q_2 T_2 P = P^\perp T_1^* T_2 P\,.$$

Since

$$T_1^*T_2 = \begin{pmatrix} I_{\operatorname{Ran} P} + X_1^*X_2 & X_1^* - X_2^* \\ X_2 - X_1 & I_{\operatorname{Ran} P^\perp} + X_1X_2^* \end{pmatrix}, \tag{5.4}$$

this proves the claim. □

Remark 5.2. In the situation of Lemma 5.1, recall that T_1 has a bounded inverse and, thus, a polar decomposition $T_1 = U_1|T_1|$ with a unitary operator $U_1 \in \mathcal{L}(\mathcal{H})$. Define the orthogonal projection $Q := U_1^*Q_2U_1$. If, in addition, $\operatorname{Ran} Q = \mathcal{G}(\operatorname{Ran} P, Z)$ for some $Z \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$, then one has

$$X_2 - X_1 = \bigl(I_{\mathcal{H}_1} + X_1X_1^*\bigr)^{1/2} Z \bigl(I_{\mathcal{H}_0} + X_1^*X_1\bigr)^{-1/2} \cdot \bigl(I_{\mathcal{H}_0} + X_1^*X_2\bigr) \tag{5.5}$$

with $\mathcal{H}_0 := \operatorname{Ran} P$ and $\mathcal{H}_1 := \operatorname{Ran} P^\perp$, see the following paragraph. In view of equation (1.18), this identity resembles the tangent angle addition formula

$$\tan\theta_2 - \tan\theta_1 = \tan(\theta_2 - \theta_1) \cdot (1 + \tan\theta_1 \tan\theta_2)\,.$$

In this sense, (5.5) can be interpreted as a non-commutative variant of this trigonometric addition formula, cf. [36, Remark 2.3].

The identity (5.5) was proved in [36, Corollary 2.2] under the additional assumption that the operator $I_{\mathcal{H}_0} + X_2^*X_1$ has full range. The preceding Lemma 5.1 allows a reasoning that does not require this additional assumption and that is simpler than the one in [36]. Indeed, identifying Z with its trivial continuation to an operator on the whole Hilbert space $\mathcal{H}$, the relation $\operatorname{Ran} Q = \mathcal{G}(\mathcal{H}_0, Z)$ means that

$$P^\perp Q = ZPQ\,.$$

Moreover, the projections P and $P^\perp$ commute with the operator $|T_1|$ since the latter is block diagonal with respect to $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$. Hence,

$$\begin{aligned} P^\perp T_1^*Q_2 = |T_1|P^\perp U_1^*Q_2 = |T_1|P^\perp QU_1^* = |T_1|ZPQU_1^* = |T_1|ZPU_1^*Q_2 \\ = |T_1|Z|T_1|^{-1}PT_1^*Q_2\,. \end{aligned}$$

Since $Q_2T_2P = T_2P$, one arrives at $P^\perp T_1^*T_2P = |T_1|Z|T_1|^{-1}PT_1^*T_2P$, which, in view of (5.4), agrees with (5.5).

An immediate corollary to Lemma 5.1 is the following result. It combines the statements of Lemmas 3.2 and 3.3 in [36]. Here, the use of the more general identity (5.2) instead of (5.5) makes the corresponding proof shorter and more transparent.

Corollary 5.3. *Let P be an orthogonal projection in a Hilbert space $\mathcal{H}$, and let $I \subset \mathbb{R}$ be an arbitrary (bounded or unbounded) interval. Furthermore, let P_t, $t \in I$, be an orthogonal projection in $\mathcal{H}$ such that*

$$\operatorname{Ran} P_t = \mathcal{G}(\operatorname{Ran} P, X_t)\,, \quad t \in I\,,$$

for some $X_t \in \mathcal{L}(\operatorname{Ran} P, \operatorname{Ran} P^\perp)$.

(a) If $I \ni t \mapsto P_t$ is continuous in norm, then so is $I \ni t \mapsto X_t$.

(b) If $I \ni t \mapsto P_t$ is a $\mathcal{C}^1$-smooth path, then so is $I \ni t \mapsto X_t$. In this case, one has

$$\|\dot{X}_t\| \le \bigl(1 + \|X_t\|^2\bigr)\|\dot{P}_t\|\,, \quad t \in I\,.$$

In particular, the path $I \ni t \mapsto X_t$ is piecewise $\mathcal{C}^1$-smooth if $I \ni t \mapsto P_t$ is.

Proof. Let $s \in I$ be arbitrary. By Lemma 5.1, the identity

$$X_t - X_s = P^\perp T_s^*(P_t - P_s)T_t P|_{\operatorname{Ran} P} \tag{5.6}$$

holds for all $t \in I$, where T_s and T_t are defined analogous to (5.3). Moreover, one has

$$\|T_t\| = \|\,|T_t|\,\| = \bigl(1 + \|X_t\|^2\bigr)^{1/2}\,, \quad t \in I\,.$$

Suppose that $I \ni t \mapsto P_t$ is continuous in norm. Since $\|P_t - P\| < 1$ for $t \in I$ and

$$\|X_t\| = \frac{\|P_t - P\|}{\sqrt{1 - \|P_t - P\|^2}}\,,$$

see Proposition 1.13, one concludes that T_t is uniformly bounded for t in a (compact) neighbourhood of s. It then follows from (5.6) that X_t converges to X_s in norm as t approaches s. This proves claim (a).

Now suppose that $I \ni t \mapsto P_t$ is $\mathcal{C}^1$-smooth in norm. In this case, it follows from part (a) that $I \ni t \mapsto T_t$ is continuous. Equation (5.6) then implies that

$$\dot{X}_s = P^\perp T_s^* \dot{P}_s T_s P|_{\operatorname{Ran} P}\,,$$

so that $I \ni t \mapsto X_t$ is $\mathcal{C}^1$-smooth. In particular, one has

$$\|\dot{X}_s\| \le \|T_s\|^2 \|\dot{P}_s\| = \bigl(1 + \|X_s\|^2\bigr) \|\dot{P}_s\| \,.$$

This proves claim (b) and, hence, completes the proof. □

5.2 The arcsine law

We are now in position to prove the arcsine law (5.1) for piecewise $\mathcal{C}^1$-smooth paths of orthogonal projections without using the triangle inequality for the maximal angle.

Proposition 5.4 ([36, Lemma 3.4]; cf. [37, Theorem 1])**.** *Let $I \ni t \mapsto P_t$ be a piecewise $\mathcal{C}^1$-smooth path of orthogonal projections. Then*

$$\arcsin\bigl(\|P_t - P_s\|\bigr) \le \int_s^t \|\dot{P}_\tau\| \,\mathrm{d}\tau \qquad \textit{whenever} \qquad s \le t \,.$$

Proof. Fix arbitrary $s, t \in I$ with $s < t$. Clearly, we may assume that $\int_s^t \|\dot{P}_\tau\| \,\mathrm{d}\tau < \frac{\pi}{2}$. Set

$$\gamma := \sup\bigl\{ r \in [s,t] \mid \|P_\tau - P_s\| < 1 \;\text{ whenever }\; s \le \tau \le r \bigr\} \,.$$

We show that

$$\arcsin\bigl(\|P_r - P_s\|\bigr) \le \int_s^r \|\dot{P}_\tau\| \,\mathrm{d}\tau < \frac{\pi}{2} \qquad \text{whenever} \qquad s \le r \le \gamma \,. \tag{5.7}$$

By the continuity of the path $I \ni r \mapsto P_r$ and the definition of γ, we then deduce that $\gamma = t$, which proves the claim.

Since $I \ni r \mapsto P_r$ is continuous, one has $\gamma > s$. Moreover, the strict inequality $\|P_r - P_s\| < 1$ holds whenever $s \le r < \gamma$, so that the range of each P_r is the graph of a bounded operator $X_r \in \mathcal{L}(\operatorname{Ran} P_s, \operatorname{Ran} P_s^\perp)$, that is,

$$\operatorname{Ran} P_r = \mathcal{G}(\operatorname{Ran} P_s, X_r) \qquad \text{for} \qquad s \le r < \gamma \,,$$

see Proposition 1.13. Define the piecewise $\mathcal{C}^1$-smooth function $F\colon [s,\gamma) \to \mathbb{R}$ by

$$F(r) := \int_s^r \bigl(1 + \|X_\tau\|^2\bigr) \|\dot{P}_\tau\| \,\mathrm{d}\tau \,, \quad s \le r < \gamma \,.$$

Let $s = \tau_0 < \dots < \tau_{n+1} = t$ be a partition of the interval $[s,t]$ such that

$[\tau_j, \tau_{j+1}] \ni \tau \mapsto P_\tau$ is $\mathcal{C}^1$-smooth for each j. By Corollary 5.3 (b), also each path $[\tau_j, \tau_{j+1}] \ni \tau \mapsto X_\tau$ is $\mathcal{C}^1$-smooth and satisfies $\|\dot{X}_\tau\| \le \bigl(1+\|X_\tau\|^2\bigr)\|\dot{P}_\tau\|$. In view of Lemma 1.16, one obtains

$$\|X_r - X_{\tau_j}\| \le \int_{\tau_j}^{r} \|\dot{X}_\tau\| \,\mathrm{d}\tau \le \int_{\tau_j}^{r} \bigl(1 + \|X_\tau\|^2\bigr)\|\dot{P}_\tau\| \,\mathrm{d}\tau = F(r) - F(\tau_j)$$

for $r \in [\tau_j, \tau_{j+1}]$, $j = 0, \dots, n$. Since $X_{\tau_0} = X_s = 0$, iterating this estimate via the triangle inequality for the operator norm yields

$$\|X_r\| = \|X_r - X_{\tau_0}\| \le F(r) - F(\tau_0) = F(r) \quad \text{for} \quad s \le r < \gamma\,. \tag{5.8}$$

For $\tau \in (\tau_j, \tau_{j+1})$, this implies that

$$F'(\tau) = \bigl(1 + \|X_\tau\|^2\bigr)\|\dot{P}_\tau\| \le \bigl(1 + F(\tau)^2\bigr)\|\dot{P}_\tau\|\,,$$

so that

$$\arctan F(r) - \arctan F(\tau_j) = \int_{\tau_j}^{r} \frac{F'(\tau)}{1 + F(\tau)^2} \,\mathrm{d}\tau \le \int_{\tau_j}^{r} \|\dot{P}_\tau\| \,\mathrm{d}\tau$$

for $r \in [\tau_j, \tau_{j+1}]$. Iterating this estimate and taking into account that $F(\tau_0) = 0$, one gets

$$\arctan F(r) \le \int_{s}^{r} \|\dot{P}_\tau\| \,\mathrm{d}\tau \quad \text{for} \quad s \le r < \gamma\,.$$

In view of $\arcsin\bigl(\|P_r - P_s\|\bigr) = \arctan\|X_r\| \le \arctan F(r)$ by (5.8), one concludes that

$$\arcsin\bigl(\|P_r - P_s\|\bigr) \le \int_{s}^{r} \|\dot{P}_\tau\| \,\mathrm{d}\tau \le \int_{s}^{t} \|\dot{P}_\tau\| \,\mathrm{d}\tau < \frac{\pi}{2} \quad \text{for} \quad s \le r < \gamma\,.$$

By continuity, this inequality also holds for $r = \gamma$, which proves (5.7) and, hence, completes the proof. □

The next lemma provides a non-trivial example for the fact that the estimate in Proposition 5.4 is sharp. It is essentially a reformulation of [16, Proposition 5] and [36, Lemma 3.5]. In its current formulation, this result will also be of interest in the forthcoming Chapter 6, see Lemma 6.11 below.

Lemma 5.5. *Let P be an orthogonal projection in a Hilbert space $\mathcal{H}$, and let $Y \in \mathcal{L}(\mathcal{H})$, $\|Y\| \le \pi/2$, be skew-symmetric and off-diagonal with respect to the orthogonal decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$, that is,*

$$Y^* = -Y \qquad \textit{and} \qquad PYP = 0 = P^\perp Y P^\perp\,.$$

Then, the path $[0,1] \ni t \mapsto P_t := \exp(tY)P\exp(-tY)$ of orthogonal projections in $\mathcal{H}$ is $\mathcal{C}^1$-smooth and satisfies

$$\arcsin\bigl(\|P_t - P_s\|\bigr) = \int_s^t \|\dot P_\tau\|\,\mathrm{d}\tau = \|Y\|\,(t-s) \qquad \textit{whenever} \qquad s \le t\,. \tag{5.9}$$

Proof. Clearly, the operator $\exp(tY)$ is unitary for every $t \in [0,1]$. Hence, each P_t is indeed an orthogonal projection in $\mathcal{H}$. Moreover, taking into account that Y satisfies $PY = YP^\perp$, the path $[0,1] \ni t \mapsto P_t$ is $\mathcal{C}^1$-smooth with

$$\dot P_t = \exp(tY)(YP - PY)\exp(-tY) = \exp(tY)Y(P - P^\perp)\exp(-tY)\,,$$

cf. Lemma 3.12. Since $\exp(tY)$ and $P - P^\perp$ are both unitary, one concludes that

$$\|\dot P_t\| = \|Y\| \qquad \text{for} \qquad t \in [0,1]\,. \tag{5.10}$$

This proves the second equality in (5.9).

Now fix s and t with $s < t$. We show that the operator angle associated with the subspaces $\operatorname{Ran} P_s$ and $\operatorname{Ran} P_t$ is given by $\Theta(P_s, P_t) = (t-s)|Y|$, so that

$$\arcsin\bigl(\|P_t - P_s\|\bigr) = \|Y\|\,(t-s)\,. \tag{5.11}$$

Combining (5.10) and (5.11) then proves the claim.

Clearly, the operator Y commutes with $\exp(sY)$ and $\exp(-sY)$. With this, it is straightforward to verify that Y is off-diagonal also with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^\perp$, that is,

$$P_s Y P_s = 0 = P_s^\perp Y P_s^\perp\,.$$

In view of Example 1.12, one arrives at the conclusion that the unitary operator

$$\exp(tY)\exp(-sY) = \exp\bigl((t-s)Y\bigr)$$

is a direct rotation from $\operatorname{Ran} P_s$ to $\operatorname{Ran} P_t$. In particular, the associated operator angle is given by $\Theta(P_s, P_t) = (t-s)|Y|$, which completes the proof. □

Remark 5.6. In Lemma 5.5, the restriction to the interval $[0,1]$ is necessary to ensure that $(t-s)\|Y\| \le \pi/2$. However, by reparametrization the inequality in Proposition 5.4 is sharp also for any other interval: If $I \subset \mathbb{R}$ is an arbitrary (bounded or unbounded) interval, then choose a $\mathcal{C}^1$-smooth function $\gamma\colon I \to [0,1]$ with $\dot\gamma > 0$. In this situation, the path $I \ni t \mapsto Q_t := P_{\gamma(t)}$ with P_τ as in Lemma 5.5 is $\mathcal{C}^1$-smooth and satisfies

$$\arcsin\bigl(\|Q_t - Q_s\|\bigr) = \int_s^t \|\dot Q_\tau\|\,\mathrm{d}\tau = \|Y\|\,\bigl(\gamma(t) - \gamma(s)\bigr)$$

whenever $s \le t$.

We have the following immediate corollary to Lemma 5.5.

Corollary 5.7 ([16, Proposition 5]; see also [36, Lemma 3.5])**.** *Let P and Q be two orthogonal projections such that* $\operatorname{Ran} P$ *and* $\operatorname{Ran} Q$ *are equivalently positioned, that is,*

$$\dim\bigl(\operatorname{Ran} P \cap \operatorname{Ran} Q^\perp\bigr) = \dim\bigl(\operatorname{Ran} P^\perp \cap \operatorname{Ran} Q\bigr)\,.$$

Then there is a $\mathcal{C}^1$-smooth path $[0,1] \ni t \mapsto P_t$ of orthogonal projections such that $P_0 = P$, $P_1 = Q$, and

$$\arcsin\bigl(\|P - Q\|\bigr) = \int_0^1 \|\dot P_\tau\|\,\mathrm{d}\tau\,.$$

Proof. Since $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ are equivalently positioned, there is a direct rotation $U = \exp(J\Theta)$ taking $\operatorname{Ran} P$ to $\operatorname{Ran} Q$, where Θ is the operator angle associated with $\operatorname{Ran} P$ and $\operatorname{Ran} Q$ and J is a suitable partial isometry, see Proposition 1.10 and equation (1.15). In particular, the operator $J\Theta$ is skew-symmetric, satisfies $\|J\Theta\| = \|\Theta\| \le \pi/2$, and is off-diagonal with respect to $\operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$. Now, set $Y := J\Theta$ and apply Lemma 5.5. □

The triangle inequality for the maximal angle between closed subspaces is now to a direct consequence of Proposition 5.4 and Corollary 5.7.

Proposition 5.8 ([16, Corollary 4]; see also [8, Lemma 2.15] and [36])**.** *Let P, Q, and R be orthogonal projections in a Hilbert space $\mathcal{H}$. Then*

$$\arcsin\bigl(\|P-Q\|\bigr) \le \arcsin\bigl(\|P-R\|\bigr) + \arcsin\bigl(\|R-Q\|\bigr)\,. \tag{5.12}$$

Proof. Clearly, we may assume that the right-hand side of (5.12) is less than $\pi/2$. In particular, $\operatorname{Ran} P$ and $\operatorname{Ran} R$, as well $\operatorname{Ran} R$ and $\operatorname{Ran} Q$, are in the acute-angle case (see Definition 1.9) and, therefore, equivalently positioned. Hence, by Corollary 5.7 there exist $\mathcal{C}^1$-smooth paths $[0,1] \ni t \mapsto R_t$ and $[0,1] \ni t \mapsto Q_t$ of orthogonal projections such that $R_0 = P$, $R_1 = R = Q_0$, and $Q_1 = Q$, as well as

$$\arcsin\bigl(\|P-R\|\bigr) = \int_0^1 \|\dot{R}_\tau\|\,\mathrm{d}\tau\,, \quad \arcsin\bigl(\|R-Q\|\bigr) = \int_0^1 \|\dot{Q}_\tau\|\,\mathrm{d}\tau\,. \tag{5.13}$$

Set $P_t := R_t$ for $t \in [0,1]$ and $P_t := Q_{t-1}$ for $t \in [1,2]$. Then, the path $[0,2] \ni t \mapsto P_t$ is piecewise $\mathcal{C}^1$-smooth with $P_0 = P$ and $P_2 = Q$. Thus, Proposition 5.4 implies that

$$\arcsin\bigl(\|P-Q\|\bigr) \le \int_0^2 \|\dot{P}_\tau\|\,\mathrm{d}\tau = \int_0^1 \|\dot{R}_\tau\|\,\mathrm{d}\tau + \int_0^1 \|\dot{Q}_\tau\|\,\mathrm{d}\tau\,,$$

which, in view of (5.13), proves inequality (5.12). □

For the sake of completeness, we close this chapter by discussing the converse line of reasoning, which establishes the arcsine law by use of the triangle inequality for the maximal angle. In this sense, the arcsine law and the triangle inequality for the maximal angle turn out to be equivalent.

Remark 5.9. If the triangle inequality for the maximal angle is taken for granted, Proposition 5.4 can also be shown directly by using the standard estimate from Lemma 1.16. Indeed, fix an arbitrary $\alpha > 1$. Since $\arcsin(x)/x \to 1$ as $x \to 0$, one has $\arcsin(x) \le \alpha x$ for sufficiently small $x \ge 0$. Hence, taking into account that the path $\tau \mapsto P_\tau$ is uniformly continuous on $[s,t]$, one can choose a partition $s = \tau_0 < \cdots < \tau_{n+1} = t$, $n \in \mathbb{N}_0$, of the interval $[s,t]$ with sufficiently small mesh size such that $[\tau_j, \tau_{j+1}] \ni \tau \mapsto P_\tau$ is $\mathcal{C}^1$-smooth and

$$\arcsin\bigl(\|P_{\tau_{j+1}} - P_{\tau_j}\|\bigr) \le \alpha\|P_{\tau_{j+1}} - P_{\tau_j}\| \quad \text{for} \quad j = 0,\dots,n\,.$$

The triangle inequality for the maximal angle and Lemma 1.16 then imply that

$$\begin{aligned}\arcsin\bigl(\|P_t - P_s\|\bigr) &\le \sum_{j=0}^{n} \arcsin\bigl(\|P_{\tau_{j+1}} - P_{\tau_j}\|\bigr) \le \alpha \sum_{j=0}^{n} \|P_{\tau_{j+1}} - P_{\tau_j}\| \\ &\le \alpha \sum_{j=0}^{n} \int_{\tau_j}^{\tau_{j+1}} \|\dot{P}_\tau\| \,\mathrm{d}\tau = \alpha \int_s^t \|\dot{P}_\tau\| \,\mathrm{d}\tau\,.\end{aligned}$$

Since $\alpha > 1$ has been chosen arbitrarily, this proves the arcsine law.

Chapter 6

Smooth variations of spectral subspaces

The present chapter is based on the joint work [37] with K. A. Makarov published in *Journal für die reine und angewandte Mathematik*. The considerations there are directly extended to unbounded operators here. Some material has also been added, and some proofs have been modified.

We study the variation of spectral subspaces associated with self-adjoint operators depending smoothly on a parameter. More precisely, let

$$I \ni t \mapsto B_t$$

be a $\mathcal{C}^1$-smooth path of possibly unbounded self-adjoint operators (cf. Section 1.6) such that the spectrum of each B_t is separated into two disjoint components, that is,

$$\operatorname{spec}(B_t) = \omega_t \cup \Omega_t \quad \text{with} \quad \operatorname{dist}(\omega_t, \Omega_t) > 0\,.$$

Under the additional assumption that the spectral components depend upper semicontinuously on the parameter (see Definition 6.2 below), based on the arcsine law for $\mathcal{C}^1$-smooth paths of orthogonal projections discussed in Chapter 5, we obtain the following a posteriori type bound on the maximal angle between the corresponding spectral subspaces (see Theorem 6.10

below):

$$\arcsin\bigl(\|P_t - P_s\|\bigr) \le \frac{\pi}{2}\int_s^t \frac{\|\dot{B}_\tau\|}{\operatorname{dist}(\omega_\tau,\Omega_\tau)}\,\mathrm{d}\tau \qquad \text{whenever} \qquad s \le t\,.$$

In the setting of unbounded operators B_t, this bound turns out to be sharp in the sense that equality can be attained, see Lemma 6.11 below. However, the constant $\pi/2$ in the estimate above is still optimal if the considerations are restricted to bounded operators B_t, see Remark 6.12. Corollary 6.13 below treats the particular case where the spectral components are additionally assumed to be subordinated or annular separated.

Finally, in Section 6.2 below, the above result is applied in the situation of the subspace perturbation problem discussed in Chapter 2. A conjecture on the optimality of the above estimate in this situation is also formulated there, see Conjecture 6.19 below and the corresponding discussion.

6.1 Paths of self-adjoint operators with separated spectra

In this section, we study paths of spectral projections associated with isolated components of the spectra of self-adjoint operators depending smoothly on a parameter. The objective is to obtain efficient estimates on the corresponding maximal angles in terms of the evolution of the operator path and the distance between the spectral components.

For notational setup we fix the following assumptions.

Hypothesis 6.1. *Let $I \ni t \mapsto B_t$ be a continuous path of self-adjoint operators on a Hilbert space $\mathcal{H}$. Suppose that the spectrum of each B_t is separated into two disjoint components, that is,*

$$\operatorname{spec}(B_t) = \omega_t \cup \Omega_t \qquad \text{with} \qquad \operatorname{dist}(\omega_t,\Omega_t) > 0\,.$$

Finally, denote by $P_t := \mathsf{E}_{B_t}(\omega_t)$, $t \in I$, the spectral projection for B_t associated with the set ω_t.

Due to the upper semicontinuity of the spectrum (see Lemma 1.17), $\operatorname{spec}(B_t)$ does not change by much for small variations of the parameter t. In view of Corollary 1.18, it is natural to require that the same holds for the spectral components ω_t and Ω_t. In this regard, we recall the concept of an upper semicontinuous family of sets depending on a parameter.

Definition 6.2. A family of sets $\{\Delta_t\}_{t\in I}$, $\Delta_t \subset \mathbb{R}$, with $I \subset \mathbb{R}$ an interval is said to be *upper semicontinuous* if for every $\varepsilon > 0$ and every $s \in I$ there is $\delta > 0$ such that

$$\Delta_t \subset \mathcal{O}_\varepsilon(\Delta_s) \quad \text{for} \quad t \in I \ \text{ with } \ |t-s| < \delta\,.$$

It turns out that, in the situation of Hypothesis 6.1, the upper semicontinuity of the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ is necessary and sufficient for the corresponding path $t \mapsto P_t$ of spectral projections to be continuous in norm, see Lemma 6.6 below and the discussion thereafter; see also Lemma 6.7. Note that the family $\{\operatorname{spec}(B_t)\}_{t\in I}$ is, by Lemma 1.17, upper semicontinuous in the sense of Definition 6.2. However, this does not guarantee that the two families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous as well. This depends on the choice of the components ω_t and Ω_t.

A corresponding choice of the spectral components is demonstrated in the following lemma, the proof of which is purely technical.

Lemma 6.3. *Let $I \ni t \mapsto B_t$ be a continuous path of self-adjoint operators. Suppose that for some $s \in I$ the spectrum of B_s is separated into two disjoint components, that is,*

$$\operatorname{spec}(B_s) = \omega_s \cup \Omega_s \quad \textit{with} \quad \operatorname{dist}(\omega_s, \Omega_s) > 0\,.$$

Moreover, assume that for all $t \in I$ there is r_t with $0 < r_t < \operatorname{dist}(\omega_s, \Omega_s)/2$ and

$$\operatorname{spec}(B_t) \subset \overline{\mathcal{O}_{r_t}\bigl(\operatorname{spec}(B_s)\bigr)}\,. \tag{6.1}$$

Then:

(a) The spectrum of each B_t is separated as

$$\operatorname{spec}(B_t) = \omega_t \cup \Omega_t\,,$$

where

$$\omega_t := \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{r_t}(\omega_s)} \quad \textit{and} \quad \Omega_t := \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{r_t}(\Omega_s)} \tag{6.2}$$

are nonempty closed sets.

(b) The two families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous.

Proof. (a). In view of (6.1), it remains to show that ω_t and Ω_t are nonempty.

Suppose that $\omega_t = \varnothing$ or $\Omega_t = \varnothing$ for some $t > s$. The case $t < s$ can be treated analogously. Define

$$\tau_0 := \inf\{\tau \in [s,t] \mid \omega_\tau = \varnothing \ \text{ or } \ \Omega_\tau = \varnothing\} \le t\,.$$

It follows from Corollary 1.18 and the continuity of the path $\tau \mapsto B_\tau$ that $\tau_0 > s$. In particular, one has $\omega_\tau \ne \varnothing \ne \Omega_\tau$ whenever $s \le \tau < \tau_0$.

Let $\varepsilon > 0$ such that $r_{\tau_0} + \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$, and choose τ with $s \le \tau < \tau_0$ and $\|B_\tau - B_{\tau_0}\| < \varepsilon$. The upper semicontinuity of the spectrum (Lemma 1.17) implies that

$$\omega_\tau \cup \Omega_\tau \subset \mathcal{O}_\varepsilon(\omega_{\tau_0}) \cup \mathcal{O}_\varepsilon(\Omega_{\tau_0})\,.$$

By (6.2) one has $\mathcal{O}_\varepsilon(\omega_{\tau_0}) \subset \mathcal{O}_{r_{\tau_0}+\varepsilon}(\omega_s)$ and $\mathcal{O}_\varepsilon(\Omega_{\tau_0}) \subset \mathcal{O}_{r_{\tau_0}+\varepsilon}(\Omega_s)$. Since $r_{\tau_0} + \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$ and the sets ω_τ and Ω_τ are nonempty, one concludes that ω_{τ_0} and Ω_{τ_0} are nonempty as well. In particular, one has $\tau_0 < t$.

Now, choose an arbitrary $\tau \in (\tau_0, t)$ with $\|B_\tau - B_{\tau_0}\| < \varepsilon$. Taking into account that $r_{\tau_0} + \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$, Corollary 1.18 then implies that

$$\omega_\tau = \operatorname{spec}(B_\tau) \cap \overline{\mathcal{O}_{\|B_\tau - B_{\tau_0}\|}(\omega_{\tau_0})} \quad \text{and} \quad \Omega_\tau = \operatorname{spec}(B_\tau) \cap \overline{\mathcal{O}_{\|B_\tau - B_{\tau_0}\|}(\Omega_{\tau_0})}$$

are nonempty. We have thus shown that there is $\tau_1 > \tau_0$ such that ω_τ and Ω_τ are nonempty whenever $s \le \tau \le \tau_1$, which is a contradiction to the definition of τ_0. Hence, ω_t and Ω_t are nonempty for all t, so that (a) holds.

(b). Let $t \in I$ be arbitrary, and choose $\varepsilon > 0$ with $r_t + \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$. By hypothesis, there is $\delta > 0$ such that $\|B_\tau - B_t\| < \varepsilon$ whenever $|\tau - t| < \delta$. It then follows from the upper semicontinuity of the spectrum (Lemma 1.17) that

$$\omega_\tau \cup \Omega_\tau \subset \overline{\mathcal{O}_{\|B_\tau - B_t\|}(\omega_t \cup \Omega_t)} \subset \mathcal{O}_\varepsilon(\omega_t) \cup \mathcal{O}_\varepsilon(\Omega_t)\,, \quad |\tau - t| < \delta\,.$$

By (6.2) one has $\mathcal{O}_\varepsilon(\omega_t) \subset \mathcal{O}_{r_t+\varepsilon}(\omega_s)$ and $\mathcal{O}_\varepsilon(\Omega_t) \subset \mathcal{O}_{r_t+\varepsilon}(\Omega_s)$. Taking into account the inequality $r_t + \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$, one deduces from the definition of ω_τ and Ω_τ that $\omega_\tau \cap \mathcal{O}_\varepsilon(\Omega_t) = \varnothing = \Omega_\tau \cap \mathcal{O}_\varepsilon(\omega_t)$ whenever

$|\tau - t| < \delta$. Hence,

$$\omega_\tau \subset \mathcal{O}_\varepsilon(\omega_t) \quad \text{and} \quad \Omega_\tau \subset \mathcal{O}_\varepsilon(\Omega_t) \quad \text{whenever} \quad |\tau - t| < \delta\,,$$

which proves (b). □

In the situation of Hypothesis 6.1, the family $\{\operatorname{spec}(B_t)\}_{t\in I}$ even is *continuous* in the following sense: For arbitrary $s \in I$ and $\varepsilon > 0$ there is $\delta > 0$ such that

$$\operatorname{spec}(B_t) \subset \mathcal{O}_\varepsilon\bigl(\operatorname{spec}(B_s)\bigr) \quad \text{and} \quad \operatorname{spec}(B_s) \subset \mathcal{O}_\varepsilon\bigl(\operatorname{spec}(B_t)\bigr)$$

for all $t \in I$ with $|t - s| < \delta$, see Lemma 1.17 and the discussion thereafter. If, in addition, the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are assumed to be upper semicontinuous, then the following observation shows that these two families are continuous in the same sense as well.

Remark 6.4. In addition to Hypothesis 6.1, suppose that the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous. Let $s \in I$ and $0 < \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$ be arbitrary. Choose $\delta > 0$ such that for all $t \in I$ with $|t - s| < \delta$ one has $\|B_t - B_s\| < \varepsilon$, as well as

$$\omega_t \subset \mathcal{O}_\varepsilon(\omega_s) \quad \text{and} \quad \Omega_t \subset \mathcal{O}_\varepsilon(\Omega_s)\,.$$

The upper semicontinuity of the spectrum (Lemma 1.17) implies that

$$\omega_s \cup \Omega_s \subset \overline{\mathcal{O}_{\|B_t - B_s\|}(\omega_t \cup \Omega_t)} \subset \mathcal{O}_\varepsilon(\omega_t) \cup \mathcal{O}_\varepsilon(\Omega_t)\,, \quad |t - s| < \delta\,.$$

Taking into account that $2\varepsilon < \operatorname{dist}(\omega_s, \Omega_s)$, the inclusions $\mathcal{O}_\varepsilon(\omega_t) \subset \mathcal{O}_{2\varepsilon}(\omega_s)$ and $\mathcal{O}_\varepsilon(\Omega_t) \subset \mathcal{O}_{2\varepsilon}(\Omega_s)$ yield that $\omega_s \cap \mathcal{O}_\varepsilon(\Omega_t) = \varnothing = \Omega_s \cap \mathcal{O}_\varepsilon(\omega_t)$. Therefore, one has

$$\omega_s \subset \mathcal{O}_\varepsilon(\omega_t) \quad \text{and} \quad \Omega_s \subset \mathcal{O}_\varepsilon(\Omega_t) \quad \text{whenever} \quad |t - s| < \delta\,.$$

The preceding observation can be used to show that the choice (6.2) of the components ω_t and Ω_t in Lemma 6.3 is unique in the following sense: In the situation of Lemma 6.3, let $\{\sigma_t\}_{t\in I}$ and $\{\Sigma_t\}_{t\in I}$ be upper semicontinuous families such that

$$\operatorname{spec}(B_t) = \sigma_t \cup \Sigma_t \quad \text{and} \quad \operatorname{dist}(\sigma_t, \Sigma_t) > 0 \quad \text{for} \quad t \in I\,.$$

If $\sigma_s = \omega_s$ and $\Sigma_s = \Omega_s$, then one has $\sigma_t = \omega_t$ and $\Sigma_t = \Omega_t$ for all $t \in I$. This can be shown with a technique similar to the one used for Lemma 6.3 (a). We omit the proof here since we do not really need this statement. Nevertheless, the inclusion (6.1) can locally always be satisfied for continuous paths of operators with separated spectra, so that Lemma 6.3 describes the typical situation for our considerations.

One immediate advantage of upper semicontinuous separated components is that the distance between the components depends continuously on the parameter.

Lemma 6.5 (cf. [36, Lemma C.1])**.** *Assume Hypothesis 6.1. Suppose, in addition, that the two families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous. Then, the mapping $I \ni t \mapsto \operatorname{dist}(\omega_t, \Omega_t)$ is continuous.*

Proof. Let $s \in I$ and $0 < \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$ be arbitrary. Taking into account Remark 6.4, there is $\delta > 0$ such that for all $t \in I$ with $|t - s| < \delta$ one has

$$\omega_t \subset \mathcal{O}_\varepsilon(\omega_s) \quad \text{and} \quad \Omega_t \subset \mathcal{O}_\varepsilon(\Omega_s) \tag{6.3}$$

as well as

$$\omega_s \subset \mathcal{O}_\varepsilon(\omega_t) \quad \text{and} \quad \Omega_s \subset \mathcal{O}_\varepsilon(\Omega_t)\,. \tag{6.4}$$

Combining (6.3) and (6.4) yields that

$$\operatorname{dist}(\omega_s, \Omega_s) - 2\varepsilon \le \operatorname{dist}(\omega_t, \Omega_t) \le \operatorname{dist}(\omega_s, \Omega_s) + 2\varepsilon$$

whenever $|t - s| < \delta$, which proves the claim. □

For the study of the variation of the spectral subspaces $\operatorname{Ran} P_t$ in Hypothesis 6.1, it is natural to require that the path $t \mapsto P_t$ of spectral projections is $\mathcal{C}^1$-smooth in norm, or at least continuous. The following lemma shows that this is the case only if the corresponding separated spectral components depend upper semicontinuously on the parameter.

Lemma 6.6. *Assume Hypothesis 6.1. If the path $I \ni t \mapsto P_t$ is continuous in norm, then the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous.*

Proof. Let $s \in I$ and $0 < \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)/2$ be arbitrary. By hypothesis, there is $\delta > 0$ such that for all $t \in I$ with $|t - s| < \delta$ one has

$$\|B_t - B_s\| < \frac{\varepsilon}{2} \quad \text{and} \quad \|P_t - P_s\| \le \frac{\sqrt{2}}{2} \,. \tag{6.5}$$

Fix an arbitrary $t \in I$ with $|t - s| < \delta$, and set $V := \overline{B_t - B_s} \in \mathcal{L}(\mathcal{H})$. Let

$$V = \begin{pmatrix} V_0 & W \\ W^* & V_1 \end{pmatrix} \quad \text{and} \quad B_s = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}$$

with $\operatorname{Dom}(B_s) = \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)$ be the representations of V and B_s as 2×2 block operator matrices with respect to the orthogonal decomposition $\mathcal{H} = \operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^\perp$. In particular, one has

$$\omega_s = \operatorname{spec}(A_0) \quad \text{and} \quad \Omega_s = \operatorname{spec}(A_1) \,.$$

Since $\|P_t - P_s\| \le \sqrt{2}/2 < 1$ by (6.5), it follows from Proposition 1.13 that there is a unique operator $X \in \mathcal{L}(\operatorname{Ran} P_s, \operatorname{Ran} P_s^\perp)$ such that $\operatorname{Ran} P_t$ is the graph of X, that is, $\operatorname{Ran} P_t = \mathcal{G}(\operatorname{Ran} P_s, X)$. This operator X satisfies

$$\|X\| = \frac{\|P_t - P_s\|}{\sqrt{1 - \|P_t - P_s\|^2}} \le 1 \,. \tag{6.6}$$

As a spectral subspace the graph $\mathcal{G}(\operatorname{Ran} P_s, X) = \operatorname{Ran} P_t$ is reducing for B_t. Considering $\operatorname{Dom}(A_0 + V_0) = \operatorname{Dom}(A_0)$, $\operatorname{Dom}(A_1 + V_1) = \operatorname{Dom}(A_1)$, and

$$B_t = B_s + V = \begin{pmatrix} A_0 + V_0 & 0 \\ 0 & A_1 + V_1 \end{pmatrix} + \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix} ,$$

it then follows from Theorem 4.7 that the operator B_t is similar to a block diagonal operator $\Lambda = \Lambda_0 \oplus \Lambda_1$ with respect to $\mathcal{H} = \operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^\perp$, where

$$\omega_t = \operatorname{spec}(\Lambda_0) = \operatorname{spec}(A_0 + V_0 + WX)$$

and

$$\Omega_t = \operatorname{spec}(\Lambda_1) = \operatorname{spec}(A_1 + V_1 - W^* X^*) \,.$$

Since by (6.5) and (6.6) one has

$$\|V_0 + WX\| \le \|V_0\| + \|W\|\,\|X\| \le 2\|V\| < \varepsilon\,,$$

the upper semicontinuity of the spectrum (Lemma 1.17) yields that

$$\omega_t = \operatorname{spec}(A_0 + V_0 + WX) \subset \mathcal{O}_\varepsilon\bigl(\operatorname{spec}(A_0)\bigr) = \mathcal{O}_\varepsilon(\omega_s)\,.$$

Analogously one obtains that

$$\Omega_t = \operatorname{spec}(A_1 + V_1 - W^*X^*) \subset \mathcal{O}_\varepsilon(\Omega_s)\,.$$

Since $t \in I$ with $|t-s| < \delta$ has been chosen arbitrarily, this proves the claim. □

The $\sin\Theta$ theorem guarantees that the converse of Lemma 6.6 is also valid, that is, the path $t \mapsto P_t$ is continuous if the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous. Indeed, let $s \in I$ and $0 < \varepsilon < \operatorname{dist}(\omega_s, \Omega_s)$ be arbitrary. The upper semicontinuity then implies that there is $\delta > 0$ such that $\omega_t \subset \mathcal{O}_\varepsilon(\omega_s)$ and $\Omega_t \subset \mathcal{O}_\varepsilon(\Omega_s)$ for $t \in I$ with $|t-s| < \delta$. In particular, for those t one has $\operatorname{dist}(\omega_t, \Omega_s) \ge d$ and $\operatorname{dist}(\Omega_t, \omega_s) \ge d$, where $d := \operatorname{dist}(\omega_s, \Omega_s) - \varepsilon > 0$. It now follows from the symmetric $\sin\Theta$ theorem (Proposition 3.7) that

$$\|P_t - P_s\| \le \frac{\pi}{2}\,\frac{\|B_t - B_s\|}{\operatorname{dist}(\omega_s, \Omega_s) - \varepsilon}\,, \quad |t-s| < \delta\,, \tag{6.7}$$

from which one concludes that $\|P_t - P_s\| \to 0$ as $t \to s$. A similar reasoning can be found in the proof of the particular case discussed in [8, Theorem 3.5].

If, in addition, the two paths $t \mapsto B_t$ and $t \mapsto P_t$ are assumed to be $\mathcal{C}^1$-smooth, then one obtains from (6.7) the estimate

$$\|\dot P_s\| \le \frac{\pi}{2}\,\frac{\|\dot B_s\|}{\operatorname{dist}(\omega_s, \Omega_s)}\,, \quad s \in I\,. \tag{6.8}$$

This follows by dividing both sides of (6.7) by $|t-s|$, taking the limit as $t \to s$, and finally letting ε approach zero.

A closer look at the proof of the $\sin\Theta$ theorem shows that the upper semicontinuity of the separated components ω_t and Ω_t does not only yield the continuity of the path $t \mapsto P_t$, it is also sufficient for this path to inherit smoothness of the path $t \mapsto B_t$.

Lemma 6.7. *In addition to Hypothesis 6.1, suppose that the two families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous. If the path $I \ni t \mapsto B_t$ is $\mathcal{C}^1$-smooth, then so is $I \ni t \mapsto P_t$, and for each t the derivative $\dot{P}_t$ is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P_t \oplus \operatorname{Ran} P_t^\perp$. Moreover, $Y = \dot{P}_t$ is a strong solution to the Sylvester equation*

$$YB_t - B_tY = P_t^\perp\, \overline{\dot{B}_t}\, P_t - P_t\, \overline{\dot{B}_t}\, P_t^\perp\,. \tag{6.9}$$

Proof. Taking into account that $B_s = B_t + (B_s - B_t) = B_t + \overline{B_s - B_t}$ with $\overline{B_s - B_t} \in \mathcal{L}(\mathcal{H})$, it follows from Lemma 3.6 that for all $s, t \in I$ the operator $Z = P_t - P_s$ is a strong solution to the Sylvester equation

$$ZB_s - B_tZ = P_t\, \overline{B_s - B_t}\, P_s^\perp - P_t^\perp\, \overline{B_s - B_t}\, P_s\,. \tag{6.10}$$

Let $t_0 \in I$ and $0 < d < \operatorname{dist}(\omega_{t_0}, \Omega_{t_0})$ be arbitrary. Since the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous by hypothesis, there is $\delta > 0$ such that for all $s, t \in (t_0 - \delta, t_0 + \delta) \cap I$ one has

$$\operatorname{dist}(\omega_t, \Omega_s) \ge d \quad\text{ and }\quad \operatorname{dist}(\Omega_t, \omega_s) \ge d\,.$$

Furthermore, the difference $P_t - P_s = P_tP_s^\perp - P_t^\perp P_s$ has an off-diagonal representation as an operator from $\operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^\perp$ to $\operatorname{Ran} P_t \oplus \operatorname{Ran} P_t^\perp$. Hence, for $s, t \in (t_0 - \delta, t_0 + \delta) \cap I$, Corollary 3.5 implies that

$$P_t - P_s = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\tau B_t}\bigl(P_t\, \overline{B_s - B_t}\, P_s^\perp - P_t^\perp\, \overline{B_s - B_t}\, P_s\bigr)\, \mathrm{e}^{-\mathrm{i}\tau B_s} f_d(\tau)\, \mathrm{d}\tau\,, \tag{6.11}$$

where the integral is understood in the weak sense and $f_d \in L^1(\mathbb{R})$ is any function as in Corollary 3.5.

We already know that P_t converges to P_s in norm as t approaches s, cf. inequality (6.7) above. Moreover, it follows from the classic theory of strongly continuous semigroups that for all $\tau \in \mathbb{R}$ the unitary operator $\mathrm{e}^{\mathrm{i}\tau B_t}$ converges to $\mathrm{e}^{\mathrm{i}\tau B_s}$ in norm as t approaches s, see, e.g., [41, Corollary 3.1.3] and [43, Theorem VIII.7]; this also follows from the more recent results on operator continuous functions due to Aleksandrov and Peller, see [3, Section 8]. Taking into account that $f_d \in L^1(\mathbb{R})$, one now concludes from (6.11) by

Lebesgue's dominated convergence theorem that the derivative $\dot{P}_s$ exists in norm sense with

$$(6.12)\quad \dot{P}_s = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\tau B_s} \Big(P_s^{\perp}\, \overline{\dot{B}_s}\, P_s - P_s\, \overline{\dot{B}_s}\, P_s^{\perp} \Big)\, \mathrm{e}^{-\mathrm{i}\tau B_s} f_d(\tau)\, \mathrm{d}\tau\,, \qquad |s - t_0| < \delta\,,$$

cf. also the proofs of Lemmas 1.16 and 3.12.

One verifies by inspection that $\dot{P}_s$ is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^{\perp}$. Moreover, comparing (6.12) with the representation (3.13) in Corollary 3.5, one infers that $Y = \dot{P}_s$ is a strong solution to (6.9) with s instead of t; this also follows by dividing both sides of (6.10) by $t - s$ and taking the limit as t approaches s.

Finally, again by Lebesgue's dominated convergence theorem, one deduces from representation (6.12) that the path $t \mapsto \dot{P}_t$ is continuous in norm on $(t_0 - \delta, t_0 + \delta) \cap I$, so that $t \mapsto P_t$ is $\mathcal{C}^1$-smooth. This completes the proof. □

Remark 6.8. Taking into account the Sylvester equation (6.9) in Lemma 6.7, inequality (6.8) now also follows from the corresponding representation of the derivative given by Corollary 3.5.

Remark 6.9. Lemma 6.7 can also be shown by means of the Daleckiĭ-Kreĭn differentiation formula from [17], see [37, Theorem 2.2 and Corollary 2.4]. The corresponding proof is more elegant, but requires the double operator integral calculus. A survey on this topic can be found, for example, in [14].

Yet another proof is available if the family $\{\omega_t\}_{t\in I}$ (resp. $\{\Omega_t\}_{t\in I}$) consists of bounded sets. In this case, the projection P_t (resp. $P_t^{\perp}$) can be represented as a contour integral for the resolvent of B_t, see [36, Lemma D.1]. The corresponding reasoning is essentially the same as the one in [25, Theorem II.5.4].

We are now ready to turn to the main result of this chapter.

Theorem 6.10 (see [37, Theorem 2.2]). *Assume Hypothesis 6.1. Suppose, in addition, that the families $\{\omega_t\}_{t\in I}$ and $\{\Omega_t\}_{t\in I}$ are upper semicontinuous and that the path $I \ni t \mapsto B_t$ is $\mathcal{C}^1$-smooth in norm. Then*

$$\arcsin\big(\|P_t - P_s\|\big) \le \frac{\pi}{2} \int_s^t \frac{\|\dot{B}_\tau\|}{\operatorname{dist}(\omega_\tau, \Omega_\tau)}\, \mathrm{d}\tau \qquad \textit{whenever} \qquad s \le t\,.$$

Proof. By Lemma 6.7, the path $I \ni t \mapsto P_t$ is $\mathcal{C}^1$-smooth. The claim then follows by combining the arcsine law (Proposition 5.4) and inequality (6.8). □

The following lemma shows that the estimate in Theorem 6.10 is sharp in the sense that equality can be attained. The corresponding example is based on Lemma 5.5 with a suitable choice of the operator $Y \in \mathcal{L}(\mathcal{H})$.

Lemma 6.11 (cf. [37, Remark 2.3]). *Let $I \subset \mathbb{R}$ be an arbitrary interval. Then, there exist non-empty closed subsets ω and Ω of $\mathbb{R}$ with $\operatorname{dist}(\omega, \Omega) = 1$ and a $\mathcal{C}^1$-smooth path $I \ni t \mapsto B_t$ of (unbounded) self-adjoint operators such that*

(i) the spectrum of each B_t is separated as

$$\operatorname{spec}(B_t) = \omega \cup \Omega\,;$$

and

(ii) one has

$$\arcsin\bigl(\|P_t - P_s\|\bigr) = \frac{\pi}{2}\int_s^t \|\dot B_\tau\|\,\mathrm{d}\tau \qquad \textit{whenever} \qquad s \le t\,,$$

where $P_t := \mathsf{E}_{B_t}(\omega)$, $t \in I$, denotes the spectral projection for B_t associated with the spectral component ω.

Proof. By reparametrization we may restrict the considerations to the case $I = [0,1]$, cf. Remark 5.6.

Let A_0, A_1, X, and K be as in Remark 3.3, and set $\omega := \operatorname{spec}(A_0)$ and $\Omega := \operatorname{spec}(A_1)$. In particular, one has

$$\operatorname{dist}(\omega, \Omega) = 1\,.$$

On $\mathcal{H} := \ell^2 \oplus \ell^2$ define the self-adjoint operator $A := A_0 \oplus A_1$ with $\operatorname{Dom}(A) := \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)$. Moreover, let the bounded operators Y and R on $\mathcal{H}$ be given by

$$Y := \frac{2}{\pi} \cdot \begin{pmatrix} 0 & -X^* \\ X & 0 \end{pmatrix} \qquad \text{and} \qquad R := \frac{2}{\pi} \cdot \begin{pmatrix} 0 & K^* \\ K & 0 \end{pmatrix}\,.$$

Since X is a strong solution to the Sylvester equation $XA_0 - A_1X = K$, the operator $Z = -X^*$ is a strong solution to $ZA_1 - A_0Z = K^*$, cf. equation (3.3). Thus, in view of (3.10), one concludes that Y is a strong solution to the Sylvester equation

$$YA - AY = R$$

satisfying

$$\|Y\| = \frac{\pi}{2}\|R\| = \frac{\pi}{2}\,. \tag{6.13}$$

It then follows from Lemma 3.12 that the path

$$[0,1] \ni t \mapsto B_t := \exp(tY)A\exp(-tY)\,, \quad \mathrm{Dom}(B_t) := \mathrm{Dom}(A)\,,$$

is $\mathcal{C}^1$-smooth with $\dot{B}_t = \exp(tY)R\exp(-tY)|_{\mathrm{Dom}(A)}$.

Since $\exp(tY)$ is unitary, the spectrum of each B_t clearly is separated as $\mathrm{spec}(B_t) = \mathrm{spec}(A) = \omega \cup \Omega$, so that (i) holds. Moreover, one has

$$\|\dot{B}_t\| = \|R\|\,, \quad t \in [0,1]\,. \tag{6.14}$$

On the other hand, Lemma 5.5 implies that the corresponding path of spectral projections

$$[0,1] \ni t \mapsto P_t = \mathsf{E}_{B_t}(\omega) = \exp(tY)\mathsf{E}_A(\omega)\exp(-tY)$$

is $\mathcal{C}^1$-smooth with

$$\arcsin\bigl(\|P_t - P_s\|\bigr) = \int_s^t \|\dot{P}_\tau\|\,\mathrm{d}\tau = \|Y\|\,(t-s) \quad \text{whenever} \quad s \le t\,. \tag{6.15}$$

The claim (ii) now follows by combining (6.13)–(6.15). □

Remark 6.12. For paths $t \mapsto B_t$ of bounded self-adjoint operators, the constant $\pi/2$ in the estimate in Theorem 6.10 is still optimal. This can be seen by truncating the above example to the finite-dimensional case, cf. Remark 3.3.

Favourable geometry

Although the estimate in Theorem 6.10 is optimal in the sense of Lemma 6.11 and Remark 6.12, one may obtain a stronger result if additional information on the mutual disposition of the spectral sets ω_t and Ω_t is available. Using the original Davis-Kahan symmetric $\sin\Theta$ theorem (see Remark 3.8), we immediately arrive at the following corollary to Theorem 6.10; in fact, it is a corollary rather to the proof of Theorem 6.10 than to its actual statement.

Corollary 6.13. *In addition to the hypotheses of Theorem 6.10, assume that for all $t \in I$ the convex hull of one of the spectral components ω_t and Ω_t is disjoint from the other component, that is,* $\operatorname{conv}(\omega_t) \cap \Omega_t = \emptyset$ *or vice versa. Then*

$$\arcsin\bigl(\|P_t - P_s\|\bigr) \le \int_s^t \frac{\|\dot{B}_\tau\|}{\operatorname{dist}(\omega_\tau, \Omega_\tau)}\,\mathrm{d}\tau \qquad \textit{whenever} \qquad s \le t\,.$$

As in Lemma 6.11, the result of Corollary 6.13 is sharp, but this time also in the setting of bounded self-adjoint operators B_t. In the particular case where the spectral components ω_t and Ω_t are subordinated, that is, $\sup\omega_t < \inf\Omega_t$ or vice versa, we even have the following illustrative example (see [37, Remark 2.5]):

Let $[0,1] \ni t \mapsto P_t$ be a $\mathcal{C}^1$-smooth path of orthogonal projections as in Lemma 5.5, and set $B_t := P_t$, $\omega_t := \{1\}$, and $\Omega_t := \{0\}$. Then, one has $\operatorname{spec}(B_t) = \omega_t \cup \Omega_t$, $\operatorname{dist}(\omega_t, \Omega_t) = 1$, and

$$\arcsin\bigl(\|P_t - P_s\|\bigr) = \int_s^t \|\dot{P}_\tau\|\,\mathrm{d}\tau = \int_s^t \frac{\|\dot{B}_\tau\|}{\operatorname{dist}(\omega_\tau, \Omega_\tau)}\,\mathrm{d}\tau$$

whenever $s \le t$.

6.2 Application to the subspace perturbation problem

In this section, we apply Theorem 6.10 in the context of the subspace perturbation problem discussed in Chapter 2.

For convenience, we recall the following assumptions.

Hypothesis 6.14. *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$ such that the spectrum of A is separated into two disjoint components, that* *is,*

$$\operatorname{spec}(A) = \sigma \cup \Sigma \quad \textit{with} \quad d := \operatorname{dist}(\sigma, \Sigma) > 0\,.$$

Let $V \in \mathcal{L}(\mathcal{H})$ be self-adjoint. Finally, denote by

$$P := \mathsf{E}_A(\sigma) \quad \textit{and} \quad Q := \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$$

the spectral projections for A and $A+V$ associated with the sets σ and $\mathcal{O}_{d/2}(\sigma)$, respectively.

The main idea for applying Theorem 6.10 in the situation of Hypothesis 6.14 is to introduce a coupling parameter on the perturbation, that is, to consider the $\mathcal{C}^1$-smooth path

$$[0,1] \ni t \mapsto B_t := A + tV\,, \quad \operatorname{Dom}(B_t) := \operatorname{Dom}(A)\,,$$

cf. Example 1.15 (a). Recall that $\dot{B}_t = V|_{\operatorname{Dom}(A)}$. For $t \in [0,1]$ define

$$\omega_t := \operatorname{spec}(B_t) \cap \mathcal{O}_{d/2}(\sigma) \quad \text{and} \quad \Omega_t := \operatorname{spec}(B_t) \cap \mathcal{O}_{d/2}(\Sigma)\,, \tag{6.16}$$

and set $P_t := \mathsf{E}_{B_t}(\omega_t)$.

Under the additional assumption that for all $t \in [0,1]$ there is r_t with $0 < r_t < d/2$ and

$$\operatorname{spec}(B_t) \subset \overline{\mathcal{O}_{r_t}(\operatorname{spec}(A))}\,, \tag{6.17}$$

the spectrum of each B_t is separated as $\operatorname{spec}(B_t) = \omega_t \cup \Omega_t$, and the two families $\{\omega_t\}_{t\in[0,1]}$ and $\{\Omega_t\}_{t\in[0,1]}$ are upper semicontinuous, see Lemma 6.3. In this case, the path $[0,1] \ni t \mapsto P_t$ is $\mathcal{C}^1$-smooth by Lemma 6.7, and, taking into account that $P_0 = P$ and $P_1 = Q$, Theorem 6.10 yields that

$$\arcsin\bigl(\|P - Q\|\bigr) \le \frac{\pi}{2}\,\|V\| \int_0^1 \frac{\mathrm{d}\tau}{\operatorname{dist}(\omega_\tau, \Omega_\tau)}\,. \tag{6.18}$$

The author's guess is that estimate (6.18) is sharp in the sense that equality can be attained, or that at least the constant $\pi/2$ in (6.18) is optimal. Yet, no rigorous proof of this guess is available so far. This is discussed in more detail after Remark 6.17 below.

Using a priori knowledge on the distance between the spectral components ω_t and Ω_t, we arrive at the following result.

Theorem 6.15 ([37, Theorems 3.2 and 3.3]; see also [8, Theorem 3.5])**.** *Assume Hypothesis 6.14.*

(a) If V satisfies $\|V\| < c_{\mathrm{gen}}d$, where $c_{\mathrm{gen}} := \frac{\sinh(1)}{\mathrm{e}} < \frac{1}{2}$ is the root of the equation

$$\frac{\pi}{4}\log\Bigl(\frac{1}{1-2x}\Bigr) = \frac{\pi}{2}\,,$$

then

$$\arcsin\bigl(\|P-Q\|\bigr) \le \frac{\pi}{4}\log\Bigl(\frac{d}{d-2\|V\|}\Bigr) < \frac{\pi}{2}\,.$$

(b) Suppose, in addition, that the perturbation V is off-diagonal with respect to the orthogonal decomposition $\mathcal{H} = \operatorname{Ran}\mathsf{E}_A(\sigma) \oplus \operatorname{Ran}\mathsf{E}_A(\Sigma)$, that is,

$$\mathsf{E}_A(\sigma)V\mathsf{E}_A(\sigma) = 0 = \mathsf{E}_A(\Sigma)V\mathsf{E}_A(\Sigma)\,.$$

If $\|V\| < c_{\mathrm{off}}d$, where $c_{\mathrm{off}} < \sqrt{3}/2$ is the unique root of the equation

$$\int_0^x \frac{\mathrm{d}\tau}{1-2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} = 1\,, \tag{6.19}$$

then

$$\arcsin\bigl(\|P-Q\|\bigr) \le \frac{\pi}{2}\int_0^{\frac{\|V\|}{d}} \frac{\mathrm{d}\tau}{1-2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} < \frac{\pi}{2}\,.$$

Proof. (a). Let $\|V\| < c_{\mathrm{gen}}d < d/2$. Then, the inclusion (6.17) above holds with $r_t = t\|V\| < d/2$, see Lemma 1.17. In particular, the spectral components ω_t and Ω_t in (6.16) satisfy $\operatorname{dist}(\omega_t,\Omega_t) \ge d - 2t\|V\| > 0$. In view of estimate (6.18), it remains to observe that

$$\int_0^1 \frac{\|V\|}{d-2\tau\|V\|}\,\mathrm{d}\tau = \frac{1}{2}\log\Bigl(\frac{d}{d-2\|V\|}\Bigr) < \frac{1}{2}\log\Bigl(\frac{1}{1-2c_{\mathrm{gen}}}\Bigr) = 1\,.$$

(b). Since the improper integral

$$\int_0^{\frac{\sqrt{3}}{2}} \frac{\mathrm{d}\tau}{1-2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} = \infty$$

diverges, the root c_{off} of equation (6.19) is well defined and less than $\sqrt{3}/2$. Thus, if $\|V\| < c_{\mathrm{off}}d < \sqrt{3}d/2$, Lemma 1.21 implies that the inclusion (6.17) holds with $r_t = \delta_{tV} < d/2$, where

$$\delta_{tV} = t\|V\| \tan\Bigl(\frac{1}{2} \arctan \frac{2t\|V\|}{d}\Bigr)\,.$$

In particular, the components ω_t and Ω_t satisfy $\operatorname{dist}(\omega_t, \Omega_t) \geq d - 2\delta_{tV}$. Now, observe that

$$\begin{aligned}\int_0^1 \frac{\|V\|}{d - 2\delta_{tV}}\,\mathrm{d}\tau &= \int_0^{\frac{\|V\|}{d}} \frac{\mathrm{d}\tau}{1 - 2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)}\\ &< \int_0^{c_{\mathrm{off}}} \frac{\mathrm{d}\tau}{1 - 2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} = 1\,.\end{aligned}$$

In view of estimate (6.18), this proves (b). □

Remark 6.16 (cf. [37, Section 3]). The estimates obtained in Theorem 6.15 are stronger than the previously known estimates (2.22) and (2.30) derived from the $\sin\Theta$ theorem. Indeed, one has

$$\frac{\pi}{4}\log\Bigl(\frac{d}{d - 2\|V\|}\Bigr) < \arcsin\Bigl(\frac{\pi}{2}\,\frac{\|V\|}{d - \|V\|}\Bigr)$$

for $0 < \|V\| \leq \frac{2d}{2+\pi} < c_{\mathrm{gen}}d$, as well as

$$\frac{\pi}{2}\int_0^{\frac{\|V\|}{d}} \frac{\mathrm{d}\tau}{1 - 2\tau\tan\bigl(\frac{1}{2}\arctan(2\tau)\bigr)} < \arcsin\Bigl(\frac{\pi}{2}\,\frac{\|V\|}{d - \delta_V}\Bigr)$$

for $0 < \|V\| \leq c_\pi d < c_{\mathrm{off}}d$ with $\delta_V = \|V\|\tan\bigl(\frac{1}{2}\arctan\frac{2\|V\|}{d}\bigr)$ and c_π as in (2.30). The corresponding proofs of these inequalities are elementary, so that we omit them. They can be found in [37, Proposition A.1].

Remark 6.17. It is clear from Corollary 6.13 that the estimates in Theorem 6.15 can be strengthened if the spectral components σ and Σ are additionally assumed to be subordinated or annular separated. However, the resulting estimates are weaker than the ones discussed in Chapter 2 for these cases, so that we omit the details here.

On the optimality of estimate (6.18)

The presented way to obtain estimate (6.18) (resp. Theorem 6.10) essentially consists of two steps: the arcsine law for the path $t \mapsto P_t$ (Proposition

5.4) and the bound (6.8) on the norm of the derivative $\dot{P}_t$. Especially the accuracy of the latter influences the quality of the whole estimate. Thus, in order to conclude that (6.18) is sharp, or that at least the constant $\pi/2$ there is optimal, one has to find examples of operators A and V for which the bound (6.8) on the corresponding derivative $\dot{P}_t$ is accurate simultaneously for all t. At this point, it is natural to start with the consideration of parameters t close to 0.

Recall that, in the current situation, for $t = 0$ the operator $Y = \dot{P}_0$ is a strong solution to the Sylvester equation

$$YA - AY = P^\perp VP - PVP^\perp\,, \tag{6.20}$$

see Lemma 6.7. Moreover, $\dot{P}_0$ is off-diagonal with respect to the decomposition $\mathcal{H} = \operatorname{Ran} P \oplus \operatorname{Ran} P^\perp$. In view of Corollary 3.5, this again motivates to consider the sharpness example for the Sylvester equation discussed in Chapter 3:

Let A_0, A_1, X, and K as in Remark 3.3. On $\mathcal{H} := \ell^2 \oplus \ell^2$ define

$$V := \frac{2}{\pi} \cdot \begin{pmatrix} 0 & K^* \\ K & 0 \end{pmatrix} \qquad \text{and} \qquad A := \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} \tag{6.21}$$

with $\operatorname{Dom}(A) := \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)$. For this choice of A and V, it follows from the Sylvester equation (6.20) and Corollary 3.5 that

$$Y = \dot{P}_0 = \frac{2}{\pi} \cdot \begin{pmatrix} 0 & X^* \\ X & 0 \end{pmatrix}\,,$$

with

$$\|\dot{P}_0\| = \frac{\pi}{2} = \frac{\pi}{2}\|V\| = \frac{\pi}{2}\|\dot{B}_0\|\,,$$

cf. the proof of Lemma 6.11. Hence, in this case, the bound (6.8) on the derivative $\dot{P}_t$ is sharp at least for $t = 0$. However, so far nothing is known on the accuracy of the bound for $t > 0$.

Nevertheless, motivated by Remark 6.12, we can truncate the above example to the finite-dimensional case and use numerical calculations to get an impression of the accuracy of the bound (6.8) and the resulting estimate (6.18) in this case. To this end, let $Q_n := \mathsf{E}_A([1, 2n])$, $n \in \mathbb{N}$, be the spectral projection for A associated with the interval $[1,2n]$. Define

$$A^{(n)} := A|_{\operatorname{Ran} Q_n} \quad \text{and} \quad V^{(n)} := Q_n V Q_n|_{\operatorname{Ran} Q_n}\,,$$

and set $B_t^{(n)} := A^{(n)} + tV^{(n)}$ for $t \in [0,1]$. Furthermore, let

$$\sigma^{(n)} := \{2j \mid j = 1,\dots,n\} \quad \text{and} \quad \Sigma^{(n)} := \{2j-1 \mid j = 1,\dots,n\}\,,$$

so that

$$\operatorname{spec}(A^{(n)}) = \sigma^{(n)} \cup \Sigma^{(n)} \quad \text{with} \quad \operatorname{dist}(\sigma^{(n)}, \Sigma^{(n)}) = 1\,.$$

Numerical calculations suggest that the spectrum of each $B_t^{(n)}$ is separated as

$$\operatorname{spec}(B_t^{(n)}) = \omega_t^{(n)} \cup \Omega_t^{(n)}\,,$$

where $\omega_t^{(n)} \subset \overline{\mathcal{O}_{t/2}(\sigma^{(n)})}$, $\Omega_t^{(n)} \subset \overline{\mathcal{O}_{t/2}(\Sigma^{(n)})}$, and

$$\operatorname{dist}(\omega_t^{(n)}, \Omega_t^{(n)}) \geq 1\,. \tag{6.22}$$

Remark 6.18. One can show that the spectrum of each $B_t^{(n)}$ is symmetrically distributed around $n+1/2$ and that the sum of two opposite eigenvalues equals $2n+1$. Moreover, numerical calculations suggest that the eigenvalues of $B_t^{(n)}$ corresponding to $1,\dots,n$ are shifted to the left as t increases, whereas the ones corresponding to $n+1,\dots,2n$ are shifted to the right. However, for large n, the perturbation of most of the eigenvalues seems to be almost negligible, so that $\operatorname{dist}(\omega_t^{(n)}, \Omega_t^{(n)})$ is close to 1.

Based on inequality (6.22), it follows from estimate (6.18) that

$$\arcsin\bigl(\|\mathsf{E}_{A^{(n)}}(\sigma^{(n)}) - \mathsf{E}_{A^{(n)}+tV^{(n)}}\bigl(\mathcal{O}_{1/2}(\sigma^{(n)})\bigr)\|\bigr) \leq \frac{\pi}{2}\|V^{(n)}\|t \leq \frac{\pi}{2}t \tag{6.23}$$

for $0 \leq t \leq 1$, where we taken into account that $\|V^{(n)}\| \leq \|V\| = 1$. Further numerical calculations suggest that the left-hand side of (6.23) gets close to $\pi t/2$ uniformly in t as n gets large, which would mean that the constant $\pi/2$ in estimate (6.18) is optimal.

The author's guess is that the non-truncated example with A and V as in (6.21) even yields equality in (6.18). However, at this moment, this is pure speculation, so that the author contents himself with the following educated guess summarizing the preceding considerations.

Conjecture 6.19. *Let $\varepsilon > 0$ be arbitrary. Then, there is a self-adjoint operator A on some Hilbert space $\mathcal{H}$ with its spectrum separated as*

$$\operatorname{spec}(A) = \sigma \cup \Sigma\,, \quad \operatorname{dist}(\sigma, \Sigma) = 1\,,$$

and a bounded self-adjoint operator $V \in \mathcal{L}(\mathcal{H})$, $\|V\| \leq 1$, such that for each $t \in [0,1]$ the following holds:

(a) The spectrum of $A + tV$ is separated as

$$\operatorname{spec}(A + tV) = \omega_t \cup \Omega_t\,,$$

where $\omega_t \subset \overline{\mathcal{O}_{t/2}(\sigma)}$, $\Omega_t \subset \overline{\mathcal{O}_{t/2}(\Sigma)}$, and

$$\operatorname{dist}(\omega_t, \Omega_t) \geq 1\,.$$

(b) One has

$$\Bigl(\frac{\pi}{2} - \varepsilon\Bigr)t \leq \arcsin\bigl(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+tV}\bigl(\mathcal{O}_{1/2}(\sigma)\bigr)\|\bigr) \leq \frac{\pi}{2}t\,.$$

In particular, the constant $\pi/2$ in (6.18) *is optimal.*

Based in Conjecture 6.19, we have the following two closing observations:

First, the maximal angle for the rotation of the spectral subspaces can be near $\pi/2$ and, at the same time, the gap in the spectrum of the perturbed operator does not shrink. In fact, allowing parameters $t > 1$, one can observe numerically in the above examples that it may happen that the gap size in the spectrum of $A + tV$ is still at least 1, whereas the maximal angle between the corresponding spectral subspaces *equals* $\pi/2$. This behaviour is rather unexpected and seems to be distinctive for the case of generic spectral disposition; a similar behaviour cannot be observed in the particular case of subordinated or annular separated spectral components, cf. [30, Theorem 2]. However, this does not contradict the conjectures on the optimal constants c_{opt} and $c_{\text{opt-off}}$ mentioned in Chapter 2 since in the examples above the norm of the perturbation exceeds $\sqrt{3}/2$ considerably when the maximal angle gets close to $\pi/2$.

Second, the rotation of the spectral subspaces can be strong when at the same time the perturbation of the spectrum is rather weak. Conversely, the following example suggests that the rotation of the subspaces tends to be rather weak if the perturbation of the spectrum is strong:

Let A, V, σ, and Σ as in Example 1.22. Set $d := \operatorname{dist}(\sigma, \Sigma) = 1$, and suppose that $0 < \|V\| = \alpha < \sqrt{3}/2$. In this case, one can easily verify that

$$\operatorname{Ran} \mathsf{E}_{A+V}\bigl(\mathcal{O}_{1/2}(\sigma)\bigr) = \mathcal{G}(\operatorname{Ran} \mathsf{E}_A(\sigma), X)$$

with

$$X = \frac{1}{\|V\|} \cdot \begin{pmatrix} 0 & \delta_V \\ \delta_V & 0 \end{pmatrix}, \quad \delta_V = \|V\| \tan\Bigl(\frac{1}{2} \arctan\bigl(2\|V\|\bigr)\Bigr),$$

so that

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{1/2}(\sigma)\bigr)\|\bigr) = \arctan\|X\| < \frac{1}{2} \arctan \sqrt{3} < \frac{\pi}{4}.$$

This observation together with Conjecture 6.19 reveals a potential disadvantage in the approach to combine a posteriori type estimates on the rotation of subspaces with a priori bounds on the perturbation of the corresponding spectral components. Both type of estimates may be optimal by themselves, but it is unlikely that they are optimal at the same time. In this regard, Conjecture 6.19 deserves further studies in order to shed some light on this matter.

Chapter 7

The $\sin 2\Theta$ theorem

In the present chapter, an analogue to the Davis-Kahan $\sin 2\Theta$ theorem from [21] is proved under a general spectral separation condition. This extends the generic $\sin 2\theta$ estimates recently shown by Albeverio and Motovilov in [8]. The result is applied to the subspace perturbation problem discussed in Chapter 2. The material here is taken to a large extent from Sections 1–3 of the author's article [50] published in *Integral Equations and Operator Theory*.

7.1 Introduction and main results

The main objective in this chapter is to show the following variant of the Davis-Kahan $\sin 2\Theta$ theorem.

Theorem 7.1. *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$ such that the spectrum of A is separated into two disjoint components, that is,*

$$\operatorname{spec}(A) = \sigma \cup \Sigma \qquad \textit{with} \qquad d := \operatorname{dist}(\sigma, \Sigma) > 0\,.$$

Moreover, let V be a bounded self-adjoint operator on $\mathcal{H}$, and let Q be an orthogonal projection in $\mathcal{H}$ onto a reducing subspace for $A+V$. Then, the operator angle $\Theta = \Theta(\mathsf{E}_A(\sigma), Q)$ associated with the subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} Q$ satisfies

$$\|\sin 2\Theta\| \le \frac{\pi}{2} \cdot 2\,\frac{\|V\|}{d}\,. \tag{7.1}$$

It should be emphasized that the projection Q in Theorem 7.1 is not assumed to be a spectral projection for $A+V$. It is also worth mentioning that the bound (7.1) is of an *a priori* type since the spectral separation condition is imposed on the unperturbed operator A only. By switching the roles of A and $A+V$, one may impose the analogous condition on the perturbed operator $A+V$ instead, which results in the corresponding *a posteriori* type estimate.

Theorem 7.1 is a direct analogue of the Davis-Kahan $\sin 2\Theta$ theorem from [21]. There, it is additionally assumed that the convex hull of one of the spectral components σ and Σ is disjoint from the other component, that is, $\operatorname{conv}(\sigma)\cap\Sigma=\varnothing$ or vice versa. The corresponding estimate is the same as (7.1), except for the constant $\pi/2$ being replaced by 1, cf. estimate (2.16). Note that the Davis-Kahan $\sin 2\Theta$ theorem is formulated in [21] for arbitrary unitary-invariant norms including the standard Schatten norms. A corresponding extension of Theorem 7.1 is discussed in Section 4 of the author's article [50].

An immediate consequence of Theorem 7.1 is the *generic* $\sin 2\theta$ *estimate* recently proved by Albeverio and Motovilov in [8],

$$\sin 2\theta \le \pi\,\frac{\|V\|}{d} \quad\text{with}\quad \theta=\|\Theta\|=\arcsin\bigl(\|\mathsf{E}_A(\sigma)-Q\|\bigr)\,. \tag{7.2}$$

This is due to the elementary inequality $\sin(2\|\Theta\|)\le\|\sin 2\Theta\|$. In this respect, we may call (7.1) the *generic* $\sin 2\Theta$ *estimate*. It should be emphasized that, in contrast to (7.1), no extension of (7.2) to norms other than the usual operator norm is at hand.

Clearly, the estimates (7.1) and (7.2) provide no useful information if $\|V\|\ge d/\pi$. On the other hand, for perturbations V satisfying $\|V\|<d/\pi$, the $\sin 2\Theta$ estimate (7.1) implies that $\|\sin 2\Theta\|<1$, so that the spectrum of Θ has a gap around $\pi/4$. This means that there is an open interval containing $\pi/4$ that belongs to the resolvent set of Θ, namely

$$\Bigl(\alpha,\frac{\pi}{2}-\alpha\Bigr)\subset\Bigl[0,\frac{\pi}{2}\Bigr]\setminus\operatorname{spec}(\Theta)\quad\text{with}\quad \alpha:=\frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr)<\frac{\pi}{4}\,.$$

Note that Θ may a priori have spectrum both in $[0,\alpha]$ and $\bigl[\frac{\pi}{2}-\alpha,\frac{\pi}{2}\bigr]$. This depends on the reducing subspace for $A+V$ that is considered, see Remark 7.4 below.

In this regard, Theorem 7.1 is in general stronger than the corresponding result of the $\sin 2\theta$ estimate (7.2) since the latter provides information only on the maximal angle $\theta = \|\Theta\|$ between the subspaces $\operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\operatorname{Ran} Q$, cf. [8, Remark 4.2]. However, if it is known that $\theta \le \pi/4$, then one has $\|\sin 2\Theta\| = \sin 2\theta$, so that, in this case, both estimates agree.

As an application to the subspace perturbation problem, we obtain the following bound on the maximal angle between the corresponding spectral subspaces for the unperturbed and perturbed operators A and $A+V$, respectively. It plays an important role in the forthcoming Chapter 8.

Corollary 7.2 (cf. [8, Remark 4.4])**.** *Let A and V be as in Theorem 7.1. If $\|V\| \le d/\pi$, then*

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)\|\bigr) \le \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr) \le \frac{\pi}{4}\,. \tag{7.3}$$

The bound (7.3) in Corollary 7.2 is not optimal if $\|V\| > \frac{4d}{\pi^2+4}$, see Theorem 8.9 below and also [8]. However, for perturbations V satisfying $\|V\| \le \frac{4d}{\pi^2+4}$, this bound on the maximal angle is the strongest one available so far, cf. [8, Remark 5.5] and also Remark 8.11 below.

The present chapter is organized as follows: The proofs of Theorem 7.1 and Corollary 7.2 are given in Section 7.2. A variant of Corollary 7.2 corresponding to the original Davis-Kahan $\sin 2\Theta$ theorem is also discussed there, see Remark 7.8.

Section 7.3 is devoted to an alternative, straightforward proof of the $\sin 2\theta$ estimate (7.2) that is not based on Theorem 7.1 and is more direct than the one by Albeverio and Motovilov in [8].

7.2 Proof of Theorem 7.1 and Corollary 7.2

We start with the following result, which has already played a crucial role in the proof of the original Davis-Kahan $\sin 2\Theta$ theorem in [21]. It is one of the key ingredients for our proof of Theorem 7.1 as well.

Lemma 7.3 (cf. [21, Section 7])**.** *Let P and Q be two orthogonal projections in a Hilbert space $\mathcal{H}$, and denote $K := Q - Q^\perp$. Then*

$$\sin\bigl(2\Theta(P,Q)\bigr) = \sin\bigl(\Theta(P,KPK)\bigr)\,.$$

Proof. For the sake of completeness, we give a proof in the current notations.

In view of (1.10), one computes

$$\begin{aligned}\sin^2\bigl(2\Theta(P,Q)\bigr) &= 4S(P,Q)C(P,Q)\\ &= 4\bigl(PQ^\perp P + P^\perp QP^\perp\bigr)\bigl(PQP + P^\perp Q^\perp P^\perp\bigr)\\ &= 4PQ^\perp PQP + 4P^\perp QP^\perp Q^\perp P^\perp\,.\end{aligned}\tag{7.4}$$

Denote $R := KPK$. Clearly, R is again an orthogonal projection in $\mathcal{H}$ since K is self-adjoint and unitary. Using $K = I_{\mathcal{H}} - 2Q^\perp = 2Q - I_{\mathcal{H}}$, one observes that

$$\begin{aligned}4PQ^\perp PQP = -4PQ^\perp P^\perp QP &= P\bigl(I_{\mathcal{H}} - 2Q^\perp\bigr)P^\perp\bigl(2Q - I_{\mathcal{H}}\bigr)P\\ &= PKP^\perp KP = PR^\perp P\end{aligned}\tag{7.5}$$

and, similarly, that

$$4P^\perp QP^\perp Q^\perp P^\perp = P^\perp RP^\perp\,.\tag{7.6}$$

Combining (7.4)–(7.6) yields

$$\sin^2\bigl(2\Theta(P,Q)\bigr) = PR^\perp P + P^\perp RP^\perp = S(P,R) = \sin^2\bigl(\Theta(P,R)\bigr)\,,$$

which proves the claim by taking the square roots. □

The preceding lemma can now be used to deduce the generic $\sin 2\Theta$ estimate from the symmetric $\sin\Theta$ theorem discussed in Section 3.2.

Proof of Theorem 7.1. In essence, we follow the proof in [21, Section 7].

As in Lemma 7.3, let K denote the self-adjoint unitary operator on $\mathcal{H}$ given by

$$K := Q - Q^\perp\,.$$

Since $\operatorname{Ran} Q$ is reducing for $A+V$, the splitting property

$$\operatorname{Dom}(A+V) = \bigl(\operatorname{Dom}(A+V)\cap\operatorname{Ran}Q\bigr) + \bigl(\operatorname{Dom}(A+V)\cap\operatorname{Ran}Q^\perp\bigr)\tag{7.7}$$

implies that K maps $\operatorname{Dom}(A+V) = \operatorname{Dom}(A)$ onto itself. It also follows from (7.7) and the invariance of the subspaces $\operatorname{Ran} Q$ and $\operatorname{Ran} Q^\perp$ that one has $K(A+V)Kx = (A+V)x$ for $x \in \operatorname{Dom}(A)$, so that $K(A+V)K = A+V$.

The operator

$$D := KAK \quad \text{on} \quad \operatorname{Dom}(D) := \operatorname{Dom}(A)$$

is therefore self-adjoint and satisfies

$$D = K(A+V)K - KVK = A + V - KVK\,. \tag{7.8}$$

Clearly, the spectra of A and D coincide, that is,

$$\operatorname{spec}(D) = \operatorname{spec}(A) = \sigma \cup \Sigma\,.$$

In particular, one has

$$\mathsf{E}_D(\sigma) = K\mathsf{E}_A(\sigma)K \quad \text{and} \quad \mathsf{E}_D(\Sigma) = K\mathsf{E}_A(\Sigma)K\,. \tag{7.9}$$

Considering D by (7.8) as a perturbation of A, and taking into account that $\operatorname{dist}(\sigma,\Sigma) = d > 0$, it now follows from the symmetric $\sin\Theta$ theorem (Proposition 3.7) that

$$\bigl\|\sin\bigl(\Theta(\mathsf{E}_A(\sigma),\mathsf{E}_D(\sigma))\bigr)\bigr\| \le \frac{\pi}{2}\,\frac{\|V - KVK\|}{d} \le \frac{\pi}{2}\cdot 2\,\frac{\|V\|}{d}\,,$$

where the last inequality is due to the fact that $\|KVK\| = \|V\|$ since K is unitary. In view of (7.9) and Lemma 7.3, this proves the claim. □

If, in the situation of Theorem 7.1, it is known that $\theta = \|\Theta\| \le \pi/4$, then one has $\|\sin 2\Theta\| = \sin 2\theta$. In this case, taking into account (1.12), the bound (7.1) can equivalently be rewritten as

$$\theta = \arcsin\bigl(\|\mathsf{E}_A(\sigma) - Q\|\bigr) \le \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr)\,, \tag{7.10}$$

see also [8, Remark 4.2].

However, the condition $\theta \le \pi/4$ does not need to be satisfied for arbitrary reducing subspaces for $A+V$, even if the perturbation V is small in norm. In fact, although the spectrum of Θ is known to have a gap around $\pi/4$ whenever $\|V\| < d/\pi$, the following observation illustrates that the operator angle Θ may a priori have spectrum everywhere else in the interval $\bigl[0,\frac{\pi}{2}\bigr]$.

Remark 7.4. In addition to the hypotheses of Theorem 7.1, assume that $\|V\| < d/\pi$ and that $\|\Theta(P,Q)\| < \pi/4$, where $P := \mathsf{E}_A(\sigma)$. Estimate (7.10) then implies that

$$\operatorname{spec}\bigl(\Theta(P,Q)\bigr) \subset [0,\alpha] \quad \text{with} \quad \alpha := \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4}\,. \tag{7.11}$$

Taking into account that $S(P,Q^\perp) = C(P,Q)$, one has the identity $\sin\bigl(\Theta(P,Q^\perp)\bigr) = \cos\bigl(\Theta(P,Q)\bigr)$. It therefore follows from (7.11) that

$$\operatorname{spec}\bigl(\Theta(P,Q^\perp)\bigr) \subset \Bigl[\frac{\pi}{2} - \alpha, \frac{\pi}{2}\Bigr]\,. \tag{7.12}$$

Now, suppose that R is an orthogonal projection onto a reducing subspace for $A+V$ such that

$$\operatorname{Ran} R \cap \operatorname{Ran} Q \neq \{0\} \neq \operatorname{Ran} R \cap \operatorname{Ran} Q^\perp\,.$$

Let $x \in \operatorname{Ran} R \cap \operatorname{Ran} Q$ with $\|x\| = 1$. Using the identity $(P-R)x = (P-Q)x$ and the inclusion (7.11), one observes that

$$\begin{aligned}\langle x, \sin^2\bigl(\Theta(P,R)\bigr)x\rangle &= \langle x, (P-R)^2x\rangle = \langle x, (P-Q)^2x\rangle \\ &= \langle x, \sin^2\bigl(\Theta(P,Q)\bigr)x\rangle \le \sin^2\alpha\,.\end{aligned} \tag{7.13}$$

Taking into account (7.12), for $y \in \operatorname{Ran} R \cap \operatorname{Ran} Q^\perp$, $\|y\| = 1$, one obtains in a similar way that

$$\langle y, \sin^2\bigl(\Theta(P,R)\bigr)y\rangle = \langle y, \sin^2\bigl(\Theta(P,Q^\perp)\bigr)y\rangle \ge \sin^2\Bigl(\frac{\pi}{2} - \alpha\Bigr)\,. \tag{7.14}$$

Combining (7.13) and (7.14) yields that $\Theta(P,R)$ has spectrum both in $[0,\alpha]$ and $\bigl[\frac{\pi}{2} - \alpha, \frac{\pi}{2}\bigr]$.

Thus, depending on the reducing subspace for $A+V$ that is considered, the operator angle has spectrum in $[0,\alpha]$, $\bigl[\frac{\pi}{2} - \alpha, \frac{\pi}{2}\bigr]$, or both.

In the situation of Corollary 7.2, the projection Q is chosen very specifically, namely $Q = \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$. It turns out that, in this case, the condition $\theta \le \pi/4$ is automatically satisfied whenever $\|V\| \le d/\pi$. Indeed, the mapping $[0,1] \ni t \mapsto \mathsf{E}_{A+tV}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ is continuous in norm for $\|V\| < d/2$ (see Section 6.2 and also [8, Theorem 3.5]), so that Corollary 7.2 is a direct consequence of the following more general statement.

Lemma 7.5 (cf. [21, Theorem 8.2]). *Let A, V, and Q be as in Theorem 7.1, and suppose that V satisfies $\|V\| \le d/\pi$. If there is a norm continuous path $[0,1] \ni t \mapsto P_t$ of orthogonal projections in $\mathcal{H}$ with $P_0 = \mathsf{E}_A(\sigma)$ and $P_1 = Q$ such that* $\operatorname{Ran} P_t$ *is reducing for $A + tV$ for all $t \in [0,1]$, then*

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - Q\|\bigr) \le \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr) \le \frac{\pi}{4}\,.$$

Proof. In view of Theorem 7.1 (or more precisely, estimate (7.10)), it suffices to show the inequality

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - Q\|\bigr) \le \frac{\pi}{4}\,. \tag{7.15}$$

Assume that (7.15) does not hold. Then, since the path $[0,1] \ni t \mapsto P_t$ is assumed to be norm continuous with $P_0 = \mathsf{E}_A(\sigma)$ and $P_1 = Q$, there is $\tau \in (0,1)$ such that

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - P_\tau\|\bigr) = \frac{\pi}{4}\,. \tag{7.16}$$

On the other hand, taking into account that $\operatorname{Ran} P_\tau$ is reducing for $A + \tau V$ and that $\tau\|V\| < d/\pi$, it follows from inequality (7.10) that

$$\arcsin\bigl(\|\mathsf{E}_A(\sigma) - P_\tau\|\bigr) \le \frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|\tau V\|}{d}\Bigr) < \frac{\pi}{4}\,,$$

which is a contradiction to (7.16). This shows inequality (7.15). □

Remark 7.6. The bound (7.3) from Corollary 7.2 has already been mentioned in [8, Remark 4.4], but only for the particular case of perturbations V satisfying $\|V\| \le \frac{\mathrm{e}-1}{2\mathrm{e}}\,d$, where $\frac{4}{\pi^2+4} < \frac{\mathrm{e}-1}{2\mathrm{e}} < \frac{1}{\pi}$. There, the condition $\theta \le \pi/4$ has been ensured by use of the bound from Theorem 6.15 (a).

Remark 7.7 (cf. [8, Remark 5.5]). The bound from Corollary 7.2 is stronger than the one from Theorem 6.15 (a). More precisely, one has

$$\frac{1}{2}\arcsin\Bigl(\pi\,\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4}\log\Bigl(\frac{d}{d - 2\|V\|}\Bigr) \quad \text{whenever} \quad 0 < \|V\| \le \frac{d}{\pi}\,.$$

Indeed, it is straightforward to verify that the difference

$$\frac{\pi}{4}\log\Bigl(\frac{1}{1-2x}\Bigr) - \frac{1}{2}\arcsin(\pi x)\,, \quad 0 \le x \le \frac{1}{\pi}\,,$$

attains its positive maximum at $x = \frac{4}{\pi^2+4}$ and its (unique) minimum at $x = 0$.

We close this section with a discussion of Theorem 7.1 and Corollary 7.2 under the additional spectral separation conditions from [21].

Remark 7.8. In addition to the hypotheses of Theorem 7.1, assume that the convex hull of one of the sets σ and Σ is disjoint from the other set. In this case, the constant $\pi/2$ in the bound (7.1) can be replaced by 1, see Remark 3.8. The resulting estimate is the bound from the Davis-Kahan $\sin 2\Theta$ theorem in [21], that is,

$$\|\sin 2\Theta\| \le 2\,\frac{\|V\|}{d}\,.$$

For the particular case of $Q = \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$, as in Corollary 7.2 (see also [21, Theorem 8.2]) this bound can equivalently be rewritten as

$$\arcsin\Bigl(\bigl\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)\bigr\|\Bigr) \le \frac{1}{2}\arcsin\Bigl(2\,\frac{\|V\|}{d}\Bigr) < \frac{\pi}{4} \tag{7.17}$$

whenever $\|V\| < d/2$. It has already been stated by Davis in [20, Theorem 5.1] that this estimate is sharp in the sense that equality can be attained. This can be seen from the following example of 2×2 matrices: Let

$$A := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{with} \quad \sigma := \{1\} \quad \text{and} \quad \Sigma := \{-1\}\,.$$

Obviously, one has $d := \operatorname{dist}(\sigma, \Sigma) = 2$. For arbitrary x with $0 < x < 1 = d/2$ consider

$$V := \begin{pmatrix} -x^2 & x\sqrt{1-x^2} \\ x\sqrt{1-x^2} & x^2 \end{pmatrix}.$$

It is easy to verify that $\|V\| = x$ and that $\operatorname{spec}(A+V) = \{\pm\sqrt{1-x^2}\}$.

Denote $\alpha := \frac{1}{2}\arcsin(x) < \frac{\pi}{4}$. Then, one has

$$\frac{1-\sqrt{1-x^2}}{x} = \frac{1-\cos(2\alpha)}{\sin(2\alpha)} = \tan\alpha \quad \text{and} \quad \frac{1+\sqrt{1-x^2}}{x} = \cot\alpha\,. \tag{7.18}$$

Using (7.18), a straightforward computation shows that

$$U^*(A+V)U = \begin{pmatrix} \sqrt{1-x^2} & 0 \\ 0 & -\sqrt{1-x^2} \end{pmatrix} \quad \text{where} \quad U = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

In particular, this implies that

$$\mathsf{E}_{A+V}\big(\mathcal{O}_1(\sigma)\big) = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \end{pmatrix} = \begin{pmatrix} \cos^2\alpha & \sin\alpha\cos\alpha \\ \sin\alpha\cos\alpha & \sin^2\alpha \end{pmatrix},$$

so that $\|\Theta\| = \alpha$ and, therefore,

$$\arcsin\big(\|\mathsf{E}_A(\sigma) - \mathsf{E}_{A+V}\big(\mathcal{O}_1(\sigma)\big)\|\big) = \frac{1}{2}\arcsin(x) = \frac{1}{2}\arcsin\Big(2\,\frac{\|V\|}{d}\Big).$$

Hence, inequality (7.17) is sharp.

7.3 The generic $\sin 2\theta$ estimate

In this section, we present an alternative, straightforward proof of the generic $\sin 2\theta$ estimate (7.2) that uses a different technique than the one presented for Theorem 7.1 and, at the same time, is more direct than the one in [8].

It is worth mentioning that inequality (7.10), and therefore also Corollary 7.2, can be deduced from estimate (7.2) as well since $\|\sin 2\Theta\| = \sin 2\theta$ whenever $\theta = \|\Theta\| \leq \pi/4$. An immediate advantage of the $\sin 2\theta$ estimate is that it can be formulated without the notion of the operator angle, see Proposition 7.9 below.

In contrast to the proof of the a priori $\sin 2\theta$ estimate (7.2) presented in [8], the one given below is direct and is not deduced from the corresponding a posteriori estimate. In addition, the key idea of the argument presented here can easily be reduced to one single equation, namely equation (7.23) below, which makes this proof very transparent.

Proposition 7.9 ([8, Corollary 4.3])**.** *Let A, V, and Q be as in Theorem 7.1. Then*

$$\sin 2\theta \leq \pi\,\frac{\|V\|}{d},$$

where $\theta = \arcsin\big(\|\mathsf{E}_A(\sigma) - Q\|\big)$ is the maximal angle between the subspaces $\operatorname{Ran}\mathsf{E}_A(\sigma)$ *and* $\operatorname{Ran} Q$.

Proof. The case $\theta = \pi/2$ is obvious. Assume that $\theta < \pi/2$, that is,

$$\|\mathsf{E}_A(\sigma) - Q\| < 1. \tag{7.19}$$

Denote $\mathcal{H}_0 := \operatorname{Ran} \mathsf{E}_A(\sigma)$ and $\mathcal{H}_1 := \mathcal{H}_0^\perp = \operatorname{Ran} \mathsf{E}_A(\Sigma)$, and let

$$V = \begin{pmatrix} V_0 & W \\ W^* & V_1 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}$$

with $\operatorname{Dom}(A) = \operatorname{Dom}(A_0) \oplus \operatorname{Dom}(A_1)$ be the representations of V and A as 2×2 block operator matrices with respect to the decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$.

In view of inequality (7.19), Proposition 1.13 implies that there is a unique operator $X \in \mathcal{L}(\mathcal{H}_0, \mathcal{H}_1)$ such that $\operatorname{Ran} Q = \mathcal{G}(\mathcal{H}_0, X)$. This operator X satisfies

$$\arctan(\|X\|) = \arcsin\bigl(\|\mathsf{E}_A(\sigma) - Q\|\bigr) = \theta\,. \tag{7.20}$$

Moreover, taking into account the identities $\operatorname{Dom}(A_0 + V_0) = \operatorname{Dom}(A_0)$, $\operatorname{Dom}(A_1 + V_1) = \operatorname{Dom}(A_1)$, and

$$A + V = \begin{pmatrix} A_0 + V_0 & 0 \\ 0 & A_1 + V_1 \end{pmatrix} + \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix},$$

it follows from Corollary 4.9 that X is a strong solution to the operator Riccati equation $X(A_0+V_0) - (A_1+V_1)X + XWX - W^* = 0$, that is,

$$\operatorname{Ran}\bigl(X|_{\operatorname{Dom}(A_0)}\bigr) \subset \operatorname{Dom}(A_1)$$

and

$$X(A_0 + V_0)g - (A_1 + V_1)Xg + XWXg - W^*g = 0 \tag{7.21}$$

for $g \in \operatorname{Dom}(A_0)$.

Defining $T \in \mathcal{L}(\mathcal{H})$ by

$$T := \begin{pmatrix} I_{\mathcal{H}_0} & -X^* \\ X & I_{\mathcal{H}_1} \end{pmatrix},$$

a straightforward calculation shows that

$$T^*VT = \begin{pmatrix} * & * \\ V_1X - XV_0 - XWX + W^* & * \end{pmatrix}. \tag{7.22}$$

Denote $P := \mathsf{E}_A(\sigma)$. Equations (7.21) and (7.22) then imply that

$$\begin{aligned} XA_0g - A_1Xg &= V_1Xg - XV_0g - XWXg + W^*g \\ &= \bigl(P^\perp T^*VTP|_{\mathcal{H}_0}\bigr)g \end{aligned} \tag{7.23}$$

for $g \in \mathrm{Dom}(A_0)$, where the restriction $P^\perp T^*VTP|_{\mathcal{H}_0}$ is understood as an operator from $\mathcal{H}_0$ to $\mathcal{H}_1$. Comparing equation (7.23) with the Sylvester equation (3.2), it follows from the bound in Theorem 3.2 given by (3.6) and (3.7) that

$$\|X\| \le \frac{\pi}{2}\,\frac{\|P^\perp T^*VTP\|}{d} \le \frac{\pi}{2}\bigl(1+\|X\|^2\bigr)\frac{\|V\|}{d}\,,$$

where we have taken into account that $\|T\| = \|T^*\| = \sqrt{1+\|X\|^2}$, cf. equation (1.4). Since $2\|X\|/(1+\|X\|^2) = 2\tan\theta/(1+\tan^2\theta) = \sin(2\theta)$ by (7.20), this proves the claim. □

Chapter 8

An optimization problem

In this chapter, we discuss a way to optimize the approach of iterating the estimate on the maximal angle thus strengthening the results from Theorem 6.15. The corresponding material in Sections 8.1 and 8.2 is taken with only minor changes from the author's article [51], whereas the material in Sections 8.3 and 8.4 is new.

The main focus of this chapter is on general perturbations, which are discussed in Sections 8.1 and 8.2 below. There, a constrained optimization problem is formulated, whose solution provides an estimate on the maximal angle between the corresponding spectral subspaces, see Definition 8.6 and Proposition 8.7 below. The explicit solution to this problem is given in Proposition 8.8, which leads to the main result of this chapter, Theorem 8.9. The proof of Proposition 8.8 is provided in the separate Section 8.2. The technique used there involves variational methods and may also be useful for solving optimization problems of a similar structure. Problems of this sort appear, for instance, when off-diagonal or semidefinite perturbations are considered. These particular cases are discussed in Sections 8.3 and 8.4, respectively.

8.1 Formulation of the optimization problem

For notational setup, we fix the following assumptions.

Hypothesis 8.1. *Let A be as in Hypothesis 6.14, and let* $V \in \mathcal{L}(\mathcal{H})$, $V \neq 0$, *be self-adjoint. For $0 \le t < \frac{1}{2}$, introduce*

$$B_t := A + td\,\frac{V}{\|V\|}\,, \quad \operatorname{Dom}(B_t) := \operatorname{Dom}(A)\,,$$

and let $P_t := \mathsf{E}_{B_t}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ denote the spectral projection for B_t associated with the open $\frac{d}{2}$-neighbourhood $\mathcal{O}_{d/2}(\sigma)$ of σ.

Under Hypothesis 8.1, one has $\|B_t - A\| = td < \frac{d}{2}$ for $0 \le t < \frac{1}{2}$. Hence, it follows from Corollary 1.18 that the spectrum of each B_t is likewise separated into two disjoint components, that is,

$$\operatorname{spec}(B_t) = \omega_t \cup \Omega_t \quad \text{for} \quad 0 \le t < \frac{1}{2}\,,$$

where

$$\omega_t = \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{td}(\sigma)} \quad \text{and} \quad \Omega_t = \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{td}(\Sigma)}\,,$$

cf. Section 6.2. In particular, one has

$$\delta_t := \operatorname{dist}(\omega_t, \Omega_t) \ge (1-2t)d > 0 \quad \text{for} \quad 0 \le t < \frac{1}{2}\,. \tag{8.1}$$

Moreover, the path $\bigl[0, \frac{1}{2}\bigr) \ni t \mapsto P_t$ is continuous in norm, see, e.g., Lemma 6.7 and also [8, Theorem 3.5]; recall that by Lemma 6.3 (b) the spectral components ω_t and Ω_t depend upper semicontinuously on the parameter.

Let $t \in \bigl(0, \frac{1}{2}\bigr)$ be arbitrary, and let $0 = t_0 < t_1 < \cdots < t_{n+1} = t$ with $n \in \mathbb{N}_0$ be a finite partition of the interval $[0, t]$. Using the triangle inequality for the maximal angle (Proposition 5.8), we obtain

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le \sum_{j=0}^{n} \arcsin\bigl(\|P_{t_j} - P_{t_{j+1}}\|\bigr)\,. \tag{8.2}$$

Clearly, we can consider $B_{t_{j+1}} = B_{t_j} + (t_{j+1} - t_j)d \cdot V/\|V\|$ as a perturbation of B_{t_j}. Taking into account the a priori bound (8.1), we observe that

$$\frac{\|B_{t_{j+1}} - B_{t_j}\|}{\operatorname{dist}(\omega_{t_j}, \Omega_{t_j})} \le \frac{t_{j+1} - t_j}{1 - 2t_j} =: \lambda_j < \frac{1}{2}\,, \quad j = 0, \dots, n\,. \tag{8.3}$$

In particular, it follows from Corollary 1.18 that $\omega_{t_{j+1}}$ is exactly the part of

$\operatorname{spec}(B_{t_{j+1}})$ that is contained in the open $\delta_{t_j}/2$-neighbourhood of ω_{t_j}, that is,

$$\omega_{t_{j+1}} = \operatorname{spec}(B_{t_{j+1}}) \cap \mathcal{O}_{\delta_{t_j}/2}(\omega_{t_j})\,, \quad j = 0, \dots, n\,.$$

Thus, each summand of the right-hand side of (8.2) can be treated in the same way as the maximal angle in the general situation discussed in Section 2.3. For example, combining (8.3) with the bound (2.22) derived from the symmetric $\sin\Theta$ theorem, one gets for each j that

$$\arcsin\bigl(\|P_{t_j} - P_{t_{j+1}}\|\bigr) \le \arcsin\Bigl(\frac{\pi}{2}\,\frac{\lambda_j}{1-\lambda_j}\Bigr)\,, \tag{8.4}$$

provided that $\lambda_j \le 2/(2+\pi)$. If partitions of the interval $[0,t]$ with arbitrarily small mesh size are considered, this once more leads to the bound (2.26) obtained in Theorem 6.15 (a):

Remark 8.2. Clearly, one has

$$\frac{\lambda_j}{1-\lambda_j} = \frac{t_{j+1}-t_j}{1-t_j-t_{j+1}} \le \frac{t_{j+1}-t_j}{1-2t_{j+1}} \le \frac{t_{j+1}-t_j}{1-2t}\,.$$

Hence, if the mesh size of the partition of the interval $[0,t]$ is sufficiently small, then for each j the right-hand side of (8.4) is small as well. At the same time, the corresponding Riemann sum

$$\sum_{j=0}^{n} \frac{t_{j+1}-t_j}{1-2t_{j+1}}$$

is close to the integral $\int_0^t \frac{1}{1-2\tau}\,\mathrm{d}\tau$. Considering partitions with arbitrarily small mesh size and taking into account that $\arcsin(x)/x \to 1$ as $x \to 0$, one then concludes from (8.2) and (8.4) that

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le \frac{\pi}{2}\int_0^t \frac{1}{1-2\tau}\,\mathrm{d}\tau = \frac{\pi}{4}\log\Bigl(\frac{1}{1-2t}\Bigr)\,, \tag{8.5}$$

which agrees with the bound from Theorem 6.15 (a). A similar, yet more technical, argument can also be used to prove the bound from Theorem 6.15 (b) for off-diagonal perturbations and even the general result from Theorem 6.10. This line of reasoning does not require the smoothness of the path $\tau \mapsto P_\tau$, but it is less elegant than the one presented in Chapter 6 above.

Albeverio and Motovilov demonstrated in [8] that a result stronger than (8.5) can be obtained from (8.2). They considered a specific finite partition of the interval $[0,t]$ and used the a priori generic $\sin 2\theta$ estimate (more precisely estimate (7.3)) to bound each summand of the corresponding right-hand side of (8.2). We follow this approach here. To this end, we require that the given partition of the interval $[0,t]$ additionally satisfies

$$\lambda_j = \frac{t_{j+1}-t_j}{1-2t_j} \le \frac{1}{\pi}\,, \quad j=0,\dots,n\,. \tag{8.6}$$

In this case, it follows from (8.2), (8.3), and Corollary 7.2 that

$$\arcsin\bigl(\|P_0-P_t\|\bigr) \le \frac{1}{2}\sum_{j=0}^{n}\arcsin(\pi\lambda_j)\,. \tag{8.7}$$

Note that the identity $1-2\tau = 1-2t_j-2(\tau-t_j)$ for $t_j\le\tau\le t_{j+1}$ guarantees that

$$\int_{t_j}^{t_{j+1}}\frac{\mathrm{d}\tau}{1-2\tau} = \int_0^{\lambda_j}\frac{\mathrm{d}\tau}{1-2\tau}\,, \quad j=0,\dots,n\,. \tag{8.8}$$

Therefore, taking into account Remark 7.7, the bound from Corollary 7.2 is for *each* summand of the right-hand side of (8.2) more accurate than the one from Theorem 6.15 (a). This justifies the approach (8.7).

Along with a specific choice of the partition of the interval $[0,t]$, estimate (8.7) is the essence of the approach by Albeverio and Motovilov in [8]. For future reference, we recall their choice of the partition in the following remark.

Remark 8.3. Let $0<t\le c_*<1/2$ be arbitrary with c_* as in (2.28). We distinguish between three cases for t:

If $t\le\frac{4}{\pi^2+4}$, then choose the trivial partition $0=t_0<t_1=t$, so that

$$\lambda_0 = \frac{t-t_0}{1-2t_0} = t\,.$$

If $\frac{4}{\pi^2+4}<t\le\frac{8\pi^2}{(\pi^2+4)^2}$, then consider the partition $0=t_0<t_1<t_2=t$ with $t_1=\frac{4}{\pi^2+4}$. In this case, one has

$$\lambda_0 = \frac{4}{\pi^2+4} \quad\text{and}\quad \lambda_1 = \frac{t-t_1}{1-2t_1} = \frac{(\pi^2+4)t-4}{\pi^2-4} \le \frac{4}{\pi^2+4}\,.$$

Finally, for $\frac{8\pi^2}{(\pi^2+4)^2} < t \le c_*$ consider $0 = t_0 < t_1 < t_2 < t_3 = t$ with $t_1 = \frac{4}{\pi^2+4}$ and $t_2 = \frac{8\pi^2}{(\pi^2+4)^2}$. Then,

$$\lambda_0 = \lambda_1 = \frac{4}{\pi^2+4} \quad \text{and} \quad \lambda_2 = \frac{t-t_2}{1-2t_2} = \frac{(\pi^2+4)^2 t - 8\pi^2}{(\pi^2-4)^2} < \frac{4}{\pi^2+4}\,.$$

In each of these cases, it is easy to verify that the corresponding right-hand side of (8.7) agrees with $M_*(t)$, where $M_*\colon [0,c_*] \to \big[0,\frac{\pi}{2}\big]$ is the function from (2.27).

We now optimize the choice of the partition of the interval $[0,t]$ such that for every fixed parameter t the right-hand side of inequality (8.7) is minimized. An equivalent and more convenient reformulation of this approach is to maximize the parameter t in estimate (8.7) over all possible choices of the parameters n and λ_j for which the right-hand side of (8.7) takes a fixed value.

Obviously, we can generalize estimate (8.7) to the case where the finite sequence $(t_j)_{j=1}^n$ is allowed to be just increasing and not necessarily strictly increasing. Altogether, this motivates the following considerations.

Definition 8.4. For $n \in \mathbb{N}_0$ let D_n denote the set of sequences $(\lambda_j)_{j\in\mathbb{N}_0}$ satisfying

$$0 \le \lambda_j \le \frac{1}{\pi} \quad \text{for } j \le n \qquad \text{and} \qquad \lambda_j = 0 \quad \text{for } j \ge n+1\,,$$

and set $D := \bigcup_{n\in\mathbb{N}_0} D_n$.

Every finite partition of the interval $[0,t]$ that satisfies condition (8.6) is related to a sequence in D in the obvious way. Conversely, the following lemma allows to regain the finite partition of the interval $[0,t]$ from this sequence.

Lemma 8.5.

(a) For $0 \le x < \frac{1}{2}$ the mapping $\big[0,\frac{1}{2}\big] \ni \tau \mapsto \tau + x(1-2\tau)$ is strictly increasing.

(b) For every $\lambda = (\lambda_j) \in D$ the sequence $(t_j) \subset \mathbb{R}$ given by the recursion

$$t_{j+1} = t_j + \lambda_j(1-2t_j)\,, \quad j \in \mathbb{N}_0\,, \quad t_0 = 0\,, \tag{8.9}$$

is increasing and satisfies $0 \le t_j < 1/2$ *for all* $j \in \mathbb{N}_0$*. Moreover, one has* $t_j = t_{n+1}$ *for* $j \ge n+1$ *if* $\lambda \in D_n$*. In particular,* (t_j) *is eventually constant.*

Proof. The proof of claim (a) is straightforward and is hence omitted.

For the proof of (b), let $\lambda = (\lambda_j) \in D$ be arbitrary and let $(t_j) \subset \mathbb{R}$ be given by (8.9). Observe that $t_0 = 0 < 1/2$ and that (a) implies that

$$0 \le t_{j+1} = t_j + \lambda_j(1-2t_j) < \frac{1}{2} + \lambda_j\Bigl(1 - 2\cdot\frac{1}{2}\Bigr) = \frac{1}{2} \quad \text{if} \quad 0 \le t_j < \frac{1}{2}\,.$$

Thus, the two-sided estimate $0 \le t_j < 1/2$ holds for all $j \in \mathbb{N}_0$ by induction. In particular, it follows that $t_{j+1} - t_j = \lambda_j(1-2t_j) \ge 0$ for all $j \in \mathbb{N}_0$, so that the sequence (t_j) is increasing. Let $n \in \mathbb{N}_0$ such that $\lambda \in D_n$. Since $\lambda_j = 0$ for $j \ge n+1$, it follows from the definition of (t_j) that $t_{j+1} = t_j$ for $j \ge n+1$, that is, $t_j = t_{n+1}$ for $j \ge n+1$. This completes the proof. □

It follows from part (b) of the preceding lemma that for every $\lambda \in D$ the sequence (t_j) given by (8.9) yields a finite partition of the interval $[0, \tilde{t}]$ with $\tilde{t} := \max_{j\in\mathbb{N}_0} t_j < 1/2$. In this respect, the approach to optimize the parameter t in (8.7) with a fixed right-hand side can now be formalized in the following way.

Definition 8.6. Let W denote the (non-linear) operator on D that assigns $\lambda = (\lambda_j) \in D$ the corresponding increasing and eventually constant sequence given by the recursion (8.9). Moreover, let the function $M\colon \bigl[0, \frac{1}{\pi}\bigr] \to \bigl[0, \frac{\pi}{4}\bigr]$ be given by

$$M(x) := \frac{1}{2}\arcsin(\pi x)\,.$$

Finally, for $\theta \in \bigl[0, \frac{\pi}{2}\bigr]$ define

$$D(\theta) := \Bigl\{(\lambda_j) \in D \Bigm| \sum_{j=0}^{\infty} M(\lambda_j) = \theta\Bigr\} \subset D$$

and

$$T(\theta) := \sup\bigl\{\max W(\lambda) \bigm| \lambda \in D(\theta)\bigr\}\,, \tag{8.10}$$

where $\max W(\lambda) := \max_{j\in\mathbb{N}_0} t_j$ with $(t_j) = W(\lambda)$.

For every fixed $\theta \in \bigl[0, \frac{\pi}{2}\bigr]$, it is easy to verify that indeed $D(\theta) \neq \varnothing$.

Moreover, one has $0 \le T(\theta) \le 1/2$ Lemma 8.5 (b), and $T(\theta) = 0$ holds if and only if $\theta = 0$. In order to compute $T(\theta)$ for $\theta \in \bigl(0, \frac{\pi}{2}\bigr]$, we have to maximize $\max W(\lambda)$ over $\lambda \in D(\theta) \subset D$. This constrained optimization problem is the central part in the approach presented here.

The following proposition shows how this optimization problem is related to the problem of estimating the maximal angle between the corresponding spectral subspaces.

Proposition 8.7. *Assume Hypothesis 8.1. Let*

$$\Bigl[0, \frac{\pi}{2}\Bigr] \ni \theta \mapsto S(\theta) \in \Bigl[0, S\Bigl(\frac{\pi}{2}\Bigr)\Bigr] \subset \Bigl[0, \frac{1}{2}\Bigr]$$

be a continuous, strictly increasing (hence invertible) mapping with

$$0 \le S(\theta) \le T(\theta) \quad \textit{for} \quad 0 \le \theta < \frac{\pi}{2}\,.$$

Then

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le S^{-1}(t) \quad \textit{for} \quad 0 \le t < S\Bigl(\frac{\pi}{2}\Bigr)\,.$$

Proof. Since the mapping $\theta \mapsto S(\theta)$ is invertible, it suffices to show the inequality

$$\arcsin\bigl(\|P_0 - P_{S(\theta)}\|\bigr) \le \theta \quad \text{for} \quad 0 \le \theta < \frac{\pi}{2}\,. \tag{8.11}$$

Considering $T(0) = S(0) = 0$, the case $\theta = 0$ in inequality (8.11) is obvious. Let $\theta \in \bigl(0, \frac{\pi}{2}\bigr)$. In particular, one has $T(\theta) > 0$. For arbitrary t with $0 \le t < T(\theta)$ choose $\lambda = (\lambda_j) \in D(\theta)$ such that $t < \max W(\lambda) \le T(\theta)$. Denote $(t_j) := W(\lambda)$. Since $t_j < 1/2$ for all $j \in \mathbb{N}_0$ by Lemma 8.5 (b), it follows from the definition of (t_j) that

$$\frac{t_{j+1} - t_j}{1 - 2t_j} = \lambda_j \le \frac{1}{\pi} \quad \text{for all} \quad j \in \mathbb{N}_0\,.$$

Moreover, taking into account that $t < \max W(\lambda) = \max_{j \in \mathbb{N}_0} t_j$, there is $k \in \mathbb{N}_0$ such that $t_k \le t < t_{k+1}$. In particular, one has

$$\frac{t - t_k}{1 - 2t_k} < \frac{t_{k+1} - t_k}{1 - 2t_k} = \lambda_k \le \frac{1}{\pi}\,.$$

Considering the partition $0 = t_0 \le \cdots \le t_k \le t$ of the interval $[0, t]$, one now obtains from estimate (8.7) that

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \leq \sum_{j=0}^{k-1} M(\lambda_j) + M\Bigl(\frac{t-t_k}{1-2t_k}\Bigr) \leq \sum_{j=0}^{\infty} M(\lambda_j) = \theta\,,$$

that is,

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \leq \theta \quad \text{for all} \quad 0 \leq t < T(\theta)\,. \tag{8.12}$$

Since $S(\theta) < S\bigl(\frac{\pi}{2}\bigr) \leq \frac{1}{2}$ and the mapping $\bigl[0,\frac{1}{2}\bigr) \ni \tau \mapsto P_\tau$ is continuous in norm, estimate (8.12) also holds for $t = S(\theta) \leq T(\theta)$. This shows (8.11) and, hence, completes the proof. □

One can show that the mapping $\bigl[0,\frac{\pi}{2}\bigr] \ni \theta \mapsto T(\theta)$ is continuous and strictly increasing without having computed $T(\theta)$ explicitly. This mapping therefore satisfies the hypotheses of Proposition 8.7. However, we omit the corresponding argument here since this is also obtained from the explicit computation of $T(\theta)$ in Section 8.2 below. For convenience, the following proposition states the result of this computation in advance.

Proposition 8.8. *In the interval* $\bigl(0,\frac{\pi}{2}\bigr]$ *the equation*

$$\Bigl(1-\frac{2}{\pi}\sin\vartheta\Bigr)^2 = \Bigl(1-\frac{2}{\pi}\sin\Bigl(\frac{2\vartheta}{3}\Bigr)\Bigr)^3$$

has a unique solution $\vartheta \in \bigl(\arcsin\bigl(\frac{2}{\pi}\bigr),\frac{\pi}{2}\bigr)$. *Moreover, the quantity* $T(\theta)$ *defined in* (8.10) *has the representation*

$$T(\theta) = \begin{cases} \frac{1}{\pi}\sin(2\theta)\,, & 0 \leq \theta \leq \frac{1}{2}\arcsin\bigl(\frac{4\pi}{\pi^2+4}\bigr)\,,\\ \frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta\,, & \frac{1}{2}\arcsin\bigl(\frac{4\pi}{\pi^2+4}\bigr) < \theta < \arcsin\bigl(\frac{2}{\pi}\bigr)\,,\\ \frac{1}{2}-\frac{1}{2}\bigl(1-\frac{2}{\pi}\sin\theta\bigr)^2\,, & \arcsin\bigl(\frac{2}{\pi}\bigr) \leq \theta \leq \vartheta\,,\\ \frac{1}{2}-\frac{1}{2}\bigl(1-\frac{2}{\pi}\sin\bigl(\frac{2\theta}{3}\bigr)\bigr)^3\,, & \vartheta < \theta \leq \frac{\pi}{2}\,. \end{cases} \tag{8.13}$$

The mapping $\bigl[0,\frac{\pi}{2}\bigr] \ni \theta \mapsto T(\theta)$ *is strictly increasing, continuous on* $\bigl[0,\frac{\pi}{2}\bigr]$, *and continuous differentiable on* $\bigl(0,\frac{\pi}{2}\bigr) \setminus \{\vartheta\}$.

We are now able to turn to the main result of this chapter.

Theorem 8.9. *Assume Hypothesis 6.14. Suppose, in addition, that V satisfies $\|V\| < c_{\mathrm{crit}} \cdot d$ with*

$$c_{\mathrm{crit}} = \frac{1}{2} - \frac{1}{2}\Bigl(1 - \frac{\sqrt{3}}{\pi}\Bigr)^3 = 0.4548399\ldots \tag{8.14}$$

Then

$$\arcsin\bigl(\|P - Q\|\bigr) \le N\Bigl(\frac{\|V\|}{d}\Bigr) < \frac{\pi}{2}\,, \tag{8.15}$$

where the function $N\colon [0, c_{\mathrm{crit}}] \to \bigl[0, \frac{\pi}{2}\bigr]$ is given by

$$N(x) = \begin{cases} \frac{1}{2}\arcsin(\pi x) & \textit{for} \quad 0 \le x \le \frac{4}{\pi^2+4}\,, \\ \arcsin\Bigl(\sqrt{\frac{2\pi^2 x - 4}{\pi^2 - 4}}\,\Bigr) & \textit{for} \quad \frac{4}{\pi^2+4} < x < 4\,\frac{\pi^2-2}{\pi^4}\,, \\ \arcsin\bigl(\frac{\pi}{2}(1 - \sqrt{1-2x}\,)\bigr) & \textit{for} \quad 4\,\frac{\pi^2-2}{\pi^4} \le x \le \kappa\,, \\ \frac{3}{2}\arcsin\bigl(\frac{\pi}{2}(1 - \sqrt[3]{1-2x}\,)\bigr) & \textit{for} \quad \kappa < x \le c_{\mathrm{crit}}\,. \end{cases} \tag{8.16}$$

Here, $\kappa \in \bigl(4\frac{\pi^2-2}{\pi^4}, 2\frac{\pi-1}{\pi^2}\bigr)$ is the unique solution to the equation

$$\arcsin\Bigl(\frac{\pi}{2}\bigl(1 - \sqrt{1-2\kappa}\,\bigr)\Bigr) = \frac{3}{2}\arcsin\Bigl(\frac{\pi}{2}\bigl(1 - \sqrt[3]{1-2\kappa}\,\bigr)\Bigr) \tag{8.17}$$

in the interval $\bigl(0, 2\frac{\pi-1}{\pi^2}\bigr]$. The function N is strictly increasing, continuous on $[0, c_{\mathrm{crit}}]$, and continuously differentiable on $(0, c_{\mathrm{crit}}) \setminus \{\kappa\}$.

Proof of Theorem 8.9. As stated in Proposition 8.8, the mapping $\theta \mapsto T(\theta)$ is strictly increasing and continuous. Hence, its range is the whole interval $[0, c_{\mathrm{crit}}]$, where $c_{\mathrm{crit}} = T\bigl(\frac{\pi}{2}\bigr)$ is given by (8.14). Moreover, using the representation (8.13) in Proposition 8.8, it is easy to verify that the inverse $N = T^{-1}\colon [0, c_{\mathrm{crit}}] \to \bigl[0, \frac{\pi}{2}\bigr]$ is given by (8.16). In particular, the constant $\kappa = T(\vartheta) = \frac{1}{2} - \frac{1}{2}\bigl(1 - \frac{2}{\pi}\sin\vartheta\bigr)^2 \in \bigl(4\frac{\pi^2-2}{\pi^4}, 2\frac{\pi-1}{\pi^2}\bigr)$ is the unique solution to equation (8.17) in the interval $\bigl(0, 2\frac{\pi-1}{\pi^2}\bigr]$. Furthermore, it is clear from Proposition 8.8 that the function $N = T^{-1}$ is strictly increasing, continuous on $[0, c_{\mathrm{crit}}]$, and continuously differentiable on $(0, c_{\mathrm{crit}}) \setminus \{\kappa\}$.

It remains to show that estimate (8.15) holds. The case $V = 0$ is obvious. Assume that $V \neq 0$. Then, $B_t := A + td \cdot V/\|V\|$, $\mathrm{Dom}(B_t) := \mathrm{Dom}(A)$, and $P_t := \mathsf{E}_{B_t}\bigl(\mathcal{O}_{d/2}(\sigma)\bigr)$ for $0 \le t < \frac{1}{2}$ satisfy Hypothesis 8.1. Moreover, one has $P = P_0$, as well as $A + V = B_\tau$ and $Q = P_\tau$ with $\tau = \frac{\|V\|}{d} < c_{\mathrm{crit}} = T\bigl(\frac{\pi}{2}\bigr)$.

Applying Proposition 8.7 to the mapping $\theta \mapsto T(\theta)$ finally gives

$$\arcsin\bigl(\|P-Q\|\bigr) = \arcsin\bigl(\|P_0 - P_\tau\|\bigr) \le N(\tau) = N\Bigl(\frac{\|V\|}{d}\Bigr),$$

which completes the proof. □

Remark 8.10. Numerical evaluations give $\vartheta = 1.1286942\ldots < \arcsin\bigl(\frac{4\pi}{\pi^2+4}\bigr)$ and $\kappa = T(\vartheta) = 0.4098623\ldots < \frac{8\pi^2}{(\pi^2+4)^2}$.

However, estimate (8.15) remains valid if the constant κ in the explicit representation for the function N is replaced by any other constant within the interval $\bigl(4\frac{\pi^2-2}{\pi^4}, 2\frac{\pi-1}{\pi^2}\bigr)$. This can be seen by applying Proposition 8.7 to each of the two mappings

$$\theta \mapsto \frac{1}{2} - \frac{1}{2}\Bigl(1 - \frac{2}{\pi}\sin\theta\Bigr)^2 \quad \text{and} \quad \theta \mapsto \frac{1}{2} - \frac{1}{2}\Bigl(1 - \frac{2}{\pi}\sin\Bigl(\frac{2\theta}{3}\Bigr)\Bigr)^3.$$

These mappings indeed satisfy the hypotheses of Proposition 8.7. Both are continuous and strictly increasing, and, by suitable choices of $\lambda \in D(\theta)$, it is easy to see that they do not exceed $T(\theta)$, cf. equation (8.22) in Section 8.2 below.

The statement of Proposition 8.8 actually goes beyond that of Theorem 8.9. As a matter of fact, instead of equality in (8.13), it is sufficient for the statement of Theorem 8.9 to hold that the right-hand side of (8.13) does not exceed $T(\theta)$. This, in turn, is rather easy to establish by suitable choices of $\lambda \in D(\theta)$, see Lemma 8.15 and the proof of Lemma 8.18 below.

However, Proposition 8.8 states that the right-hand side of (8.13) provides an exact representation for $T(\theta)$, and most of the considerations in Section 8.2 are required to show this stronger result. As a consequence, the bound from Theorem 8.9 is optimal within the framework of the approach based on estimate (8.7).

In fact, the following observation shows that one requires a bound substantially stronger than the one from Corollary 7.2, at least for perturbations small in norm, in order to improve on Theorem 8.9.

Remark 8.11. One can modify the approach (8.7) by replacing the term $M(\lambda_j) = \frac{1}{2}\arcsin(\pi\lambda_j)$ with $N(\lambda_j)$, where N is the function from Theorem 8.9. In this case, the condition (8.6) can be relaxed to $\lambda_j \le c_{\mathrm{crit}}$. Yet, the corresponding optimization problem has exactly the same solution given by

(8.13). This can be seen from the fact that each $N(\lambda_j)$ is of the form of the right-hand side of (8.7) (cf. the computation of $T(\theta)$ in Section 8.2 below), so that we are actually dealing with essentially the same optimization problem. In this sense, the function N is a fixed point in the approach presented here.

We close this section with a comparison between the bound from Theorem 8.9 and the strongest previously known one (2.27), proved by Albeverio and Motovilov in [8].

Remark 8.12. The function $M_*\colon [0,c_*] \to \bigl[0,\frac{\pi}{2}\bigr]$ from (2.27) agrees on the interval $\bigl[0,\frac{4}{\pi^2+4}\bigr]$ with N. For $\frac{4}{\pi^2+4} < t \le c_*$, however, the strict inequality $N(t) < M_*(t)$ holds. Indeed, it follows from the computation of $T(\theta)$ in Section 8.2 that

$$t < T(M_*(t)) \le c_{\mathrm{crit}} \quad \text{for} \quad \frac{4}{\pi^2+4} < t \le c_*\,,$$

see Remark 8.22 below. Since the function $N = T^{-1}\colon [0,c_{\mathrm{crit}}] \to \bigl[0,\frac{\pi}{2}\bigr]$ is strictly increasing, this implies that

$$N(t) < N\bigl(T(M_*(t))\bigr) = M_*(t) \quad \text{for} \quad \frac{4}{\pi^2+4} < t \le c_*\,.$$

8.2 Proof of Proposition 8.8

We split the proof of Proposition 8.8 into several steps. We first reduce the problem of computing $T(\theta)$ to the problem of solving suitable finite-dimensional constrained optimization problems, see equations (8.18) and (8.20) below. The corresponding critical points are then characterized in Lemma 8.15 using Lagrange multipliers. The crucial tool to reduce the set of relevant critical points is provided by Lemma 8.16. Finally, the finite-dimensional optimization problems are solved in Lemmas 8.18, 8.20 and Proposition 8.21.

Throughout this section, we make use of the notations introduced in Definitions 8.4 and 8.6. In addition, we fix the following notations.

Definition 8.13. For $n \in \mathbb{N}_0$ and $\theta \in \bigl[0,\frac{\pi}{2}\bigr]$ define $D_n(\theta) := D(\theta) \cap D_n$. Moreover, let

$$T_n(\theta) := \sup\bigl\{\max W(\lambda) \bigm| \lambda \in D_n(\theta)\bigr\} \quad \text{if} \quad D_n(\theta) \ne \varnothing\,,$$

and set $T_n(\theta) := 0$ if $D_n(\theta) = \varnothing$.

Since $D(0) = D_n(0)$ contains only the sequence identical to zero, one has $T(0) = T_n(0) = 0$ for every $n \in \mathbb{N}_0$. Let $\theta \in \left(0, \frac{\pi}{2}\right]$ be arbitrary. In view of the inclusions $D_0(\theta) \subset D_1(\theta) \subset D_2(\theta) \subset \dots$, the sequence $(T_n(\theta))_n$ is increasing, that is,

$$T_0(\theta) \le T_1(\theta) \le T_2(\theta) \le \dots$$

Moreover, we observe that

$$T(\theta) = \sup_{n \in \mathbb{N}_0} T_n(\theta)\,. \tag{8.18}$$

In fact, we show below that $T_n(\theta) = T_2(\theta)$ for every $n \ge 2$, so that $T(\theta)$ agrees with $T_2(\theta)$, see Proposition 8.21.

Let $n \in \mathbb{N}$ be arbitrary and let $\lambda = (\lambda_j) \in D_n$. Denote $(t_j) := W(\lambda)$. It follows from Lemma 8.5 (b) that $\max W(\lambda) = t_{n+1}$. Moreover, we have

$$1 - 2t_{j+1} = 1 - 2t_j - 2\lambda_j(1 - 2t_j) = (1 - 2t_j)(1 - 2\lambda_j)\,, \quad j = 0, \dots, n\,.$$

Since $t_0 = 0$, this implies that

$$1 - 2t_{n+1} = \prod_{j=0}^{n} (1 - 2\lambda_j)\,.$$

In particular, we obtain the explicit representation

$$\max W(\lambda) = t_{n+1} = \frac{1}{2}\left(1 - \prod_{j=0}^{n} (1 - 2\lambda_j)\right). \tag{8.19}$$

An immediate conclusion of representation (8.19) is the following statement.

Lemma 8.14. *For $\lambda = (\lambda_j) \in D_n$ the value of $\max W(\lambda)$ does not depend on the order of the entries $\lambda_0, \dots, \lambda_n$.*

Another implication of representation (8.19) is that $\max W(\lambda) = t_{n+1}$ can be considered as a continuous function of the variables $\lambda_0, \dots, \lambda_n$. Since the set $D_n(\theta)$ is compact as a closed bounded subset of an $(n+1)$-dimensional subspace of the sequences with finite support, we deduce that $T_n(\theta)$ can be written as

$$T_n(\theta) = \max\bigl\{t_{n+1} \bigm| (t_j) = W(\lambda)\,,\ \lambda \in D_n(\theta)\bigr\}\,. \tag{8.20}$$

Hence, $T_n(\theta)$ is determined by a finite-dimensional constrained optimization problem, which can be studied by use of Lagrange multipliers.

Taking into account the definition of the set $D_n(\theta)$, it follows from (8.19) and (8.20) that there is some point $(\lambda_0, \dots, \lambda_n) \in \bigl[0, \frac{1}{\pi}\bigr]^{n+1}$ satisfying

$$T_n(\theta) = t_{n+1} = \frac{1}{2}\Bigl(1 - \prod_{j=0}^{n}(1-2\lambda_j)\Bigr) \qquad \text{and} \qquad \sum_{j=0}^{n} M(\lambda_j) = \theta\,,$$

where $M(x) = \frac{1}{2}\arcsin(\pi x)$. In particular, if $(\lambda_0, \dots, \lambda_n) \in \bigl(0, \frac{1}{\pi}\bigr)^{n+1}$, then the method of Lagrange multipliers gives a constant $r \in \mathbb{R}$, $r \neq 0$, with

$$\frac{\partial t_{n+1}}{\partial \lambda_k} = r \cdot M'(\lambda_k) = r \cdot \frac{\pi}{2\sqrt{1-\pi^2\lambda_k^2}} \qquad \text{for} \qquad k = 0, \dots, n\,.$$

Hence, in this case, for every $k \in \{0, \dots, n-1\}$ we obtain

$$\frac{\sqrt{1-\pi^2\lambda_k^2}}{\sqrt{1-\pi^2\lambda_{k+1}^2}} = \frac{\dfrac{\partial t_{n+1}}{\partial \lambda_{k+1}}}{\dfrac{\partial t_{n+1}}{\partial \lambda_k}} = \frac{\prod\limits_{\substack{j=0\\ j\neq k+1}}^{n} (1-2\lambda_j)}{\prod\limits_{\substack{j=0\\ j\neq k}}^{n} (1-2\lambda_j)} = \frac{1-2\lambda_k}{1-2\lambda_{k+1}}\,. \tag{8.21}$$

This leads to the following characterization of critical points of the mapping $\lambda \mapsto \max W(\lambda)$ on $D_n(\theta)$.

Lemma 8.15. *For $n \geq 1$ and $\theta \in \bigl(0, \frac{\pi}{2}\bigr]$ let $\lambda = (\lambda_j) \in D_n(\theta)$ satisfy $T_n(\theta) = \max W(\lambda)$. Assume that $\lambda_0 \geq \dots \geq \lambda_n$. If, in addition, $\lambda_0 < \frac{1}{\pi}$ and $\lambda_n > 0$, then one has*

$$\lambda_0 = \dots = \lambda_n = \frac{1}{\pi}\sin\Bigl(\frac{2\theta}{n+1}\Bigr)\,,$$

so that

$$\max W(\lambda) = \frac{1}{2} - \frac{1}{2}\left(1 - \frac{2}{\pi}\sin\Bigl(\frac{2\theta}{n+1}\Bigr)\right)^{n+1}\,, \tag{8.22}$$

or there is $l \in \{0, \dots, n-1\}$ *with*

$$\frac{4}{\pi^2+4} > \lambda_0 = \dots = \lambda_l > \frac{2}{\pi^2} > \lambda_{l+1} = \dots = \lambda_n > 0\,. \tag{8.23}$$

In the latter case, λ_0 *and* λ_n *satisfy*

$$\lambda_0 + \lambda_n = \frac{4\alpha^2}{\pi^2+4\alpha^2} \qquad \textit{and} \qquad \lambda_0\lambda_n = \frac{\alpha^2-1}{\pi^2+4\alpha^2} \tag{8.24}$$

with

$$\alpha = \frac{\sqrt{1-\pi^2\lambda_0^2}}{1-2\lambda_0} = \frac{\sqrt{1-\pi^2\lambda_n^2}}{1-2\lambda_n}\,, \tag{8.25}$$

and α *lies within the bounds*

$$1 < \alpha < m := \frac{\pi}{2}\tan\Bigl(\arcsin\Bigl(\frac{2}{\pi}\Bigr)\Bigr)\,.$$

Proof. Let $\lambda_0 < \frac{1}{\pi}$ and $\lambda_n > 0$. In particular, the point $(\lambda_0, \dots, \lambda_n)$ lies in $\bigl(0, \frac{1}{\pi}\bigr)^{n+1}$. Hence, it follows from (8.21) that

$$\alpha := \frac{\sqrt{1-\pi^2\lambda_k^2}}{1-2\lambda_k} \tag{8.26}$$

does not depend on $k \in \{0, \dots, n\}$.

If $\lambda_0 = \lambda_n$, then all λ_j coincide and one has $\theta = (n+1)M(\lambda_0)$, that is, $\lambda_0 = \dots = \lambda_n = \frac{1}{\pi}\sin\bigl(\frac{2\theta}{n+1}\bigr)$. Inserting this into (8.19) yields representation (8.22).

Now assume that $\lambda_0 > \lambda_n$. A straightforward calculation shows that $x = 2/\pi^2$ is the only critical point of the mapping

$$\Bigl[0, \frac{1}{\pi}\Bigr] \ni x \mapsto \frac{\sqrt{1-\pi^2x^2}}{1-2x}\,, \tag{8.27}$$

cf. Fig. 8.1. The image of this point is $\bigl(1-\frac{4}{\pi^2}\bigr)^{-1/2} = m > 1$. Moreover, 0 and $\frac{4}{\pi^2+4}$ are mapped to 1, and $\frac{1}{\pi}$ is mapped to 0. In particular, every value in the interval $(1, m)$ has exactly two preimages under the mapping (8.27), and all the other values in the range $[0, m]$ have only one preimage. Since $\lambda_0 > \lambda_n$ by assumption, it follows from (8.26) that α has two preimages. Hence, $\alpha \in (1, m)$ and $\frac{4}{\pi^2+4} > \lambda_0 > \frac{2}{\pi^2} > \lambda_n > 0$. Furthermore, there is

$l \in \{0, \dots, n-1\}$ with $\lambda_0 = \cdots = \lambda_l$ and $\lambda_{l+1} = \cdots = \lambda_n$. This proves (8.23) and (8.25).

Finally, the equation $\frac{\sqrt{1-\pi^2 z^2}}{1-2z} = \alpha$ can be rewritten as

$$0 = z^2 - \frac{4\alpha^2}{\pi^2 + 4\alpha^2}\, z + \frac{\alpha^2 - 1}{\pi^2 + 4\alpha^2} = (z-\lambda_0)(z-\lambda_n) = z^2 - (\lambda_0 + \lambda_n) z + \lambda_0 \lambda_n \,,$$

which shows the relations (8.24). □

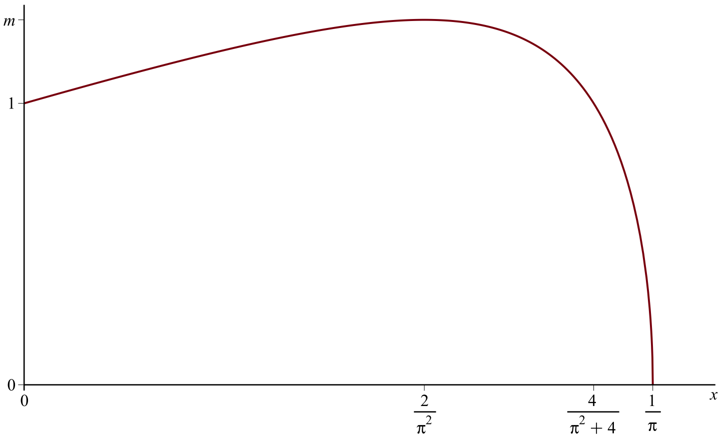

Fig. 8.1: The mapping $\left[0, \frac{1}{\pi}\right] \ni x \mapsto \frac{\sqrt{1-\pi^2 x^2}}{1-2x}$.

The preceding lemma is one of the main ingredients for solving the constrained optimization problem that defines the quantity $T_n(\theta)$ in (8.20). However, it is still a hard task to compute $T_n(\theta)$ from the corresponding critical points. Especially the case (8.23) in Lemma 8.15 is difficult to handle and needs careful treatment. An efficient computation of $T_n(\theta)$ therefore requires a technique that allows to narrow down the set of relevant critical points. The following result provides an adequate tool for this and is thus crucial for the remaining considerations.

Lemma 8.16. *For $n \geq 1$ and $\theta \in \left(0, \frac{\pi}{2}\right]$ let $\lambda = (\lambda_j) \in D_n(\theta)$ satisfy $T_n(\theta) = \max W(\lambda)$. Then, for every $k \in \{0, \dots, n\}$ one has*

$$\max W\big((\lambda_0, \dots, \lambda_k, 0, \dots)\big) = T_k(\theta_k) \qquad \textit{with} \qquad \theta_k := \sum_{j=0}^{k} M(\lambda_j) \leq \theta \,.$$

Proof. The case $k = n$ agrees with the hypothesis that $T_n(\theta) = \max W(\lambda)$.

Let $k \in \{0, \dots, n-1\}$ be arbitrary and denote $(t_j) := W(\lambda)$. It follows from Lemma 8.5 (b) that $t_{k+1} = \max W\big((\lambda_0, \dots, \lambda_k, 0, \dots)\big)$. In particular, one has $t_{k+1} \le T_k(\theta_k)$ since $(\lambda_0, \dots, \lambda_k, 0, \dots) \in D_k(\theta_k)$.

Assume that $t_{k+1} < T_k(\theta_k)$, and choose $\gamma = (\gamma_j) \in D_k(\theta_k)$ such that $\max W(\gamma) = T_k(\theta_k)$. Denote $\mu := (\gamma_0, \dots, \gamma_k, \lambda_{k+1}, \dots, \lambda_n, 0, \dots) \in D_n(\theta_n)$ and $(s_j) := W(\mu)$. Again by Lemma 8.5 (b), $s_{k+1} = \max W(\gamma) > t_{k+1}$ and $s_{n+1} = \max W(\mu) \le T_n(\theta_n)$. Taking into account Lemma 8.5 (a) and the definition of the operator W, one concludes that

$$t_{k+2} = t_{k+1} + \lambda_{k+1}(1 - 2t_{k+1}) < s_{k+1} + \lambda_{k+1}(1 - 2s_{k+1}) = s_{k+2}\,.$$

Iterating this estimate eventually gives $t_{n+1} < s_{n+1} \le T_n(\theta_n)$, which contradicts the case $k = n$ from above. Thus,

$$\max W\big((\lambda_0, \dots, \lambda_k, 0, \dots)\big) = t_{k+1} = T_k(\theta_k)$$

as claimed. □

Lemma 8.16 states that if a sequence $\lambda \in D_n(\theta)$ solves the optimization problem for $T_n(\theta)$, then every truncation of λ solves the corresponding reduced optimization problem. This allows to exclude many sequences in $D_n(\theta)$ from the considerations once the optimization problem is solved for small n. The number of parameters in (8.20) can thereby be reduced considerably.

The following lemma demonstrates this technique. It implies that the condition $\lambda_0 < \frac{1}{\pi}$ in Lemma 8.15 is always satisfied except for one single case, which can be treated separately.

Lemma 8.17. *For $n \ge 1$ and $\theta \in \big(0, \frac{\pi}{2}\big]$ let $\lambda = (\lambda_j) \in D_n(\theta)$ satisfy $T_n(\theta) = \max W(\lambda)$ and $\lambda_0 \ge \cdots \ge \lambda_n$. If $\theta < \pi/2$ or $n \ge 2$, then $\lambda_0 < 1/\pi$.*

Proof. Suppose that $\lambda_0 = 1/\pi$. We have to show that $\theta = \pi/2$ and $n = 1$.

Define $\theta_1 := M(\lambda_0) + M(\lambda_1) \le \theta$. It is clear that $\lambda \in D_1\big(\frac{\pi}{2}\big)$ is equivalent to $\theta_1 = \theta = \pi/2$. Assume that $\theta_1 < \pi/2$. Then, one has $\theta_1 \ge M(\lambda_0) = \pi/4$ and $\lambda_1 = \frac{1}{\pi}\sin\big(2\theta_1 - \frac{\pi}{2}\big) = -\frac{1}{\pi}\left(1 - 2\sin^2\theta_1\right) \in \big[0, \frac{1}{\pi}\big)$. Taking into account representation (8.19), for $\mu := (\lambda_0, \lambda_1, 0, \dots) \in D_1(\theta_1)$ one computes

$$\begin{aligned}\max W(\mu) &= \frac{1}{2} - \frac{1}{2}(1-2\lambda_0)(1-2\lambda_1) = (\lambda_0+\lambda_1) - 2\lambda_0\lambda_1 \\ &= \frac{2}{\pi}\sin^2\theta_1 + \frac{2}{\pi^2}\big(1 - 2\sin^2\theta_1\big) = \frac{2}{\pi^2} + \frac{2\pi-4}{\pi^2}\sin^2\theta_1\,.\end{aligned}$$

Since $\arcsin\big(\frac{1}{\pi-1}\big) < \frac{\pi}{4} \le \theta_1 < \frac{\pi}{2}$, it now follows from Lemma A.1 (a) that

$$\max W(\mu) < \frac{2}{\pi}\Big(1 - \frac{1}{\pi}\sin\theta_1\Big)\sin\theta_1 = \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{2}{\pi}\sin\theta_1\Big)^2 \le T_1(\theta_1)\,,$$

where the last inequality is due to representation (8.22). This is a contradiction to Lemma 8.16. Thus, $\theta_1 = \theta = \frac{\pi}{2}$ and, in particular, $\lambda = \mu \in D_1\big(\frac{\pi}{2}\big)$.

Obviously, one has $D_1\big(\frac{\pi}{2}\big) = \big\{\big(\frac{1}{\pi}, \frac{1}{\pi}, 0, \dots\big)\big\}$, so that $\lambda = \big(\frac{1}{\pi}, \frac{1}{\pi}, 0, \dots\big)$. Taking into account that $\sin\big(\frac{\pi}{3}\big) = \frac{\sqrt{3}}{2}$, it follows from representations (8.19) and (8.22) that

$$\max W(\lambda) = \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{2}{\pi}\Big)^2 < \frac{1}{2} - \frac{1}{2}\bigg(1 - \frac{\sqrt{3}}{\pi}\bigg)^3 \le T_2\Big(\frac{\pi}{2}\Big)\,.$$

Since $\max W(\lambda) = T_n(\theta)$ by hypothesis, this implies that $n = 1$, which completes the proof. □

We are now able to solve the finite-dimensional constrained optimization problem in (8.20) for every $\theta \in \big[0, \frac{\pi}{2}\big]$ and $n \in \mathbb{N}$. We start with the case $n = 1$.

Lemma 8.18. *The quantity $T_1(\theta)$ has the representation*

$$T_1(\theta) = \begin{cases} T_0(\theta) = \frac{1}{\pi}\,\sin(2\theta) & \textit{for} \quad 0 \le \theta \le \frac{1}{2}\arcsin\big(\frac{4\pi}{\pi^2+4}\big)\,, \\ \frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta & \textit{for} \quad \frac{1}{2}\arcsin\big(\frac{4\pi}{\pi^2+4}\big) < \theta < \arcsin\big(\frac{2}{\pi}\big)\,, \\ \frac{1}{2} - \frac{1}{2}\big(1 - \frac{2}{\pi}\sin\theta\big)^2 & \textit{for} \quad \arcsin\big(\frac{2}{\pi}\big) \le \theta \le \frac{\pi}{2}\,. \end{cases}$$

In particular, if $0 < \theta < \arcsin\big(\frac{2}{\pi}\big)$ and $\lambda = (\lambda_0, \lambda_1, 0, \dots) \in D_1(\theta)$ with $\lambda_0 = \lambda_1$, then the strict inequality $\max W(\lambda) < T_1(\theta)$ holds.

The mapping $\big[0, \frac{\pi}{2}\big] \ni \theta \mapsto T_1(\theta)$ is strictly increasing, continuous on $\big[0, \frac{\pi}{2}\big]$, and continuously differentiable on $\big(0, \frac{\pi}{2}\big)$.

Proof. Since $T_1(0) = T_0(0) = 0$, the representation is obviously correct for $\theta = 0$. For $\theta = \frac{\pi}{2}$ one has $D_1\big(\frac{\pi}{2}\big) = \big\{\big(\frac{1}{\pi}, \frac{1}{\pi}, 0, \dots\big)\big\}$, so that representation

(8.19) gives $T_1\big(\frac{\pi}{2}\big) = \frac{1}{2} - \frac{1}{2}\big(1 - \frac{2}{\pi}\big)^2$. This also agrees with the claim.

Now let $\theta \in \big(0, \frac{\pi}{2}\big)$ be arbitrary. Obviously, $D_0(\theta)$ contains only the sequence $\big(\frac{1}{\pi}\sin(2\theta), 0, \dots\big)$ if $\theta \le \frac{\pi}{4}$, and one has $D_0(\theta) = \varnothing$ if $\theta > \frac{\pi}{4}$. Hence,

$$T_0(\theta) = \frac{1}{\pi}\sin(2\theta) \quad \text{if} \quad 0 < \theta \le \frac{\pi}{4}\,, \tag{8.28}$$

and $T_0(\theta) = 0$ if $\theta > \pi/4$.

By Lemmas 8.14, 8.15, and 8.17 there are only two sequences in the set $D_1(\theta) \setminus D_0(\theta)$ that need to be considered in order to compute $T_1(\theta)$. One of them is given by $\mu = (\mu_0, \mu_1, 0, \dots)$ with $\mu_0 = \mu_1 = \frac{1}{\pi}\sin\theta \in \big(0, \frac{1}{\pi}\big)$. For this sequence, representation (8.22) yields

$$\max W(\mu) = \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{2}{\pi}\sin\theta\Big)^2 = \frac{2}{\pi}\Big(1 - \frac{1}{\pi}\sin\theta\Big)\sin\theta\,. \tag{8.29}$$

The other sequence in the set $D_1(\theta) \setminus D_0(\theta)$ that needs to be considered is $\lambda = (\lambda_0, \lambda_1, 0, \dots)$ with λ_0 and λ_1 satisfying $\frac{4}{\pi^2+4} > \lambda_0 > \frac{2}{\pi^2} > \lambda_1 > 0$ and

$$\lambda_0 + \lambda_1 = \frac{4\alpha^2}{\pi^2 + 4\alpha^2}\,, \quad \lambda_0\lambda_1 = \frac{\alpha^2 - 1}{\pi^2 + 4\alpha^2}\,, \tag{8.30}$$

where

$$\alpha = \frac{\sqrt{1 - \pi^2\lambda_0^2}}{1 - 2\lambda_0} = \frac{\sqrt{1 - \pi^2\lambda_1^2}}{1 - 2\lambda_1} \in \Big(1, \frac{\pi}{2}\tan\Big(\arcsin\frac{2}{\pi}\Big)\Big)\,. \tag{8.31}$$

It turns out shortly that this sequence λ exists if and only if θ satisfies the two-sided estimate $\arctan\big(\frac{2}{\pi}\big) < \theta < \arcsin\big(\frac{2}{\pi}\big)$.

Using representation (8.19) and the relations in (8.30), one obtains

$$\begin{aligned} \max W(\lambda) &= \frac{1}{2} - \frac{1}{2}(1 - 2\lambda_0)(1 - 2\lambda_1) = (\lambda_0 + \lambda_1) - 2\lambda_0\lambda_1 \\ &= 2\,\frac{\alpha^2 + 1}{\pi^2 + 4\alpha^2}\,. \end{aligned} \tag{8.32}$$

The objective is to rewrite the right-hand side of (8.32) in terms of θ.

It follows from

$$2\theta = \arcsin(\pi\lambda_0) + \arcsin(\pi\lambda_1) \tag{8.33}$$

and the relations (8.30) and (8.31) that

$$\begin{aligned}\sin(2\theta) &= \pi\lambda_0\sqrt{1-\pi^2\lambda_1^2} + \pi\lambda_1\sqrt{1-\pi^2\lambda_0^2} \\ &= \alpha\pi\lambda_0(1-2\lambda_1) + \alpha\pi\lambda_1(1-2\lambda_0) \\ &= \alpha\pi\left(\lambda_0+\lambda_1-4\lambda_0\lambda_1\right) = \frac{4\alpha\pi}{\pi^2+4\alpha^2}\,.\end{aligned} \tag{8.34}$$

Taking into account that $\sin(2\theta) > 0$, equation (8.34) can be rewritten as

$$\alpha^2 - \frac{\pi}{\sin(2\theta)}\alpha + \frac{\pi^2}{4} = 0\,.$$

In turn, this gives

$$\alpha = \frac{\pi}{2\sin(2\theta)}\left(1 \pm \sqrt{1-\sin^2(2\theta)}\right) = \frac{\pi}{2}\,\frac{1 \pm |\cos^2\theta - \sin^2\theta|}{2\sin\theta\cos\theta}\,,$$

that is,

$$\alpha = \frac{\pi}{2}\tan\theta \qquad \text{or} \qquad \alpha = \frac{\pi}{2}\cot\theta\,. \tag{8.35}$$

We show that the second case in (8.35) does not occur.

Since $1 < \alpha < \frac{\pi}{2}\tan\bigl(\arcsin\bigl(\frac{2}{\pi}\bigr)\bigr) < \frac{\pi}{2}$, by equation (8.34) one has $\sin(2\theta) < 1$, which implies that $\theta \neq \frac{\pi}{4}$. Moreover, combining relations (8.30) and (8.31), λ_1 can be expressed in terms of λ_0 alone. Hence, by equation (8.33) the quantity θ can be written as a continuous function of the sole variable $\lambda_0 \in \bigl(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\bigr)$. Taking the limit $\lambda_0 \to \frac{4}{\pi^2+4}$ in equation (8.33) then implies that $\lambda_1 \to 0$ and, therefore, $\theta \to \frac{1}{2}\arcsin\bigl(\frac{4\pi}{\pi^2+4}\bigr) < \frac{\pi}{4}$. This yields $\theta < \frac{\pi}{4}$ for every value of λ_0 in $\bigl(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\bigr)$ by continuity, that is, the sequence λ can exist only if $\theta < \frac{\pi}{4}$. Taking into account that α satisfies $1 < \alpha < \frac{\pi}{2}\tan\bigl(\arcsin\bigl(\frac{2}{\pi}\bigr)\bigr)$, it now follows from (8.35) that the sequence λ exists if and only if θ satisfies $\arctan\bigl(\frac{2}{\pi}\bigr) < \theta < \arcsin\bigl(\frac{2}{\pi}\bigr)$, and, in this case, one has

$$\alpha = \frac{\pi}{2}\tan\theta\,. \tag{8.36}$$

Combining equations (8.32) and (8.36) finally gives

$$\max W(\lambda) = \frac{1}{2}\,\frac{\frac{4}{\pi^2}+\tan^2\theta}{1+\tan^2\theta} = \frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta \tag{8.37}$$

for $\arctan\left(\frac{2}{\pi}\right) < \theta < \arcsin\left(\frac{2}{\pi}\right)$.

As a result of Lemmas 8.14, 8.15, and 8.17, the quantities (8.28), (8.29), and (8.37) are the only possible values for $T_1(\theta)$, and we have to determine which of them is the greatest.

The easiest case is $\theta > \frac{\pi}{4}$ since then (8.29) is the only possibility for $T_1(\theta)$.

The quantity (8.37) is relevant only if $\arctan\left(\frac{2}{\pi}\right) < \theta < \arcsin\left(\frac{2}{\pi}\right) < \frac{\pi}{4}$. In this case, it follows from parts (b) and (c) of Lemma A.1 that (8.37) gives the greatest value of the three possibilities and, hence, is the correct term for $T_1(\theta)$ here.

For $0 < \theta \le \arctan\left(\frac{2}{\pi}\right) < 2\arctan\left(\frac{1}{\pi}\right)$, by Lemma A.1 (d) the quantity (8.28) is greater than (8.29). Therefore, $T_1(\theta)$ is given by (8.28) in this case.

Finally, consider the case $\arcsin\left(\frac{2}{\pi}\right) \le \theta \le \frac{\pi}{4}$. It follows from Lemma A.1 (e) and the inequality $2\arctan\left(\frac{1}{\pi}\right) < \arcsin\left(\frac{2}{\pi}\right)$ that (8.29) is then greater than (8.28) and, hence, coincides with $T_1(\theta)$.

Upon observing the identity $\arctan\left(\frac{2}{\pi}\right) = \frac{1}{2}\arcsin\left(\frac{4\pi}{\pi^2+4}\right)$, this completes the computation of $T_1(\theta)$ for $\theta \in \left[0, \frac{\pi}{2}\right]$. In particular, it follows from the discussion of the two cases $0 < \theta \le \arctan\left(\frac{2}{\pi}\right)$ and $\arctan\left(\frac{2}{\pi}\right) < \theta < \arcsin\left(\frac{2}{\pi}\right)$ that $\max W(\mu)$ is always strictly less than $T_1(\theta)$ if $0 < \theta < \arcsin\left(\frac{2}{\pi}\right)$.

The piecewise defined mapping $\left[0, \frac{\pi}{2}\right] \ni \theta \mapsto T_1(\theta)$ is continuously differentiable on each of the corresponding subintervals. It remains to show that the mapping is continuous and continuously differentiable at the points $\theta = \arctan\left(\frac{2}{\pi}\right) = \frac{1}{2}\arcsin\left(\frac{4\pi}{\pi^2+4}\right)$ and $\theta = \arcsin\left(\frac{2}{\pi}\right)$.

Taking into account that $\sin^2\theta = \frac{4}{\pi^2+4}$ for $\theta = \frac{1}{2}\arcsin\left(\frac{4\pi}{\pi^2+4}\right)$, the continuity is straightforward to verify. The continuous differentiability follows from the relations

$$\frac{\pi^2-4}{\pi^2}\sin\theta\cos\theta = \frac{2}{\pi}\Bigl(1 - \frac{2}{\pi}\sin\theta\Bigr)\cos\theta\,, \quad \theta = \arcsin\Bigl(\frac{2}{\pi}\Bigr)\,,$$

and

$$\frac{2}{\pi}\cos(2\theta) = \frac{\pi^2-4}{2\pi^2}\sin(2\theta) = \frac{\pi^2-4}{\pi^2}\sin\theta\cos\theta\,, \quad \theta = \frac{1}{2}\arcsin\Bigl(\frac{4\pi}{\pi^2+4}\Bigr)\,,$$

where the latter is due to

$$\cot\Bigl(\arcsin\Bigl(\frac{4\pi}{\pi^2+4}\Bigr)\Bigr) = \frac{\sqrt{1-\frac{16\pi^2}{(\pi^2+4)^2}}}{\frac{4\pi}{\pi^2+4}} = \frac{\pi^2-4}{4\pi}\,.$$

This completes the proof. □

So far, Lemma 8.16 has been used only to prove Lemma 8.17. Its whole strength becomes apparent in connection with Lemma 8.14. This is demonstrated in the following corollary to Lemma 8.18. It states that in (8.23) the sequences with $l \in \{0, \dots, n-2\}$ do not need to be considered.

Corollary 8.19. *In the case* (8.23) *in Lemma 8.15 one has* $l = n - 1$.

Proof. The case $n = 1$ is obvious. For $n \geq 2$ let $\lambda = (\lambda_0, \dots, \lambda_n, 0, \dots)$ be a sequence in $D_n(\theta)$ with

$$\frac{4}{\pi^2 + 4} > \lambda_0 = \dots = \lambda_l > \frac{2}{\pi^2} > \lambda_{l+1} = \dots = \lambda_n > 0$$

for some $l \in \{0, \dots, n-2\}$. In particular, one has $0 < \lambda_{n-1} = \lambda_n < 2/\pi^2$, which implies that $0 < \tilde{\theta} := M(\lambda_{n-1}) + M(\lambda_n) < \arcsin\bigl(\frac{2}{\pi}\bigr)$. Hence, it follows from Lemma 8.18 that

$$\max W\bigl((\lambda_{n-1}, \lambda_n, 0, \dots)\bigr) < T_1(\tilde{\theta}) .$$

By Lemmas 8.14 and 8.16 one concludes that

$$\max W(\lambda) = \max W\bigl((\lambda_{n-1}, \lambda_n, \lambda_0, \dots, \lambda_{n-2}, 0, \dots)\bigr) < T_n(\theta) ,$$

which leaves $l = n - 1$ as the only possibility in (8.23). □

We now turn to the computation of $T_2(\theta)$ for $\theta \in \bigl[0, \frac{\pi}{2}\bigr]$.

Lemma 8.20. *In the interval* $\bigl(0, \frac{\pi}{2}\bigr]$ *the equation*

$$\Bigl(1 - \frac{2}{\pi} \sin\vartheta\Bigr)^2 = \Bigl(1 - \frac{2}{\pi} \sin\Bigl(\frac{2\vartheta}{3}\Bigr)\Bigr)^3 \tag{8.38}$$

has a unique solution $\vartheta \in \bigl(\arcsin\bigl(\frac{2}{\pi}\bigr), \frac{\pi}{2}\bigr)$. *Moreover, the quantity* $T_2(\theta)$ *has the representation*

$$T_2(\theta) = \begin{cases} T_1(\theta) & \textit{for} \quad 0 \leq \theta \leq \vartheta , \\ \dfrac{1}{2} - \dfrac{1}{2}\Bigl(1 - \dfrac{2}{\pi} \sin\Bigl(\dfrac{2\theta}{3}\Bigr)\Bigr)^3 & \textit{for} \quad \vartheta < \theta \leq \frac{\pi}{2} . \end{cases}$$

In particular, one has $T_1(\theta) < T_2(\theta)$ *if* $\theta > \vartheta$, *and the strict inequality* $\max W(\lambda) < T_2(\theta)$ *holds for* $\theta \in \bigl(0, \frac{\pi}{2}\bigr]$ *and* $\lambda = (\lambda_0, \lambda_1, \lambda_2, 0, \dots) \in D_2(\theta)$ *with* $\lambda_0 = \lambda_1 > \lambda_2 > 0$.

The mapping $\left[0,\frac{\pi}{2}\right] \ni \theta \mapsto T_2(\theta)$ *is strictly increasing, continuous on* $\left[0,\frac{\pi}{2}\right]$*, and continuously differentiable on* $\left(0,\frac{\pi}{2}\right) \setminus \{\vartheta\}$*.*

Proof. Since $T_2(0) = T_1(0) = 0$, the case $\theta = 0$ in the representation for $T_2(\theta)$ is obvious. Let $\theta \in \left(0,\frac{\pi}{2}\right]$ be arbitrary. It follows from Lemmas 8.14, 8.15, and 8.17 and Corollary 8.19 that there are only two sequences in $D_2(\theta) \setminus D_1(\theta)$ that need to be considered in order to compute $T_2(\theta)$. One of them is $\mu = (\mu_0, \mu_1, \mu_2, 0, \dots)$ with $\mu_0 = \mu_1 = \mu_2 = \frac{1}{\pi}\sin\left(\frac{2\theta}{3}\right)$. For this sequence representation (8.22) yields

$$\max W(\mu) = \frac{1}{2} - \frac{1}{2}\left(1 - \frac{2}{\pi}\sin\left(\frac{2\theta}{3}\right)\right)^3 . \tag{8.39}$$

The other sequence in the set $D_2(\theta) \setminus D_1(\theta)$ that needs to be considered is $\lambda = (\lambda_0, \lambda_1, \lambda_2, 0, \dots)$, where $\frac{4}{\pi^2+4} > \lambda_0 = \lambda_1 > \frac{2}{\pi^2} > \lambda_2 > 0$ and λ_0 and λ_2 are given by (8.24) and (8.25). Using representation (8.19), one obtains

$$\max W(\lambda) = \frac{1}{2} - \frac{1}{2}(1 - 2\lambda_0)^2(1 - 2\lambda_2) . \tag{8.40}$$

According to Lemma A.3, this sequence λ can exist only if θ satisfies the two-sided estimate $\frac{3}{2}\arcsin\left(\frac{2}{\pi}\right) < \theta \le \arcsin\left(\frac{12+\pi^2}{8\pi}\right) + \frac{1}{2}\arcsin\left(\frac{12-\pi^2}{4\pi}\right)$. However, if λ exists, combining Lemma A.3 with equations (8.39) and (8.40) yields

$$\max W(\lambda) < \max W(\mu) .$$

Therefore, in order to compute $T_2(\theta)$ for $\theta \in \left(0,\frac{\pi}{2}\right]$, it remains to compare (8.39) with $T_1(\theta)$. In particular, for every sequence $\lambda = (\lambda_0, \lambda_1, \lambda_2, 0, \dots)$ in $D_2(\theta)$ with $\lambda_0 = \lambda_1 > \lambda_2 > 0$ the strict inequality $\max W(\lambda) < T_2(\theta)$ holds.

According to Lemma A.2, there is a unique $\vartheta \in \left(\arcsin\left(\frac{2}{\pi}\right), \frac{\pi}{2}\right)$ such that

$$\left(1 - \frac{2}{\pi}\sin\theta\right)^2 < \left(1 - \frac{2}{\pi}\sin\left(\frac{2\theta}{3}\right)\right)^3 \quad \text{for} \quad 0 < \theta < \vartheta$$

and

$$\left(1 - \frac{2}{\pi}\sin\theta\right)^2 > \left(1 - \frac{2}{\pi}\sin\left(\frac{2\theta}{3}\right)\right)^3 \quad \text{for} \quad \vartheta < \theta \le \frac{\pi}{2} .$$

These inequalities imply that ϑ is the unique solution to equation (8.38) in the interval $\left(0,\frac{\pi}{2}\right]$. Moreover, taking into account Lemma 8.18, equation (8.39), and the inequality $\vartheta > \arcsin\left(\frac{2}{\pi}\right)$, it follows that $T_1(\theta) < \max W(\mu)$

if and only if $\theta > \vartheta$. This proves the claimed representation for $T_2(\theta)$.

By Lemma 8.18 and the choice of ϑ it is obvious that the mapping $\big[0, \frac{\pi}{2}\big] \ni \theta \mapsto T_2(\theta)$ is strictly increasing, continuous on $\big[0, \frac{\pi}{2}\big]$, and continuously differentiable on $\big(0, \frac{\pi}{2}\big) \setminus \{\vartheta\}$. □

In order to prove Proposition 8.8, it remains to show that $T(\theta)$ coincides with $T_2(\theta)$.

Proposition 8.21. *For every $\theta \in \big[0, \frac{\pi}{2}\big]$ and $n \geq 2$ one has the identities $T(\theta) = T_n(\theta) = T_2(\theta)$.*

Proof. Since $T(0) = 0$, the case $\theta = 0$ is obvious. Let $\theta \in \big(0, \frac{\pi}{2}\big]$ be arbitrary. As a result of equation (8.18), it suffices to show that $T_n(\theta) = T_2(\theta)$ for all $n \geq 3$. Let $n \geq 3$ and let $\lambda = (\lambda_j) \in D_n(\theta) \setminus D_{n-1}(\theta)$. The objective is to show that $\max W(\lambda) < T_n(\theta)$.

First, assume that $\lambda_0 = \cdots = \lambda_n = \frac{1}{\pi}\sin\big(\frac{2\theta}{n+1}\big) > 0$. We examine the two cases $\lambda_0 < \frac{2}{\pi^2}$ and $\lambda_0 \geq \frac{2}{\pi^2}$. If $\lambda_0 < \frac{2}{\pi^2}$, then $2M(\lambda_0) < \arcsin\big(\frac{2}{\pi}\big)$. In this case, it follows from Lemma 8.18 that $\max W\big((\lambda_0, \lambda_0, 0, \dots)\big) < T_1(\tilde\theta)$ with $\tilde\theta = 2M(\lambda_0)$. Hence, by Lemma 8.16 one has $\max W(\lambda) < T_n(\theta)$. If $\lambda_0 \geq \frac{2}{\pi^2}$, then

$$(n+1)\arcsin\Big(\frac{2}{\pi}\Big) \leq 2(n+1)M(\lambda_0) = 2\theta \leq \pi\,,$$

which is possible only if $n \leq 3$, that is, $n = 3$. In this case, one has $\lambda_0 = \frac{1}{\pi}\sin\big(\frac{\theta}{2}\big)$. Taking into account representation (8.22), it follows from Lemma A.4 that

$$\max W(\lambda) = \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{2}{\pi}\sin\Big(\frac{\theta}{2}\Big)\Big)^4 < \frac{1}{2} - \frac{1}{2}\Big(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta}{3}\Big)\Big)^3 \leq T_2(\theta)\,.$$

Since $T_2(\theta) \leq T_n(\theta)$, one concludes that $\max W(\lambda) < T_n(\theta)$ again.

Now, assume that the sequence $\lambda = (\lambda_j) \in D_n(\theta) \setminus D_{n-1}(\theta)$ satisfies $\lambda_0 = \cdots = \lambda_{n-1} > \lambda_n > 0$. Since, in particular, $\lambda_{n-2} = \lambda_{n-1} > \lambda_n > 0$, Lemma 8.20 implies that

$$\max W\big((\lambda_{n-2}, \lambda_{n-1}, \lambda_n, 0, \dots)\big) < T_2(\tilde\theta) \quad \text{with} \quad \tilde\theta = \sum_{j=n-2}^{n} M(\lambda_j)\,.$$

It follows from Lemmas 8.14 and 8.16 that

$$\max W(\lambda) = \max W\bigl((\lambda_{n-2}, \lambda_{n-1}, \lambda_n, \lambda_0, \dots, \lambda_{n-3}, 0, \dots)\bigr) < T_n(\theta)\,,$$

that is, $\max W(\lambda) < T_n(\theta)$ once again.

Hence, by Lemmas 8.14, 8.15, and 8.17 and Corollary 8.19 the inequality $\max W(\lambda) < T_n(\theta)$ holds for all $\lambda \in D_n(\theta) \setminus D_{n-1}(\theta)$, which implies that $T_n(\theta) = T_{n-1}(\theta)$. Now the claim follows by induction. □

We close this section with the following observation, which, together with Remark 8.12 above, shows that the estimate from Theorem 8.9 is indeed stronger than the best previously known estimate from [8].

Remark 8.22. It follows from the previous considerations that

$$t < T(M_*(t)) \qquad \text{for} \qquad \frac{4}{\pi^2+4} < t \le c_*\,,$$

where $M_*\colon [0, c_*] \to \bigl[0, \frac{\pi}{2}\bigr]$ is the function from the bound (2.27). Indeed, for $\frac{4}{\pi^2+4} < t \le c_*$, set $\theta := M_*(t)$, and let $\lambda = (\lambda_j) \in D(\theta)$ be given according to Remark 8.3. In particular, one has $t = \max W(\lambda)$. One then observes by inspection that only $\lambda = \bigl(\frac{4}{\pi^2+4}, \frac{4}{\pi^2+4}, 0, \dots\bigr)$ for $t = \frac{8\pi^2}{(\pi^2+4)^2}$ is one of the critical points in Lemma 8.15, so that $\max W(\lambda) < T(\theta)$ for $t \neq \frac{8\pi^2}{(\pi^2+4)^2}$. If, however, $t = \frac{8\pi^2}{(\pi^2+4)^2}$, then $\theta = M_*(t) = \arcsin\bigl(\frac{4\pi}{\pi^2+4}\bigr) > \vartheta$ (cf. Remark 8.10), and it follows from Lemma 8.20 that $\max W(\lambda) \le T_1(\theta) < T_2(\theta)$. So, in either case one has $t = \max W(\lambda) < T(\theta) = T(M_*(t))$.

8.3 Off-diagonal perturbations

In this section, we discuss the particular case of off-diagonal perturbations V. In this situation one has additional a priori knowledge on the components of the spectrum of the perturbed operator $A + tV$, and the optimization problem from Definition 8.6 can be modified accordingly, see Definition 8.24 below. However, it turns out that the treatment of this problem is more difficult. Not only is an explicit representation of the critical points for the associated finite-dimensional problems not at hand, but also the supremum corresponding to (8.10) is not attained this time, see Definition 8.24 below and the discussion thereafter. In fact, the optimization problem for off-diagonal perturbations is not solved explicitly yet. Nevertheless, by a suitable choice of the parameters involved, a lower bound on the optimal constant $c_{\text{opt-off}}$ can be obtained that is stronger than the previously known bounds, see Example 8.25 below. A corresponding estimate on the maximal angle is also given there.

Throughout this section, assume, in addition to Hypothesis 8.1, that the perturbation V is off-diagonal with respect to the orthogonal decomposition $\mathcal{H} = \operatorname{Ran} \mathsf{E}_A(\sigma) \oplus \operatorname{Ran} \mathsf{E}_A(\Sigma)$. In this situation, we extend the definition of the operator B_t and the orthogonal projection P_t to parameters t satisfying $0 \le t < \sqrt{3}/2$. It follows from Proposition 1.21 that the spectrum of each B_t is separated as $\operatorname{spec}(B_t) = \omega_t \cup \Omega_t$ with

$$\omega_t = \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{\delta_t \cdot d}(\sigma)} \quad \text{and} \quad \Omega_t = \operatorname{spec}(B_t) \cap \overline{\mathcal{O}_{\delta_t \cdot d}(\Sigma)}\,,$$

where

$$\delta_t := t \tan\Bigl(\frac{1}{2}\arctan(2t)\Bigr) = \frac{1}{2}\sqrt{1+4t^2} - \frac{1}{2} < \frac{1}{2}\,.$$

In particular, for $0 \le t < \sqrt{3}/2$ one has $P_t = \mathsf{E}_{B_t}(\omega_t)$ and

$$\operatorname{dist}(\omega_t, \Omega_t) \ge (1 - 2\delta_t)d = \bigl(2 - \sqrt{1+4t^2}\bigr)d\,. \tag{8.41}$$

One of the main differences to the situation discussed in the preceding sections is that the function $\tau \mapsto 1 - 2\delta_\tau$ from the lower bound (8.41) is not affine, in contrast to the function $\tau \mapsto 1 - 2\tau$ in the case of general perturbations. This corresponds to the fact that for $B_t = B_s + (B_t - B_s)$ with $0 < s < t$ the perturbation $B_t - B_s$ does not need to be (and usually is not) off-diagonal with respect to $\mathcal{H} = \operatorname{Ran} P_s \oplus \operatorname{Ran} P_s^\perp$. As a consequence, an identity analogous to (8.8) is not at hand in the case of off-diagonal perturbations; recall that (8.8), in combination with Remark 7.7, justified the approach (8.7). The following lemma illustrates the importance of this observation very clearly.

Lemma 8.23.

(a) One has

$$\frac{\pi}{2}\int_0^t \frac{\mathrm{d}\tau}{1 - 2\delta_\tau} < \frac{1}{2}\arcsin(\pi t) \quad \textit{for} \quad 0 < t \le \frac{1}{\pi}\,.$$

(b) For every $0 < s < \sqrt{3}/2$ *there is* ε *with* $0 < \varepsilon \le (1 - 2\delta_s)/\pi$ *and*

$s+\varepsilon<\sqrt{3}/2$ *such that*

$$\frac{1}{2}\arcsin\Bigl(\pi\,\frac{t-s}{1-2\delta_s}\Bigr)<\frac{\pi}{2}\int_s^t\frac{\mathrm{d}\tau}{1-2\delta_\tau}\qquad\textit{whenever}\qquad s<t\le s+\varepsilon\,.$$

Proof. Let s with $0\le s<\sqrt{3}/2$ be arbitrary, and define

$$h_s(t):=\frac{\pi}{2}\int_s^t\frac{\mathrm{d}\tau}{1-2\delta_\tau}-\frac{1}{2}\arcsin\Bigl(\pi\,\frac{t-s}{1-2\delta_s}\Bigr)\,.$$

Taking into account that $1-2\delta_\tau=2-\sqrt{1+4\tau^2}$, one computes

$$h_s'(t)=\frac{\pi}{2}\biggl(\frac{1}{2-\sqrt{1+4t^2}}-\frac{1}{\sqrt{\bigl(2-\sqrt{1+4s^2}\,\bigr)^2-\pi^2(t-s)^2}}\biggr)\,.$$

First, suppose that $s=0$. In this case, the inequality $h_0'(t)<0$ is equivalent to $\sqrt{1-\pi^2t^2}<2-\sqrt{1+4t^2}$, and it is easy to verify that the latter is valid for $0<t\le 1/\pi$. Since $h_0(0)=0$, this implies that $h_0(t)<0$ for $0<t\le 1/\pi$, which proves (a).

Now, let $s>0$. In this case, the inequality $h_s'(t)>0$ is equivalent to

$$\bigl(2-\sqrt{1+4s^2}\bigr)^2-\pi^2(t-s)^2>\bigl(2-\sqrt{1+4t^2}\bigr)^2\,,$$

which, in turn, can be rewritten as

$$4(t^2-s^2)\Bigl(\frac{4}{\sqrt{1+4s^2}+\sqrt{1+4t^2}}-1\Bigr)>\pi^2(t-s)^2\,.$$

Dividing the latter inequality for $t>s$ by $t-s$ and then letting t approach s, one arrives at the inequality

$$8s\cdot\Bigl(\frac{2}{\sqrt{1+4s^2}}-1\Bigr)>0\,,$$

which is obviously valid for $0<s<\sqrt{3}/2$. Hence, by continuity, one concludes that $h_s'(t)>0$ if $t>s$ is sufficiently close to s. Since $h_s(s)=0$, this proves (b). □

Part (a) of the preceding lemma implies that the bound from Theorem 6.15 (b) is stronger than the one from Corollary 7.2 derived from the generic $\sin 2\Theta$ estimate. However, part (b) of Lemma 8.23 indicates that the situation somehow changes when the estimate on the maximal angle is iterated. More precisely, for given $0 < s < t < \sqrt{3}/2$, the bound on the maximal angle $\arcsin\bigl(\|P_t - P_s\|\bigr)$ obtained from Corollary 7.2 is more accurate than the one obtained from Theorem 6.15 (b), provided that t is sufficiently close to s. In fact, even if t is not close to s, a deeper analysis of the proof of Lemma 8.23 (b) shows that one has sufficient control over the quantity ε to find a finite partition $s = \tau_0 < \cdots < \tau_{l+1} = t$, $l \in \mathbb{N}_0$, such that

$$\frac{1}{2}\sum_{j=0}^{l} \arcsin\Bigl(\pi\,\frac{\tau_{j+1}-\tau_j}{1-2\delta_{\tau_j}}\Bigr) < \frac{\pi}{2}\int_s^t \frac{\mathrm{d}\tau}{1-2\delta_\tau}\,. \tag{8.42}$$

Let $t \in \bigl(0, \frac{\sqrt{3}}{2}\bigr)$ be arbitrary, and let $0 = t_0 < \cdots < t_{n+1} = t$, $n \in \mathbb{N}_0$, be a finite partition of the interval $[0,t]$. Analogously to (8.3), define

$$\lambda_j := \frac{t_{j+1}-t_j}{1-2\delta_{t_j}}\,, \quad j = 0,\dots,n\,.$$

Now, Lemma 8.23 (a) and inequality (8.42) motivate to consider the approach

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le \frac{\pi}{2}\int_0^{\lambda_0} \frac{\mathrm{d}\tau}{1-2\delta_\tau} + \frac{1}{2}\sum_{j=1}^{n} \arcsin(\pi\lambda_j)\,, \tag{8.43}$$

provided that $\lambda_0 = t_1 \le c_{\mathrm{off}}$ with c_{off} from Theorem 6.15 (b) and that $\lambda_j \le 1/\pi$ for $j = 1,\dots,n$.

Yet, there is one difficulty in this approach: Whenever $\lambda_0 > 0$, one can write

$$\frac{\pi}{2}\int_0^{\lambda_0} \frac{\mathrm{d}\tau}{1-2\delta_\tau} = \frac{\pi}{2}\int_0^{\lambda_0'} \frac{\mathrm{d}\tau}{1-2\delta_\tau} + \frac{\pi}{2}\int_{\lambda_0'}^{\lambda_0} \frac{\mathrm{d}\tau}{1-2\delta_\tau} \tag{8.44}$$

with $0 < \lambda_0' < \lambda_0$. In this case, one obtains a more accurate estimate in (8.43) if the summand $\frac{\pi}{2}\int_{\lambda_0'}^{\lambda_0} \frac{\mathrm{d}\tau}{1-2\delta_\tau}$ in (8.44) is replaced with a suitable term of the form of the left-hand side of (8.42). In other words, one can find another partition of the interval $[0,t]$ for which the corresponding estimate (8.43) is tighter. In fact, iterating this argument, it is easy to see that the inequality (8.42) can be ensured also if $s = 0$, so that one does not benefit substantially from using the bound from Theorem 6.15 (b) at all.

As a consequence, we do not modify the set $D(\theta)$ from Definition 8.6

for the optimization problem for off-diagonal perturbations, but only the operator W.

Definition 8.24. Let $\widetilde{W}$ denote the (non-linear) operator on D that assigns $\lambda = (\lambda_j) \in D$ the sequence (t_j) given by the recursion

$$t_{j+1} := t_j + \lambda_j(1 - 2\delta_{t_j})\,, \quad j \in \mathbb{N}_0\,, \quad t_0 = 0\,. \tag{8.45}$$

For $\theta \in \bigl[0, \frac{\pi}{2}\bigr]$ set

$$\widetilde{T}(\theta) := \sup\bigl\{\max \widetilde{W}(\lambda) \mid \lambda \in D(\theta)\bigr\}\,. \tag{8.46}$$

One can show in a way completely analogous to the proof of Lemma 8.5 that every sequence (t_j) defined by (8.45) is increasing and eventually constant with $0 \le t_j < \sqrt{3}/2$ for all $j \in \mathbb{N}_0$. In particular, the quantity $\max \widetilde{W}(\lambda) = \max_{j\in\mathbb{N}_0} t_j$ is well defined and less than $\sqrt{3}/2$. One can also show that the mapping $\bigl[0, \frac{\pi}{2}\bigr] \ni \theta \mapsto \widetilde{T}(\theta) \in \bigl[0, \frac{\sqrt{3}}{2}\bigr]$ is continuous and strictly increasing and that a corresponding variant of Proposition 8.7 is valid. We omit the details here. For now, it suffices to note that $\widetilde{T}\bigl(\frac{\pi}{2}\bigr)$ yields a lower bound on the optimal constant $c_{\text{opt-off}}$ introduced in Section 2.3, that is, one has $c_{\text{opt-off}} \ge \widetilde{T}\bigl(\frac{\pi}{2}\bigr)$.

It is a direct consequence of the preceding considerations (in particular Lemma 8.23 (a) and inequality (8.42) with $s = 0$) that, in contrast to the case of general perturbations, the supremum in (8.46) is *not* attained. This is one of the fundamental differences between the optimization problems defined by (8.10) and (8.46).

Another important difference between these two problems is that for $\lambda = (\lambda_j) \in D_n$ no explicit representation for $\max \widetilde{W}(\lambda)$ as in (8.19) is at hand, which makes it hard to characterize the corresponding critical points on $D_n(\theta)$ in form of an analogue to Lemma 8.15. Moreover, in contrast to Lemma 8.14, one has to expect that the value of $\max \widetilde{W}(\lambda)$ depends on the order of the entries $\lambda_0, \dots, \lambda_n$.

In fact, the optimization problem given by (8.46) is not solved explicitly yet. So far, the author can only guess a choice of $\lambda \in D\bigl(\frac{\pi}{2}\bigr)$ guaranteeing that $\widetilde{T}\bigl(\frac{\pi}{2}\bigr) > 0.694$, see Example 8.25 below. This guess is based on the approach (8.43). It seems to be a reasonable compromise between the complexity of the choice of the parameters and the strength of the result.

Example 8.25. Choose $a > 0$ such that

$$\frac{\pi}{2}\int_0^a \frac{\mathrm{d}\tau}{1-2\delta_\tau} = \frac{1}{3}$$

and $b \in \bigl(0, \frac{1}{\pi}\bigr]$ such that $2\arcsin(\pi b) = \frac{\pi}{2} - \frac{1}{3}$, that is,

$$b = \frac{1}{\pi}\sin\Bigl(\frac{3\pi-2}{12}\Bigr) = 0.1846204\ldots$$

Define $\lambda = (\lambda_j) \in D_4$ by

$$\lambda = (a, b, b, b, b, 0, \dots)\,.$$

For this choice of λ, the right-hand side of (8.43) equals $\pi/2$, and one easily concludes from inequality (8.42) (for $s = 0$) that there is $\theta < \pi/2$ and some $\mu \in D(\theta)$ with $\max\widetilde{W}(\mu) = \max\widetilde{W}(\lambda)$. Hence, $\widetilde{T}\bigl(\frac{\pi}{2}\bigr) > \widetilde{T}(\theta) \ge \max\widetilde{W}(\lambda)$.

We now estimate the value of $\max\widetilde{W}(\lambda) = t_5$, where $(t_j) := \widetilde{W}(\lambda)$. Numerical calculations give

$$t_1 = a \ge 0.2062031\,.$$

Upon observing that the mapping $\bigl[0, \frac{\sqrt{3}}{2}\bigr] \ni t \mapsto t + x(1-2\delta_t)$ is strictly increasing for all $0 \le x < 1/2$ (cf. Lemma 8.5 (a)), one verifies that

$$t_2 = t_1 + \lambda_1(1-2\delta_{t_1}) \ge 0.3757396$$

and, in the same way, that

$$t_3 \ge 0.5140409\,, \quad t_4 \ge 0.6184976\,,$$

and

$$t_5 \ge 0.6940725\,.$$

These considerations also yield a corresponding estimate on the maximal angle. Namely, in view of (8.43), one has

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le \frac{\pi}{2}\int_0^t \frac{\mathrm{d}\tau}{1-2\delta_\tau} \qquad \text{for} \qquad 0 \le t \le t_1\,,$$

and

$$\arcsin\bigl(\|P_0 - P_t\|\bigr) \le \frac{1}{3} + (j-1)\,\frac{3\pi - 2}{24} + \frac{1}{2}\arcsin\Bigl(\pi\,\frac{t - t_j}{1 - 2\delta_{t_j}}\Bigr)$$

for $t_j < t \le t_{j+1}$ with $j \in \{1, \dots, 4\}$, where we have taken into account that $\arcsin(\pi b) = (3\pi - 2)/12$. Numerical calculations suggest that this bound is indeed stronger than the one from Theorem 6.15 (b).

Corollary 8.26. *The optimal constant $c_{\text{opt-off}}$ for off-diagonal perturbations satisfies the lower bound $c_{\text{opt-off}} > 0.6940725$.*

Despite the mentioned difficulties of computing $\widetilde{T}(\theta)$, Example 8.25 and Corollary 8.26 demonstrate that the approach based on the optimization problem defined by (8.46) has great potential and deserves to be studied in future research.

8.4 Semidefinite perturbations. An outlook

In this final section, we briefly discuss how the optimization problem can be modified in the case of semidefinite perturbations; computing the corresponding solution is left for future studies.

To this end, in addition to Hypothesis 8.1, suppose that $V \ge 0$. The case $V \le 0$ can be treated analogously. We extend the definition of B_t to parameters t satisfying $0 \le t < 1$. Then, it follows from Proposition 1.20 that the spectrum of each B_t is separated as $\operatorname{spec}(B_t) = \omega_t \cup \Omega_t$ with ω_t and Ω_t defined analogously to (2.33). In particular, one has

$$\text{(8.47)} \qquad \operatorname{dist}(\omega_t, \Omega_t) \ge (1-t)d \quad \text{for} \quad 0 \le t < 1\,.$$

Since the inclusion $\omega_t \subset \mathcal{O}_{d/2}(\sigma)$ does not need to hold anymore for $t \ge 1/2$, we replace the definition of the spectral projection P_t by $P_t := \mathsf{E}_{B_t}(\omega_t)$, $0 \le t < 1$.

In view of (8.47), let $\widehat{W}$ denote the (non-linear) operator on D that assigns $\lambda = (\lambda_j) \in D$ the sequence (t_j) given by the recursion

$$t_{j+1} := t_j + \lambda_j(1 - t_j)\,, \quad j \in \mathbb{N}_0\,, \quad t_0 = 0\,.$$

The optimization problem for semidefinite perturbations can then be defined by

$$\widehat{T}(\theta) := \sup\bigl\{\max \widehat{W}(\lambda) \mid \lambda \in D(\theta)\bigr\}\,.$$

For $\lambda = (\lambda_j) \in D_n$, in a way analogous to (8.19), one obtains the explicit representation

$$\max \widehat{W}(\lambda) = t_{n+1} = 1 - \prod_{j=0}^{n}(1 - \lambda_j)\,.$$

In particular, Lemma 8.14 remains valid for $\widehat{W}$ instead of W. Moreover, from the explicit representation for t_{n+1}, one infers that the condition for the critical points of the mapping $\lambda \mapsto \max \widehat{W}(\lambda)$ on $D_n(\theta)$ analogous to (8.21) now reads

$$\frac{\sqrt{1 - \pi^2 \lambda_k^2}}{\sqrt{1 - \pi^2 \lambda_{k+1}^2}} = \frac{1 - \lambda_k}{1 - \lambda_{k+1}} \qquad \text{for} \qquad k = 0, \ldots, n-1\,.$$

This allows to obtain a corresponding variant of Lemma 8.15, which has been one of key ingredients in Section 8.2. The other main tool, Lemma 8.16, also remains available, provided that W is replaced with $\widehat{W}$. Thus, the optimization problem for semidefinite perturbations can be handled essentially in the same way as in the case of general perturbations. Yet, numerical experiments suggest that one will need more than three parameters for the solution, so that the problem here seems to be more involved than the one for general perturbations. The author hopes to return to this matter in future research.

Appendix A

Proof of some inequalities

Lemma A.1. *The following inequalities hold:*

(a) $\frac{2}{\pi^2} + \frac{2\pi-4}{\pi^2}\sin^2\theta < \frac{2}{\pi}\left(1 - \frac{1}{\pi}\sin\theta\right)\sin\theta$ *for* $\arcsin\left(\frac{1}{\pi-1}\right) < \theta < \frac{\pi}{2}$,

(b) $\frac{1}{\pi}\sin(2\theta) < \frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta$ *for* $\arctan\left(\frac{2}{\pi}\right) < \theta \le \frac{\pi}{4}$,

(c) $\frac{2}{\pi}\left(1 - \frac{1}{\pi}\sin\theta\right)\sin\theta < \frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta$ *for* $\theta \ne \arcsin\left(\frac{2}{\pi}\right)$,

(d) $\frac{2}{\pi}\left(1 - \frac{1}{\pi}\sin\theta\right)\sin\theta < \frac{1}{\pi}\sin(2\theta)$ *for* $0 < \theta < 2\arctan\left(\frac{1}{\pi}\right)$,

(e) $\frac{2}{\pi}\left(1 - \frac{1}{\pi}\sin\theta\right)\sin\theta > \frac{1}{\pi}\sin(2\theta)$ *for* $2\arctan\left(\frac{1}{\pi}\right) < \theta < \pi$.

Proof. One has

$$\begin{aligned}
\frac{2}{\pi}\Big(1 - \frac{1}{\pi}\sin\theta\Big)\sin\theta - \Big(\frac{2}{\pi^2} + \frac{2\pi-4}{\pi^2}\sin^2\theta\Big)& \\
&= -\frac{2(\pi-1)}{\pi^2}\Big(\sin^2\theta - \frac{\pi}{\pi-1}\sin\theta + \frac{1}{\pi-1}\Big) \\
&= -\frac{2(\pi-1)}{\pi^2}\Big(\Big(\sin\theta - \frac{\pi}{2(\pi-1)}\Big)^2 - \frac{(\pi-2)^2}{4(\pi-1)^2}\Big),
\end{aligned}$$

which is strictly positive if and only if

$$\Big(\sin\theta - \frac{\pi}{2(\pi-1)}\Big)^2 < \frac{(\pi-2)^2}{4(\pi-1)^2}\,.$$

A straightforward analysis shows that the last inequality holds for θ with $\arcsin\left(\frac{1}{\pi-1}\right) < \theta < \frac{\pi}{2}$. This proves (a).

For $\theta_0 := \arctan\left(\frac{2}{\pi}\right) = \frac{1}{2}\arcsin\left(\frac{4\pi}{\pi^2+4}\right)$ one has $\sin(2\theta_0) = \frac{4\pi}{\pi^2+4}$ and $\sin^2\theta_0 = \frac{4}{\pi^2+4}$. Thus, the inequality in (b) becomes an equality for $\theta = \theta_0$. Therefore, in order to show (b), it suffices to show that the corresponding estimate holds for the derivatives of both sides of the inequality, that is,

$$\frac{2}{\pi}\cos(2\theta) < \frac{\pi^2-4}{2\pi^2}\sin(2\theta) \quad \text{for} \quad \theta_0 < \theta < \frac{\pi}{4}\,.$$

This inequality is equivalent to $\tan(2\theta) > \frac{4\pi}{\pi^2-4}$ for $\theta_0 < \theta < \frac{\pi}{4}$, which, in turn, follows from $\tan(2\theta_0) = \frac{2\tan\theta_0}{1-\tan^2\theta_0} = \frac{4\pi}{\pi^2-4}$. This implies (b).

The claim (c) follows immediately from

$$\frac{2}{\pi^2} + \frac{\pi^2-4}{2\pi^2}\sin^2\theta - \frac{2}{\pi}\Big(1 - \frac{1}{\pi}\sin\theta\Big)\sin\theta = \frac{1}{2}\Big(\frac{2}{\pi} - \sin\theta\Big)^2 .$$

Finally, observe that

$$\text{(A.1)}\qquad \frac{1}{\pi}\sin(2\theta) - \frac{2}{\pi}\Big(1 - \frac{1}{\pi}\sin\theta\Big)\sin\theta = \frac{2}{\pi}\Big(\cos\theta - 1 + \frac{1}{\pi}\sin\theta\Big)\sin\theta .$$

For $0 < \theta < \pi$, the right-hand side of (A.1) is positive if and only if the term $\frac{1-\cos\theta}{\sin\theta} = \tan\big(\frac{\theta}{2}\big)$ is less than $\frac{1}{\pi}$. This is the case if and only if θ satisfies $\theta < 2\arctan\big(\frac{1}{\pi}\big)$, which proves (d). The proof of claim (e) is analogous. □

Lemma A.2. *There is a unique* $\vartheta \in \big(\arcsin\big(\frac{2}{\pi}\big), \frac{\pi}{2}\big)$ *such that*

$$\Big(1 - \frac{2}{\pi}\sin\theta\Big)^2 < \Big(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta}{3}\Big)\Big)^3 \qquad \textit{for} \quad 0 < \theta < \vartheta$$

and

$$\Big(1 - \frac{2}{\pi}\sin\theta\Big)^2 > \Big(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta}{3}\Big)\Big)^3 \qquad \textit{for} \quad \vartheta < \theta \le \frac{\pi}{2} .$$

Proof. Define $u, v, w\colon \mathbb{R} \to \mathbb{R}$ by

$$u(\theta) := \sin\Big(\frac{2\theta}{3}\Big), \quad v(\theta) := \frac{\pi}{2} - \frac{\pi}{2}\Big(1 - \frac{2}{\pi}\sin\theta\Big)^{2/3}, \quad w(\theta) := u(\theta) - v(\theta) .$$

Obviously, the claim is equivalent to the existence of $\vartheta \in \big(\arcsin\big(\frac{2}{\pi}\big), \frac{\pi}{2}\big)$ such that $w(\theta) < 0$ for $0 < \theta < \vartheta$ and $w(\theta) > 0$ for $\vartheta < \theta \le \frac{\pi}{2}$.

Observe that $u'''(\theta) = -\frac{8}{27}\cos\big(\frac{2\theta}{3}\big) < 0$ for $0 \le \theta \le \frac{\pi}{2}$. In particular, u'' is strictly decreasing on the interval $\big[0, \frac{\pi}{2}\big]$. Moreover, u''' is strictly increasing on $\big[0, \frac{\pi}{2}\big]$, so that the inequality $u''' \ge u'''(0) = -\frac{8}{27} > -\frac{1}{2}$ holds on $\big[0, \frac{\pi}{2}\big]$.

One computes

$$\text{(A.2)}\qquad v^{(4)}(\theta) = \frac{2\pi^{1/3}}{81}\,\frac{p(\sin\theta)}{(\pi - 2\sin\theta)^{10/3}} \qquad \text{for} \quad 0 \le \theta \le \frac{\pi}{2} ,$$

where

$$p(x) = 224 - 72\pi^2 + 27\pi^3 x - (160 + 36\pi^2)x^2 + 108\pi x^3 - 64x^4 .$$

The polynomial p is strictly increasing on $[0,1]$ and has exactly one root in the interval $(0,1)$. Combining this with equation (A.2), one obtains that $v^{(4)}$ has a unique zero in $\big(0, \frac{\pi}{2}\big)$ and that $v^{(4)}$ changes its sign from minus to plus there. Observing that $v'''(0) < -\frac{1}{2}$ and $v'''\big(\frac{\pi}{2}\big) = 0$, this yields $v''' < 0$

on $\big[0,\frac{\pi}{2}\big)$, that is, v'' is strictly decreasing on $\big[0,\frac{\pi}{2}\big]$. Moreover, it is easy to verify that $v'''\big(\frac{\pi}{3}\big) < v'''(0)$, so that $v''' \le v'''(0) < -\frac{1}{2}$ on $\big[0,\frac{\pi}{3}\big]$. Since $u''' > -\frac{1}{2}$ on $\big[0,\frac{\pi}{2}\big]$ as stated above, it follows that $w''' = u''' - v''' > 0$ on $\big[0,\frac{\pi}{3}\big]$, that is, w'' is strictly increasing on $\big[0,\frac{\pi}{3}\big]$.

Recall that u'' and v'' are both decreasing functions on $\big[0,\frac{\pi}{2}\big]$. Observing the inequality $u''\big(\frac{\pi}{2}\big) > v''\big(\frac{\pi}{3}\big)$, one deduces that

$$w''(\theta) = u''(\theta) - v''(\theta) \ge u''\Big(\frac{\pi}{2}\Big) - v''\Big(\frac{\pi}{3}\Big) > 0 \quad \text{for} \quad \theta \in \Big[\frac{\pi}{3},\frac{\pi}{2}\Big]. \tag{A.3}$$

Moreover, one has $w''(0) < 0$. Combining this with (A.3) and the fact that w'' is strictly increasing on $\big[0,\frac{\pi}{3}\big]$, one concludes that w'' has a unique zero in the interval $\big(0,\frac{\pi}{2}\big)$ and that w'' changes its sign from minus to plus there. Since $w'(0) = 0$ and $w'\big(\frac{\pi}{2}\big) = \frac{1}{3} > 0$, it follows that w' has a unique zero in $\big(0,\frac{\pi}{2}\big)$, where it changes its sign from minus to plus. Finally, observing that $w(0) = 0$ and $w\big(\frac{\pi}{2}\big) > 0$, in the same way one arrives at the conclusion that w has a unique zero $\vartheta \in \big(0,\frac{\pi}{2}\big)$ such that $w(\theta) < 0$ for $0 < \theta < \vartheta$ and $w(\theta) > 0$ for $\vartheta < \theta < \frac{\pi}{2}$. As a result of $w\big(\arcsin\big(\frac{2}{\pi}\big)\big) < 0$, one has $\vartheta > \arcsin\big(\frac{2}{\pi}\big)$. □

Lemma A.3. *For* $x \in \big(\frac{2}{\pi^2},\frac{4}{\pi^2+4}\big)$ *let*

$$\alpha := \frac{\sqrt{1-\pi^2x^2}}{1-2x} \quad \textit{and} \quad y := \frac{4\alpha^2}{\pi^2+4\alpha^2} - x\,. \tag{A.4}$$

Then, $\theta := \arcsin(\pi x) + \frac{1}{2}\arcsin(\pi y)$ *satisfies the inequalities*

$$\frac{3}{2}\arcsin\Big(\frac{2}{\pi}\Big) < \theta \le \arcsin\Big(\frac{12+\pi^2}{8\pi}\Big) + \frac{1}{2}\arcsin\Big(\frac{12-\pi^2}{4\pi}\Big) \tag{A.5}$$

and

$$\Big(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta}{3}\Big)\Big)^3 < (1-2x)^2(1-2y)\,. \tag{A.6}$$

Proof. One has $1 < \alpha < m := \frac{\pi}{2}\tan\big(\arcsin\big(\frac{2}{\pi}\big)\big)$, as well as $y \in \big(0,\frac{2}{\pi^2}\big)$ and $\alpha = \frac{\sqrt{1-\pi^2y^2}}{1-2y}$, cf. Lemma 8.15. Moreover, taking into account that $\alpha^2 = \frac{1-\pi^2x^2}{(1-2x)^2}$ by (A.4), one computes

$$y = \frac{4-(\pi^2+4)x}{\pi^2+4-4\pi^2x}\,. \tag{A.7}$$

Observe that $\alpha \to m$ and $y \to \frac{2}{\pi^2}$ as $x \to \frac{2}{\pi^2}$, and that $\alpha \to 1$ and $y \to 0$ as $x \to \frac{4}{\pi^2+4}$. With this and taking into account (A.7), it is convenient to consider $\alpha = \alpha(x)$, $y = y(x)$, and $\theta = \theta(x)$ as continuous functions of the variable $x \in \big[\frac{2}{\pi^2},\frac{4}{\pi^2+4}\big]$.

Straightforward calculations show that

$$1 - 2y(x) = \frac{\pi^2 - 4}{\pi^2 + 4 - 4\pi^2 x} \cdot (1 - 2x) \quad \text{for} \quad \frac{2}{\pi^2} \le x \le \frac{4}{\pi^2 + 4},$$

so that

$$y'(x) = -\frac{(\pi^2 - 4)^2}{(\pi^2 + 4 - 4\pi^2 x)^2} = -\frac{(1 - 2y(x))^2}{(1 - 2x)^2} \quad \text{for} \quad \frac{2}{\pi^2} < x < \frac{4}{\pi^2 + 4}.$$

Taking into account that $\alpha(x) = \alpha\big(y(x)\big)$, that is, $\frac{\sqrt{1-\pi^2 x^2}}{\sqrt{1-\pi^2 y(x)^2}} = \frac{1-2x}{1-2y(x)}$, this leads to

$$\begin{aligned} \theta'(x) &= \frac{\pi}{\sqrt{1 - \pi^2 x^2}} + \frac{\pi y'(x)}{2\sqrt{1 - \pi^2 y(x)^2}} \\ &= \frac{\pi}{2\sqrt{1 - \pi^2 x^2}} \left(2 + \frac{1 - 2x}{1 - 2y(x)} \cdot y'(x) \right) \\ &= \frac{\pi}{2\sqrt{1 - \pi^2 x^2}} \left(2 - \frac{\pi^2 - 4}{\pi^2 + 4 - 4\pi^2 x} \right) \\ &= \frac{\pi}{2\sqrt{1 - \pi^2 x^2}} \cdot \frac{12 + \pi^2 - 8\pi^2 x}{\pi^2 + 4 - 4\pi^2 x}. \end{aligned} \tag{A.8}$$

In particular, $x = \frac{12+\pi^2}{8\pi^2}$ is the only critical point of θ in the interval $\left(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\right)$ and θ' changes its sign from plus to minus there. Moreover, using $y\left(\frac{2}{\pi^2}\right) = \frac{2}{\pi^2}$ and $y\left(\frac{4}{\pi^2+4}\right) = 0$, one has

$$\theta\Big(\frac{2}{\pi^2}\Big) = \frac{3}{2} \arcsin\Big(\frac{2}{\pi}\Big) < \arcsin\Big(\frac{4\pi}{\pi^2 + 4}\Big) = \theta\Big(\frac{4}{\pi^2 + 4}\Big),$$

so that

$$\frac{3}{2} \arcsin\Big(\frac{2}{\pi}\Big) < \theta(x) \le \theta\Big(\frac{12 + \pi^2}{8\pi^2}\Big) \quad \text{for} \quad \frac{2}{\pi^2} < x < \frac{4}{\pi^2 + 4}.$$

Since $y\left(\frac{12+\pi^2}{8\pi^2}\right) = \frac{12-\pi^2}{4\pi^2}$, this proves the two-sided inequality (A.5).

Further calculations show that

$$\theta''(x) = \frac{\pi^3}{2} \frac{p(x)}{(1 - \pi^2 x^2)^{3/2} \, (\pi^2 + 4 - 4\pi^2 x)^2}, \quad \frac{2}{\pi^2} < x < \frac{4}{\pi^2 + 4}, \tag{A.9}$$

where

$$p(x) = 16 - 4\pi^2 + (48 + 16\pi^2 + \pi^4)x - 8\pi^2(12 + \pi^2)x^2 + 32\pi^4 x^3.$$

The polynomial p is strictly negative on the interval $\left[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\right]$, so that θ' is strictly decreasing.

Define $w\colon \big[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big] \to \mathbb{R}$ by

$$w(x) := (1-2x)^2 \cdot \big(1-2y(x)\big) - \left(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta(x)}{3}\Big)\right)^3 .$$

The claim (A.6) is equivalent to the inequality $w(x) > 0$ for $\frac{2}{\pi^2} < x < \frac{4}{\pi^2+4}$. Since $y\big(\frac{2}{\pi^2}\big) = \frac{2}{\pi^2}$ and, hence, $\theta\big(\frac{2}{\pi^2}\big) = \frac{3}{2}\arcsin\big(\frac{2}{\pi}\big)$, one has $w\big(\frac{2}{\pi^2}\big) = 0$. Moreover, a numerical evaluation gives $w\big(\frac{4}{\pi^2+4}\big) > 0$. Therefore, in order to prove $w(x) > 0$ for $\frac{2}{\pi^2} < x < \frac{4}{\pi^2+4}$, it suffices to show that w has exactly one critical point in the interval $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$ and that w takes its maximum there.

Using (A.8) and taking into account that $\sqrt{1-\pi^2x^2} = \alpha(x)(1-2x)$, one computes

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(1-2x)^2\big(1-2y(x)\big) &= -4(1-2x)\big(1-2y(x)\big) - 2(1-2x)^2 y'(x)\\ &= -2(1-2x)\big(1-2y(x)\big)\left(2 + \frac{1-2x}{1-2y(x)}\cdot y'(x)\right)\\ &= -\frac{4}{\pi}(1-2x)^2\big(1-2y(x)\big)\alpha(x)\theta'(x)\,.\end{aligned}$$

Hence, for $\frac{2}{\pi^2} < x < \frac{4}{\pi^2+4}$ one obtains

$$w'(x) = -\frac{4}{\pi}\theta'(x)\cdot\big(u(x) - v(x)\big)\,,$$

where $u, v\colon \big[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big] \to \mathbb{R}$ are given by

$$u(x) := \alpha(x)(1-2x)^2\big(1-2y(x)\big)$$

and

$$v(x) := \left(1 - \frac{2}{\pi}\sin\Big(\frac{2\theta(x)}{3}\Big)\right)^2 \cos\Big(\frac{2\theta(x)}{3}\Big)\,.$$

Suppose that for all $x \in \big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$ the difference $u(x) - v(x)$ is strictly negative. In this case, w' and θ' have the same zeros on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$, and $w'(x)$ and $\theta'(x)$ have the same sign for all $x \in \big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$. Combining this with (A.8), one concludes that $x = \frac{12+\pi^2}{8\pi^2}$ is the only critical point of w in the interval $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$ and that w takes its maximum in this point.

Hence, it remains to show that the difference $u - v$ is indeed strictly negative on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$. Since $\alpha\big(\frac{2}{\pi^2}\big) = \frac{\pi}{2}\tan\big(\arcsin\big(\frac{2}{\pi}\big)\big)$, $y\big(\frac{2}{\pi^2}\big) = \frac{2}{\pi^2}$, and $\theta\big(\frac{2}{\pi^2}\big) = \frac{3}{2}\arcsin\big(\frac{2}{\pi}\big)$, it is easy to verify that one has $u\big(\frac{2}{\pi^2}\big) = v\big(\frac{2}{\pi^2}\big)$ and $u'\big(\frac{2}{\pi^2}\big) = v'\big(\frac{2}{\pi^2}\big) < 0$. Therefore, it suffices to show that $u' < v'$ holds on the whole interval $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$.

One computes

$$u''(x) = \frac{(\pi^2-4)q(x)}{(1-\pi^2x^2)^{3/2}(\pi^2+4-4\pi^2x)^3} \tag{A.10}$$

where

$$q(x) = (128-80\pi^2-\pi^6) + 12\pi^2(\pi^2+4)^2x - 12\pi^2(7\pi^4+24\pi^2+48)x^2 + 32\pi^4(5\pi^2+12)x^3 + 24\pi^4(\pi^4+16)x^4 - 96\pi^6(\pi^2+4)x^5 + 128\pi^8x^6\,.$$

A further analysis shows that q'', which is a polynomial of degree 4, has exactly one root in the interval $\big[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big]$ and that q'' changes its sign from minus to plus there. Moreover, q' takes a positive value in this root of q'', so that $q' > 0$ on $\big[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big]$, that is, q is strictly increasing on this interval. Since $q\big(\frac{4}{\pi^2+4}\big) < 0$, one concludes that $q < 0$ on $\big[\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big]$. It follows from (A.10) that $u'' < 0$ on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$, so that u' is strictly decreasing. In particular, one has $u' < u'\big(\frac{2}{\pi^2}\big) < 0$ on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$.

A straightforward calculation yields

$$v'(x) = -\frac{2}{3}\Big(1-\frac{2}{\pi}\sin\Big(\frac{2\theta(x)}{3}\Big)\Big)\cdot\theta'(x)\cdot r\Big(\sin\Big(\frac{2\theta(x)}{3}\Big)\Big)\,, \tag{A.11}$$

where $r(t) = \frac{4}{\pi}+t-\frac{6}{\pi}t^2$. The polynomial r is positive and strictly decreasing on the interval $\big[\frac{1}{2},1\big]$. Furthermore, taking into account (A.5), $\theta(x)$ satisfies $\frac{1}{2} < \sin\big(\frac{2\theta(x)}{3}\big) < 1$. Combining this with equation (A.11), one deduces that $v'(x)$ has the opposite sign of $\theta'(x)$ for all $\frac{2}{\pi^2} < x < \frac{4}{\pi^2+4}$. In particular, by (A.8) it follows that $v'(x) \geq 0$ if $x \geq \frac{12+\pi^2}{8\pi^2}$. Since $u' < 0$ on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$, this implies that $v'(x) > u'(x)$ for $\frac{12+\pi^2}{8\pi^2} \leq x < \frac{4}{\pi^2+4}$. If $\frac{2}{\pi^2} < x < \frac{12+\pi^2}{8\pi^2}$, then one has $\theta'(x) > 0$. In particular, θ is strictly increasing on $\big(\frac{2}{\pi^2}, \frac{12+\pi^2}{8\pi^2}\big)$. Recall, that θ' is strictly decreasing by (A.9). Combining all this with equation (A.11) again, one deduces that on the interval $\big(\frac{2}{\pi^2}, \frac{12+\pi^2}{8\pi^2}\big)$ the function $-v'$ can be expressed as a product of three positive, strictly decreasing terms. Hence, on this interval v' is negative and strictly increasing. Recall that $u' < u'\big(\frac{2}{\pi^2}\big) = v'\big(\frac{2}{\pi^2}\big)$ on $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$, which now implies that

$$u'(x) < u'\Big(\frac{2}{\pi^2}\Big) = v'\Big(\frac{2}{\pi^2}\Big) < v'(x) \quad \text{for} \quad \frac{2}{\pi^2} < x < \frac{12+\pi^2}{8\pi^2}\,.$$

Since the inequality $u'(x) < v'(x)$ has already been shown for $x \geq \frac{12+\pi^2}{8\pi^2}$, one concludes that $u' < v'$ holds on the whole interval $\big(\frac{2}{\pi^2}, \frac{4}{\pi^2+4}\big)$. This completes the proof. □

Lemma A.4. *One has*

$$\left(1-\frac{2}{\pi}\sin\left(\frac{2\theta}{3}\right)\right)^3 < \left(1-\frac{2}{\pi}\sin\left(\frac{\theta}{2}\right)\right)^4 \qquad \textit{for} \quad 0<\theta\le\frac{\pi}{2}\,.$$

Proof. The proof is similar to the one of Lemma A.2. Define $u, v, w\colon \mathbb{R}\to\mathbb{R}$ by

$$u(\theta) := \sin\left(\frac{\theta}{2}\right), \quad v(\theta) := \frac{\pi}{2}-\frac{\pi}{2}\left(1-\frac{2}{\pi}\sin\left(\frac{2\theta}{3}\right)\right)^{3/4},$$

and

$$w(\theta) := u(\theta)-v(\theta)\,.$$

Obviously, the claim is equivalent to the inequality $w(\theta)<0$ for $0<\theta\le\frac{\pi}{2}$.

Observe that $u'''(\theta) = -\frac{1}{8}\cos\left(\frac{\theta}{2}\right) < 0$ for $0\le\theta\le\frac{\pi}{2}$. In particular, u''' is strictly increasing on $\left[0,\frac{\pi}{2}\right]$ and satisfies $u'''\ge u'''(0) = -\frac{1}{8}$.

One computes

$$\text{(A.12)} \qquad v^{(4)}(\theta) = \frac{\pi^{1/4}}{54}\,\frac{p\Bigl(\sin\left(\frac{2\theta}{3}\right)\Bigr)}{\Bigl(\pi-2\sin\left(\frac{2\theta}{3}\right)\Bigr)^{13/4}} \qquad \text{for} \quad 0\le\theta\le\frac{\pi}{2}\,,$$

where

$$p(x) = 45-16\pi^2+4\pi(1+2\pi^2)x-(34+20\pi^2)x^2+44\pi x^3-27x^4\,.$$

The polynomial p is strictly increasing on $\left[0,\frac{\sqrt{3}}{2}\right]$ and has exactly one root in the interval $\left(0,\frac{\sqrt{3}}{2}\right)$. Combining this with equation (A.12), one obtains that $v^{(4)}$ has a unique zero in the interval $\left(0,\frac{\pi}{2}\right)$ and that $v^{(4)}$ changes its sign from minus to plus there. Moreover, it is easy to verify that $v'''\left(\frac{\pi}{2}\right)< v'''(0) < -\frac{1}{8}$. Hence, one has $v''' < -\frac{1}{8}$ on $\left[0,\frac{\pi}{2}\right]$. Since $u'''\ge-\frac{1}{8}$ on $\left[0,\frac{\pi}{2}\right]$ as stated above, this implies that $w''' = u'''-v''' > 0$ on $\left[0,\frac{\pi}{2}\right]$, that is, w'' is strictly increasing on $\left[0,\frac{\pi}{2}\right]$.

With $w''(0) < 0$ and $w''\left(\frac{\pi}{2}\right) > 0$ one deduces that w'' has a unique zero in $\left(0,\frac{\pi}{2}\right)$ and that w'' changes its sign from minus to plus there. Since $w'(0)=0$ and $w'\left(\frac{\pi}{2}\right)>0$, it follows that w' has a unique zero in $\left(0,\frac{\pi}{2}\right)$, where it changes its sign from minus to plus. Finally, observing that $w(0)=0$ and $w\left(\frac{\pi}{2}\right)<0$, one concludes that $w(\theta)<0$ for $0<\theta\le\frac{\pi}{2}$. □

Bibliography

[1] V. M. Adamjan, H. Langer, *Spectral properties of a class of rational operator valued functions*, J. Operator Theory **33** (1995), 259–277.

[2] N. I. Akhiezer, I. M. Glazman, *Theory of Linear Operators in Hilbert Space*, Dover Publications, New York, 1993.

[3] A. B. Aleksandrov, V. V. Peller, *Operator Hölder-Zygmund functions*, Adv. Math. **224** (2010), 910–966.

[4] V. Adamjan, H. Langer, C. Tretter, *Existence and uniqueness of contractive solutions of some Riccati equations*, J. Funct. Anal. **179**, 2001, 448–473.

[5] S. Albeverio, K. A. Makarov, A. K. Motovilov, *Graph subspaces and the spectral shift function*, Canad. J. Math. **55** (2003), 449–503.

[6] S. Albeverio, A. K. Motovilov, *Operator Stieltjes integrals with respect to a spectral measure and solutions to some operator equations*, Trans. Moscow Math. Soc. **72** (2011), 45–77.

[7] S. Albeverio, A. K. Motovilov, *The a priori* $\tan\Theta$ *theorem for spectral subspaces*, Integral Equations Operator Theory **73** (2012), 413–430.

[8] S. Albeverio, A. K. Motovilov, *Sharpening the norm bound in the subspace perturbation theory*, Complex Anal. Oper. Theory **7** (2013), 1389–1416.

[9] W. Arendt, F. Räbiger, A. Sourour, *Spectral properties of the operator equation* $AX + XB = Y$, Q. J. Math. **45** (1994), 133–149.

[10] R. Bhatia, C. Davis, P. Koosis, *An extremal problem in Fourier analysis with applications to operator theory*, J. Funct. Anal. **82** (1989), 138–150.

[11] R. Bhatia, C. Davis, A. McIntosh, *Perturbation of spectral subspaces and solution of linear operator equations*, Linear Algebra Appl. **52/53** (1983), 45–67.

[12] R. Bhatia, P. Rosenthal, *How and why to solve the operator equation $AX - XB = Y$*, Bull. Lond. Math. Soc. **29** (1997), 1–21.

[13] M. S. Birman, M. Z. Solomjak, *Spectral Theory of Self-Adjoint Operators in Hilbert space*, Mathematics and its Applications (Soviet Series), D. Reidel Publishing Co., Dordrecht, 1987. Translation of the 1980 Russian original.

[14] M. S. Birman, M. Z. Solomyak, *Double operator integrals in a Hilbert space*, Integral Equations Operator Theory **47** (2003), 131–168.

[15] M. R. Bridson, A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Grundlehren der mathematischen Wissenschaften, vol. 319, Springer, Berlin, 1999.

[16] L. G. Brown, *The rectifiable metric on the set of closed subspaces of Hilbert space*, Trans. Amer. Math. Soc. **337** (1993), 279–289.

[17] Y. L. Daleckiĭ, S. G. Kreĭn, *Integration and differentiation of functions of Hermitian operators and applications to the theory of perturbations*, Amer. Math. Soc. Transl. Ser. 2 **47** (1965), 1–30. Translation of the 1956 Russian original.

[18] J. Daughtry, *Isolated solutions of quadratic matrix equations*, Linear Algebra Appl. **21** (1978), 89–94.

[19] C. Davis, *Separation of two linear subspaces*, Acta Sci. Math. (Szeged) **19** (1958), 172–187.

[20] C. Davis, *The rotation of eigenvectors by a perturbation*, J. Math. Anal. Appl. **6** (1963), 159–173.

[21] C. Davis, W. M. Kahan, *The rotation of eigenvectors by a perturbation. III*, SIAM J. Numer. Anal. **7** (1970), 1–46.

[22] I. C. Gohberg, M. G. Kreĭn, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Transl. Math. Monogr., vol. 18, Amer. Math. Soc., Providence, RI, 1969.

[23] L. Grubišić, V. Kostrykin, K. A. Makarov, K. Veselić, *The* $\tan 2\Theta$ *theorem for indefinite quadratic forms*, J. Spectr. Theory **3** (2013), 83–100.

[24] P. R. Halmos, *Two subspaces*, Trans. Amer. Math. Soc. **144** (1969), 381–389.

[25] T. Kato, *Perturbation Theory for Linear Operators*, Die Grundlehren der mathematischen Wissenschaften, vol. 132, Springer, Berlin, 1966.

[26] V. Kostrykin, K. A. Makarov, A. K. Motovilov, *On a subspace perturbation problem*, Proc. Amer. Math. Soc. **131** (2003), 3469–3476.

[27] V. Kostrykin, K. A. Makarov, A. K. Motovilov, *Existence and uniqueness of solutions to the operator Riccati equation. A geometric approach*, In: Advances in Differential Equations and Mathematical Physics (Birmingham, AL, 2002), Contemp. Math., vol. 327, Amer. Math. Soc., Providence, RI, 2003, pp. 181–198.

[28] V. Kostrykin, K. A. Makarov, A. K. Motovilov, *A generalization of the* $\tan 2\Theta$ *theorem*, In: Current Trends in Operator Theory and Its Applications, Oper. Theory Adv. Appl., vol. 149, Birkhäuser, Basel, 2004, pp. 349–372.

[29] V. Kostrykin, K. A. Makarov, *The singularly continuous spectrum and non-closed invariant subspaces*, In: Recent Advances in Operator Theory and Its Applications, Oper. Theory Adv. Appl., vol. 160, Birkhäuser, Basel, 2005, pp. 299–309.

[30] V. Kostrykin, K. A. Makarov, A. K. Motovilov, *On the existence of solutions to the operator Riccati equation and the* $\tan\Theta$ *theorem*, Integral Equations Operator Theory **51** (2005), 121–140.

[31] V. Kostrykin, K. A. Makarov, A. K. Motovilov, *Perturbation of spectra and spectral subspaces*, Trans. Amer. Math. Soc. **359** (2007), 77–89.

[32] H. Langer, C. Tretter, *Diagonalization of certain block operator matrices and applications to Dirac operators*, In: Operator Theory and Analysis (Amsterdam, 1997), Oper. Theory Adv. Appl., vol. 122, Birkhäuser, Basel, 2001, pp. 331–358.

[33] R. McEachin, *A sharp estimate in an operator inequality*, Proc. Amer. Math. Soc. **115** (1992), 161–165.

[34] R. McEachin, *Closing the gap in a subspace perturbation bound*, Linear Algebra Appl. **180** (1993), 7–15.

[35] R. McEachin, *Analyzing specific cases of an operator inequality*, Linear Algebra Appl. **208/209** (1994), 343–365.

[36] K. A. Makarov, A. Seelmann, *Metric properties of the set of orthogonal projections and their applications to operator perturbation theory*, e-print arXiv:1007.1575v1 [math.SP] (2010).

[37] K. A. Makarov, A. Seelmann, *The length metric on the set of orthogonal projections and new estimates in the subspace perturbation problem*, J. Reine Angew. Math. (2013). DOI: 10.1515/crelle-2013-0099

[38] K. A. Makarov, S. Schmitz, A. Seelmann, *Reducing graph subspaces and strong solutions to operator Riccati equations*, e-print arXiv:1307.6439 [math.SP] (2013).

[39] R. Mennicken, A. A. Shkalikov, *Spectral decomposition of symmetric operator matrices*, Math. Nachr. **179** (1996), 259–273.

[40] A. K. Motovilov, A. V. Selin, *Some sharp norm estimates in the subspace perturbation problem*, Integral Equations Operator Theory **56** (2006), 511–542.

[41] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Appl. Math. Sci., vol. 44, Springer, New York, 1983.

[42] V. Q. Phóng, *The operator equation $AX - XB = C$ with unbounded operators A and B and related abstract Cauchy problems*, Math. Z. **208** (1991), 567–588.

[43] M. Reed, B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1980.

[44] M. Reed, B. Simon, *Methods of Modern Mathematical Physics. IV. Analysis of Operators*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1978.

[45] F. Riesz, B. Sz-Nagy, *Vorlesungen über Funktionalanalysis*, Hochschulbücher für Mathematik, vol. 27, VEB Deutscher Verlag der Wissenschaften, Berlin, 1956. Translation from the French 2nd and 3rd editions.

[46] M. Rosenblum, *On the operator equation $BX - XA = Q$*, Duke Math. J. **23** (1956), 263–269.

[47] M. Rosenblum, *The operator equation $BX - XA = Q$ with selfadjoint A and B*, Proc. Amer. Soc. **20** (1969), 115-120.

[48] S. Schmitz, *Representation theorems for indefinite quadratic forms and applications*, Dissertation, Johannes Gutenberg-Universität Mainz, 2014.

[49] K. Schmüdgen, *Unbounded Self-Adjoint Operators on Hilbert Space*, Grad. Texts in Math., vol. 265, Springer, Dordrecht, 2012.

[50] A. Seelmann, *Notes on the $\sin 2\Theta$ theorem*, Integral Equations Operator Theory **79** (2014), 579–597.

[51] A. Seelmann, *On an estimate in the subspace perturbation problem*, e-print arXiv:1310.4360 [math.SP] (2013). (submitted)

[52] B. Sz.-Nagy, *Über die Ungleichung von H. Bohr*, Math. Nachr. **9** (1953), 255–259.

[53] B. Sz.-Nagy, *Bohr inequality and an operator equation*, In: Operators in Indefinite Metric Spaces, Scattering Theory and Other Topics (Bucharest, 1985), Oper. Theory Adv. Appl., vol. 24, Birkhäuser, Basel, 1987, pp. 321–327.

[54] C. Tretter, *Spectral Theory of Block Operator Matrices and Applications*, Imperial College Press, London, 2008.

[55] K. Veselić, *Spectral perturbation bounds for selfadjoint operators. I*, Oper. Matrices **2** (2008), 307–339.

[56] J. Weidmann, *Linear Operators in Hilbert Spaces*, Grad. Texts in Math., vol. 68, Springer, New York, 1980. Translation of the 1976 German original.

[57] J. Weidmann, *Lineare Operatoren in Hilberträumen. Teil 1. Grundlagen*, Mathematische Leitfäden, B. G. Teubner, Stuttgart, 2000.

[58] C. Wyss, *Hamiltonians with Riesz bases of generalised eigenvectors and Riccati equations*, Indiana Univ. Math. J. **60** (2011), 1723–1765.