Spectral Methods for Quantum Markov Chains

Oleg Szehr

Zentrum Mathematik M5

Technische Universität München

A thesis submitted for the degree of

Doctor rerum naturalium

2013 December

TECHNISCHE UNIVERSITÄT MÜNCHEN
ZENTRUM MATHEMATIK

Institut für mathematische Physik

SPECTRAL METHODS FOR QUANTUM MARKOV CHAINS

Oleg Szehr

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzende :
Univ.-Prof. Dr. Nina Gantert

Prüfer der Dissertation:
1. Univ.-Prof. Dr. Michael M. Wolf
2. Prof. Dr. David Pérez-García,
   Universidad Complutense de Madrid,
   Spanien
3. Univ.-Prof. Dr. Dr. Holger Boche

Die Dissertation wurde am 19.12.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 08.05.2014 angenommen.

Работа посвящается

В. П. Прохоровой и В. В. Прохорову

A soul in tension
That's learning to fly.
Condition grounded
But determined to try.
Can't keep my eyes
From the circling skies.
Tongue-tied and twisted
Just an earth-bound misfit,
I.

Acknowledgements

First and foremost, I would like to thank my advisor Prof. Michael Wolf for the supervision of this project and for always lending an ear to my endless questions and concerns. Without his dedicated support this work would not have been possible. I thank Prof. Michael Keyl for providing me with an introduction to the theory of C^*-algebras. I thank all co-authors of my publications for fruitful collaboration. For this dissertation I would like to thank my committee members, Prof. Holger Boche, Prof. Nina Gantert and Prof. David Pérez-García for their time, effort and helpful comments. I gratefully acknowledge the funding sources that made my dissertation possible, first of all project QCCC of the Elite Network of Bavaria. Particular thanks go to O. Zorina and my office mate D. Lercher.

List of contributed articles

I) O. Szehr, D. Reeb. and M.M. Wolf,
Spectral convergence bounds for classical and quantum Markov processes,
arXiv:1301.4827

II) O. Szehr and M.M. Wolf,
Perturbation bounds for quantum Markov processes and their fixed points,
J. Math. Phys. **54**, 032203 (2013)

III) O. Szehr
Eigenvalue estimates for the resolvent of a non-normal matrix,
EMS J. Spec. Theory., To appear, arXiv:1305.7208

IV) O. Szehr and M.M. Wolf,
Perturbation Theory for Parent Hamiltonians of Matrix Product States,
arXiv:1402.4175

V) F. Dupuis, O. Szehr and M. Tomamichel
A decoupling approach to classical data transmission over quantum channels,
IEEE TIT **60:3**, 1562 - 1572 (2014)

VI) M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel
On quantum Renyi entropies: a new definition and some properties,
J. Math. Phys. **54**, 122203 (2013)

Contents

1

Introduction

The theoretical understanding as well as the practical development of quantum information tasks build on our understanding of *quantum evolutions*. Significant technological challenges such as the construction of devices for quantum computation, quantum communication and quantum data storage boil down to our ability to engineer a certain quantum evolution. The power of the mentioned quantum processes and their superiority over their classical counterparts are expressions of the fact that the underlying evolution is subject to the laws of quantum mechanics rather than classical physics. A paradigmatic quantum evolution is the evolution of a *closed* system i.e. a system which is isolated ideally from any interaction with the rest of the universe. Such evolutions are modelled as unitary operators on some Hilbert space. The most basic model of a quantum computer as a pattern of wires and unitary gates (that are applied successively to the wires) falls into this class. However, in any *physical* scenario some suitable measurement will recover the influence of the environment and prove that the isolation of the system is not ideal. In fact, closed systems simply do not exist[1] and the unitary description is insufficient. *Quantum channels* (i.e linear, trace-preserving/unital and completely positive maps) constitute the most general framework to describe the evolution of a physical system within quantum mechanics. In particular, irreversible dynamics and the dynamics of *open* quantum systems are modelled by quantum channels. It is, however, crucial that the evolution itself does neither depend on the state

[1] It is a *philosophical* question if the universe as such constitutes a closed system and if the corresponding dynamics are unitary. However, this question cannot be answered by any experiment and hence it is irrelevant from the point of view of physics.

of the system nor on its history. The latter assumption is often made implicitly, when speaking about evolutions of open systems and the corresponding scenario is referred to as *Markovian*. For a time evolution this assumption entails that it can be appropriately described as a quantum Markov process. Such processes provide a natural generalization of the ubiquitous concept of a classical Markov chain to quantum mechanics. They can either arise naturally from relaxation or equilibration, or they may be engineered for instance for the purpose of dissipative quantum computation, dissipative quantum state preparation or quantum Metropolis sampling. With technological advance quantum effects occur and are exploited in an increasing number of real-world situations. Our theoretical understanding of quantum Markov processes, however, is much less developed than in the classical case. This is contrasted by the fact that ultimately nature is better described by the laws of quantum mechanics. It is presumable that quantum Markov processes will develop a similar universal importance as their classical counterparts. The purpose of the main research line of this dissertation is to contribute to their understanding.

The content is mathematical in nature but the ubiquitous occurrence of quantum evolutions in theoretical physics and the increasingly sophisticated methods of implementation demand for a more detailed understanding, also on the purely theoretical side.

1.1 Our contribution

This is a cumulative dissertation based on publications achieved during the period of the project. More than containing the treatment of one particular topic this transcript reflects the scientific work of the author during a period of about two and a half years. Although the emphasis lies on the study of spectral estimates in the theory of Markov processes some of the contributed articles treat different topics. Naturally, the referee process takes its time such that some of the contributed articles are already published in scientific journals while others are published on ArXiv and currently under review. We start with a sketch of our main contributions.

1. *Spectral convergence estimates for Markov chains: Article I*

   The main purpose of this work is to contribute to our understanding of quan-

tum channels and quantum Markov processes. We do this in providing the first rigorous and in-depth analysis of *spectral estimates* for quantum Markov evolutions and by applying our results to various problems ranging from quantum information theory to solid state physics. We introduce a new mathematical framework to the theory of quantum and classical Markov chains, which appears to be a powerful and elegant machinery for bounding norms of holomorphic functions of operators. With the techniques developed we obtain the *strongest* so far spectral convergence estimates for homogeneous Markov chains in the classical as well as the quantum case. We improve the known spectral convergence bounds significantly for instance by replacing an exponential with a polynomial dimension-dependent prefactor. The main bounds only rely on the spectrum and on no other assumption (like irreducibility, aperiodicity, detailed-balance, etc.). We emphasize that the developed methods are not merely extensions of an established framework for classical Markov chains but are new even to the extremely well developed classical theory. A core conceptual observation in our approach is to view a spectral estimate for a Markov transition map as an interpolation problem in a particular function space, the so-called Wiener algebra. The solution of this problem involves deep results from interpolation theory, functional analysis and operator theory and provides a nice tangible application for the abstract methods (cf. Chapter 3).

2. *Spectral stability estimates for Markov chains: Articles II, III*
   It is a well-studied question how sensitive the steady states of a classical homogeneous Markov chain are with respect to perturbations in the transition map (1, 2, 3, 4, 5). In contrast, this topic is essentially untouched in the context of quantum Markov chains, where stability estimates are necessary at least for the following reasons. In the theory of quantum phase transitions it was recently shown that phase transitions can occur as a result of a dissipative process. It is natural to ask for conditions under which a perturbation of the generator of a Markov process leads to a phase transition in the asymptotic state. In (6) novel schemes for the preparation of quantum states and for quantum computation are proposed. For state preparation the core idea is to engineer a dissipative quantum process that drives a certain many-body system to a desired steady state. In case

of quantum computation a dissipative process performs a given computation on the input. Clearly, in any practical implementation the underlying processes can be engineered only up to small errors. For this reason, it is crucial to provide a priori estimates for the quality of asymptotic states given a certain level of error in the transition map. We study two types of stability estimates in Article II. The first directly bounds the distance of asymptotic states in terms of the distance of transition matrices and resembles in flavour the classical approach of (1). This method relies on strong resolvent estimates, which we develop in Article III, see below. The second class of estimates bounds the distance of states resulting from quantum evolutions at finite time and extends the results of (3, 7) to the quantum context. This is motivated by the fact that in laboratory implementation only a finite period of time can be observed. Not only do our results provide the first rigorous discussion in the quantum case, but our methods also *significantly improve* established stability estimates (3, 7) for classical Markov chains.

Motivated by the aforementioned sensitivity analysis for asymptotic states, we study the problem of bounding the resolvent of an operator in Article III. In this work we provide the *strongest* spectral estimates for the resolvents of finite-dimensional Hilbert space contractions and power-bounded Banach space operators improving on previous work by E.B. Davies and B. Simon (8) as well as N. Nikolski (9). In the case of Hilbert space contractions we present explicit operators that achieve equality in our bounds. In this sense the estimates obtained are *optimal*.

3. *Stability of quantum matter: Article IV*

We contribute to the theory of quantum many-body systems and quantum phase transitions. We study the stability of quantum phases in the *parent Hamiltonian* model (10, 11), which appears to be one of the most important rigorous models for quantum many-body systems. This model describes the dynamics of a closed quantum spin chain (i.e. systems of numerous quantum particles located in a row) with local interaction. The parent Hamiltonian model is constructed with respect to a so-called *Matrix product state* in a way such that the letter naturally emerges as the ground state of the system. Matrix product states provide an efficient description of the states of quantum spin chains. With their local structure these

states provide a natural framework to study ground states arising from local particle interaction. The parent Hamiltonian model is important for the following reasons. On the one hand it contains the mathematical structure necessary for a rigorous analysis. On the other hand it is an extremely general model such that *basically all* models for quantum spin chains discussed in the literature fall into this class (11). We provide conditions under which certain properties of the model change smoothly when a perturbation is added that is a sum of local interactions. In particular, we prove that the system remains in the same ground state if the local interactions are sufficiently weak. This is particularly delicate when considering the limit of an infinite number of particles in the chain as in this case the total perturbation is not bounded. Our result is of great importance in terms of the physical relevance of the model, since the contrary assertion would predict that arbitrary small perturbations result in a phase transition of the quantum matter. Clearly, this is in conflict with physical observation. In addition, the parent Hamiltonian model is an idealization, which cannot capture the behaviour of a messy physical system. It is more natural to describe such systems using a perturbed model. Our result shows that if perturbed models are close to the original model then they are stable under sufficiently weak local perturbations, which augments the range where the parent Hamiltonian model can be applied. The physical implication of our result lies in a theoretical prediction that quantum matter remains stable against phase transitions if the added interaction is small enough.

4. *Transmission of classical data over noisy quantum channels: Article V*
   Remark: *This contribution was achieved in a scientific collaboration, where the author contributed but did* not *take the leading role.*
   Suppose we have a quantum channel and that we want to use this channel to send information from Alice (who has access to the input) to Bob (who receives the output system). To achieve this Alice encodes her message into the input of the channel and after its execution Bob applies a decoding operation to recover the original message. Generally, the noise of the channel results in a part of the message being sent to the environment and thus being lost from Bob's point of view. (See (12) for a detailed discussion of the data-transmission scenario.)

It is a basic task of (quantum) Shannon theory to investigate and quantify the theoretical limits for the transmission of data in this scenario. One of the most fruitful ideas that arose in this area of research in the past few years is that of *decoupling*: the fact that, in quantum mechanics, the absence of correlations between two systems implies perfect correlations of those two systems with a third one. (See (13) for an introduction.) In other words, if Alice and Bob do not share any correlations, then the system that purifies Alice's and Bob's quantum systems can be split in a part that posses full correlations with Alice and a part that is fully correlated with Bob. Based on the concept of decoupling we present a new approach to the problem of transmitting classical data over a quantum channel and present a new proof of the Holevo-Schumacher-Westmoreland theorem (14, 15).

5. *Quantum Rényi divergence: Article VI*

   Remark: *This contribution was achieved in a scientific collaboration, where the author contributed but did* not *take the leading role.*

   We introduce a new quantum generalization of the family of classical Rényi's divergences (16) and show that our quantum divergences exhibits various natural properties. In particular, we show that key properties of the classical divergences can be translated in natural quantum counterparts: We prove that our new quantity can only decrease under the action of a quantum channel on its arguments and show that our family of quantum divergences is monotonically increasing in the family parameter. We introduce derived conditional entropies via a specific optimization of our quantum divergence. The resulting entropic measures include the von Neumann entropy and the well-established quantum generalizations of min-, max- and collision-entropy as special cases (17, 18). Our conditional entropies satisfy a duality relation and naturally generalize Maassen and Uffink's uncertainty relation (19) to the situation when quantum side information present. Following conjectures in a previous version of our article, some of our results were achieved (and extended) in the following two independent contributions: R. Frank and E. Lieb (20) prove that our quantum divergences can only decrease under the action of a quantum channel and extend the parameter range

where this property holds. S. Beigi (21) proves monotonicity in the family parameter, monotonicity under the action of quantum channels and a duality relation for the conditional entropy. Moreover, M. Wilde, A. Winter and D. Yang have applied our quantum Rényi divergences to solve an open problem in quantum information theory (22) and M. Mosonyi and T. Ogawa provided an operational interpretation of our divergences as cut-off rates in the strong converse problem of hypothesis testing (23).

1.2 Outline

The chapters 2,3 and 4 serve to facilitate the access of the interested reader to the contributed articles. Chapter 2 contains an extremely brief introduction to quantum mechanics with particular eyesight to the theory of quantum Markovian evolutions and some glance at an algebraic formulation of quantum mechanics. Chapter 3 introduces the mathematical framework that is required in order to understand the contributions I and III. Important constituents are the theory of interpolation in function spaces and the theory of model operators, which was developed mostly in the 1960s and was promoted by various mathematicians including A. Beurling, H. Helson, B. Sz.-Nagy, C. Foiaş, D. Sarason, N. Nikolski. Chapter 3 does not contain any new mathematics (developed by the author), however the techniques introduced are new to quantum information theory and to the theory of Markov chains. One standard text (24) covering the topic of Chapter 3 refers to itself as "An easy reading''. Chapter 4 contains and introduction to the theory of Matrix product states and provides the background needed to access Article IV. While a general understanding of quantum mechanics, Chapter 2 is essential for all articles, the content of Chapter 3 and Chapter 4 is independent.

2

Quantum mechanics

2.1 The role of Mathematics in theoretical Physics

The aim of theoretical physics is to forecast, based on a developed understanding, the future behaviour of physical systems. While this is an intractably hard task in general, our description of idealized scenarios, so-called *experiments* is extremely successful. A theory by definition is a framework comprising a pattern of rules that allow the prediction of *measurement outcomes* of such experiments. The process of setting up a theory naturally goes along with *approximation* and *simplification*. The obvious reason is that we cannot describe everything. The role of mathematics in theoretical physics is to provide an efficient approximate description of physical observations. If one is faced with the task to describe in a most simple way (say with least Kholomogorov-complexity) the appearance of a tree, then one might choose the following: "In first approximation a tree looks like a cylinder on whose top there is placed a sphere." The mathematical structures (cylinder and sphere) occurring in this description are products of our demand for a simple description. Similarly, we can describe the dynamics of a stone bouncing against the trunk approximately as the scattering of a small sphere off a cylinder. The endless number of acting forces in the real world scenario can be approximated by neglecting all of them but the (homogeneous) gravitational force and a force the sphere experiences in a direction radially off the trunk. More generally, one chooses an approximation, which one optimizes according to given constraints (say computational power, desired precision, a priori knowledge) and after a computation obtains a prediction. The computation is nothing but an equivalent reformulation of

the given data. The validity of this prediction is however limited to a certain *parameter range*, where the initial approximation is sensible. If the results of our theoretical considerations are not satisfactory i.e. if the predicted outcomes are not close to the observed experimental reality, one goes back to the initial mathematical description and tries to refine it. The process of building up a *theory* thus consists of an iteration of three steps: First, setting up a mathematical approximation; second, computation and third, comparison of the obtained results to experimental outcomes. Theoretical physics is an optimization task, which iterates the above steps to recover better and better approximations of the observed reality. In this sense there might not be so much "Unreasonable [in the] Effectiveness of Mathematics in the Natural Sciences'' (25). The mathematical description is optimized such that it provides an effective approximation. However, the more precise the description should be, the more involved the mathematics must get.

In the following sections we present a brief introduction to quantum mechanics with a particular eyesight on the theory of C^*-algebras. We emphasize some of the deep connections between operator theory and the theory of quantum evolutions that might be useful for the further development of our contributions. In Section 3.4 we will study the dilation theory for Hilbert space contractions in a general operator-theoretic framework. We suspect that based on Stinespring's representation theorem a similar theory might be developed for quantum channels. However, a general C^*-algebraic formulation of the formalism will be inevitable for such an extension.

General introductions to quantum mechanics can be found in endless textbooks, university scripts and on Wikipedia. Here, we just mention the extraordinarily good writings (26, 27, 28). See (29, 30) for a general introduction to C^*-algebras and (31) for an introduction with emphasis on completely positive and bounded evolutions.

2.2 The static structure of quantum mechanics

A quantum mechanical *observable* is a quantity that represents a measurable property of a quantum system. It determines the set of *measurement outcomes* for an experiment on the system. The measurement process attaches an expectation value to the observable weighting the measurement outcomes with respect to some probability measure. The

measurement is associated with (a possibly zero) gain of information: The observer obtains an expectation value for a previously undetermined physical quantity.

In the mathematical formulation of quantum mechanics the observable is represented by a self-adjoint element $A = A^\dagger$ of a unital[1] C^*-algebra $\mathcal{A}$[2]. The possible measurement outcomes are elements of the spectrum of A. The measurement process is implemented via a *quantum state*, which maps the observable to its expectation value[3]. In other words, the quantum state is described via a linear functional $\omega \in \mathcal{A}^* := \{f : \mathcal{A} \to \mathbb{C}, \textit{linear}\}$, which takes observables A to their expectation values $\omega(A)$. To make sense of $\omega(A)$ as an expectation value we require ω to be normalized, $\omega(\mathbb{1}) = 1$ and positive $\omega(A^\dagger A) \geq 0$. It is not hard to show (29) that for any $\lambda \in [0, 1]$ and any two quantum states ω_1, ω_2 the linear combination $\lambda\omega_1 + (1 - \lambda)\omega_2$ is a quantum state again, that is the set of quantum states is convex. We call ω *pure* if $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$ with $\omega_1 \neq \omega_2$ implies that $\lambda \in \{0, 1\}$ and *mixed* otherwise. In quantum mechanics one commonly interprets a convex combination of pure states as a *probabilistic mixture*. If ω_1 and ω_2 are pure then in terms of expectation values $\omega = \lambda\omega_1 + (1-\lambda)\omega_2$ corresponds to the situation, where the underlying state is ω_1 with probability λ and ω_2 with probability $1 - \lambda$.

The famous Gelfand-Naimark theorem builds the bridge to the more familiar Hilbert space based formulation of quantum mechanics. It states that in a certain sense any C^*-algebra can be seen as an algebra of bounded operators on some Hilbert space. Recall that a $*$-homomorphism of C^*-algebras $\mathcal{A}, \mathcal{B}$ is a map $\pi : \mathcal{A} \to \mathcal{B}$ that is linear, multiplicative $\pi(AB) = \pi(A)\pi(B)$ and that commutes with the $\cdot^\dagger$-operation $\pi(A^\dagger) = \pi(A)^\dagger$ (29, 30). A $*$-homomorphism into the C^*-algebra $\mathcal{B}(\mathcal{H})$ of bounded operators on some Hilbert space is often called a *representation*.

Proposition 1 (Gelfand-Naimark (29, 32))**.** *For any C^*-algebra $\mathcal{A}$ there is a Hilbert space $\mathcal{H}$ and an isometric [4] $*$-isomorphism from $\mathcal{A}$ to a C^*-subalgebra of the bounded operators on $\mathcal{H}$.*

[1] This means that the C^*-algebra contains a unit element.

[2] We shall write $A^\dagger$ instead of A^*, which is the natural notation in the framework of C^*-algebras. However, $\cdot^\dagger$ is more common in quantum information theory.

[3] The splitting of the measurement process into notions of "states" and "observables" is not unique. However, this splitting is irrelevant since in the end the observer is left with an expectation value, which captures everything the theory can describe.

[4] A representation is isometric if and only if it is injective.

In other words, every C^*-algebra can be isometrically represented as an algebra of operators on Hilbert space. In our consecutive discussion we will not go into the details of the representation theory of C^*-algebras. We just note the particularly important special case, where $\mathcal{A}$ is a D^2-dimensional vector space and there is no non-trivial observable commuting with all elements of the observable algebra.[1] From a physical point of view these assumptions mean that all observables are confined to carry a finite number of measurement outcomes and that the description of the physical system is such that there is no environmental system, whose observables commute with all elements in the observable algebra. In the depicted scenario we can view $\mathcal{A}$ as an algebra of complex $D \times D$ matrices $\mathcal{M}_D$. This algebra carries a natural scalar product $\langle X|Y\rangle = \mathrm{tr}(X^\dagger Y)$ and it follows from Riesz' representation theorem (33) (or simple linear algebra) that any state can be written as $\omega(\cdot) = \mathrm{tr}(\rho\ \cdot)$ with a positive semi-definite and trace one matrix ρ. Hence, any quantum state can be identified with a so-called density matrix $\rho \in \mathcal{M}_D$. Pure quantum states correspond to rank-one projectors $\rho = \psi\psi^\dagger$, $\psi \in \mathbb{C}^D$ and are often identified with the vector ψ. In quantum mechanics the natural notation for an element of a Hilbert space is $|\psi\rangle$, whereas $\langle\psi|$ denotes the corresponding[2] element of the dual space $\mathcal{H}^*$. Hence, we will often write $\rho = \psi\psi^\dagger = |\psi\rangle\langle\psi|$ such that $\omega(A) = \mathrm{tr}(\rho|\psi\rangle\langle\psi|) = \langle\psi|A\psi\rangle$ and say that pure quantum states are given by a vector in a Hilbert space. The latter representation is possible in finite dimensions but in the general context of C^*-algebras *not every pure state can be written in this form*[3].

2.3 Quantum mechanical evolutions

To a physical evolution of a quantum system corresponds an evolution of expectation values of observables. Mathematically this can be implemented as a linear map $\mathcal{T}$ acting on observables such that the expectation value of A after the evolution is given by $\omega(\mathcal{T}(A))$. Note, that the same expectation value can be achieved via an evolution of the state ω, while A remains unchanged. The whole physical information is contained in the expectation values of observables such that both descriptions are equivalent. The

[1] More generally, any finite-dimensional C^*-algebra can be represented as a direct sum of matrix algebras.

[2] If $\mathcal{H}$ has infinite dimension this correspondence is due to Riesz-Fréchet representation theorem (33).

[3] States that have such structure are commonly referred to as *vector states* (29).

dynamical picture of evolving observables is referred to as the *Heisenberg picture* while in the *Schrödinger picture* the evolution of the system is reflected on the states. In finite dimensions this can be realized via a duality relation between maps describing the evolution on observables and states. We have that $\mathrm{tr}(\rho\mathcal{T}(A)) = \mathrm{tr}(\mathcal{T}^*(\rho)A)$, where $\mathcal{T}^*$ denotes the dual or adjoint of $\mathcal{T}$ and describes the evolution of ρ. Of course, $\mathcal{T}$ and $\mathcal{T}^*$ are mutually dual maps and their depiction as $\mathcal{T}$ and $\mathcal{T}^*$ is arbitrary[1]. In order for $\mathcal{T}$ (or $\mathcal{T}^*$) to correspond to a physically sensible evolution we have to make sure that $\mathcal{T}(A)$ is a valid quantum observable (or accordingly $\mathcal{T}^*(\rho)$ a valid quantum state). In quantum mechanics the following properties are required for $\mathcal{T} : \mathcal{A} \to \mathcal{A}$.

1. $\mathcal{T}$ is linear. This is an axiom of quantum mechanics, although intense research was put into a derivation of this property from physically motivated assumptions (34, 35, 36). It seems that the *static* structure of quantum mechanics sketched in Section 2.2 *implies* that $\mathcal{T}$ must be linear (36).

2. $\mathcal{T}$ is unital i.e. $\mathcal{T}(\mathbb{1}) = \mathbb{1}$. For finite systems one can equivalently demand that $\mathcal{T}^*$ preserves the trace $\mathrm{tr}(\mathcal{T}^*(A)) = \mathrm{tr}(A)\ \forall A \in \mathcal{A}$. This requirement states that after the evolution the probability of measuring any outcome is 1.

3. $\mathcal{T}$ is completely positive. The following definition makes this notion precise.

Definition 1 ((31))**.** Let $\mathcal{A}$ and $\mathcal{B}$ denote two C^*-algebras and let $\mathcal{T} : \mathcal{A} \to \mathcal{B}$ be linear. Let $\mathcal{M}_s(\mathcal{A})$ and $\mathcal{M}_s(\mathcal{B})$ denote $s \times s$ matrices with entries in $\mathcal{A}$ and $\mathcal{B}$, respectively. Consider the map $\mathcal{T}_s : \mathcal{M}_s(\mathcal{A}) \to \mathcal{M}_s(\mathcal{B})$ defined via

$$(\mathcal{T}_s(X))_{ij} = \mathcal{T}\left((X)_{ij}\right).$$

$\mathcal{T}$ is called positive if it maps positive elements[2] of $\mathcal{A}$ to positive elements of $\mathcal{B}$ and $\mathcal{T}$ is completely positive (CP) iff for any s the map $\mathcal{T}_s$ is positive.

There are positive maps that are not CP (28, 37). Completely positive and trace-preserving (CPTP) maps or completely positive and unital (CPU) maps are referred to as quantum channels in the realm of quantum information theory (12). In the context of matrix algebras $\mathcal{T}$ is CP iff $\mathcal{T}^*$ is. Assumption *3.* has the following *physical*

[1] In the attached contributions we mostly deal with finite dimensional systems and $\mathcal{T}$ will denote the evolution of the quantum *state*. However, in the more general context the choice above is natural.

[2] An element $X \in \mathcal{A}$ is called positive iff there is $Y \in \mathcal{A}$ with $X = Y^\dagger Y$. Equivalently, one can demand that $X = X^\dagger$ and the spectrum of X is real and non-negative.

interpretation. Consider a state ρ of a bipartite system and an evolution that acts as $\mathcal{T}$ on the first share of the system while the second share "remains frozen" i.e. an operator identity $\mathcal{I}$ describes the evolution of this subsystem. For any dimension of the Hilbert space of the auxiliary (frozen) system we require the outcome $(\mathcal{T} \otimes \mathcal{I})(\rho)$ of the total evolution to be a valid state, in particular to be positive semi-definite. In mathematical terms we demand that $\mathcal{T} \otimes \mathcal{I}_s$, where $\mathcal{I}_s$ denotes the identity operation on $s \times s$ is positive for any s, see Definition 1. Particular interest lies in the dual of maps of the form $\mathcal{T} \otimes \mathcal{I}$. The dual map $\mathrm{tr}_2(\cdot)$ defined via $\mathrm{tr}(X \mathrm{tr}_2(Y)) := \mathrm{tr}((\mathcal{T} \otimes \mathcal{I})(X)Y)$ for all X, Y is called the partial trace (over the second system). It has the physical interpretation of a process that erases or discards the second system (12, 28).

2.4 Norms and spectral properties

We denote by $\|\cdot\|$ any particular norm coming from a normed space (for example the C^*-algebra $\mathcal{A}$). For $X \in \mathcal{M}_D$ and $p \in [1, \infty]$ we denote by $\|X\|_p$ the Schatten p-norm of X. The Schatten p-norms naturally induce norms for linear maps (38) $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_{D'}$ (super-operators) we have the p-to-q-norms, which are defined as

$$\|\mathcal{T}\|_{p \to q} = \sup_{\substack{X \in \mathcal{M}_D \\ X \neq 0}} \frac{\|\mathcal{T}(X)\|_q}{\|X\|_p}.$$

However, the distances $\|\mathcal{T} - \mathcal{E}\|_{p \to q}$ do not capture the *physical* distinguishability of two quantum channels $\mathcal{T}, \mathcal{E}$. Surprisingly, distinguishing the channels $\mathcal{E}$, $\mathcal{T}$ may become easier when one applies them to one part of a bipartite state. We have that

$$\|\mathcal{T} - \mathcal{E}\|_{p \to q} \leq \|\mathcal{T} \otimes \mathcal{I} - \mathcal{E} \otimes \mathcal{I}\|_{p \to q},$$

where strict inequality may occur (12, 38, 39). To better describe the situation, where the channel is applied on part of a bipartite state one considers *stabilzed* norms defined by (39)

$$\sup_{s \in \mathbb{N}} \|\mathcal{T} \otimes \mathcal{I}_s\|_{p \to q}.$$

Particular interest lies in the stabilized $\|\cdot\|_{\infty \to \infty}$-norm. Operationally, it is the natural norm to measure how well two CPU maps $\mathcal{E}, \mathcal{T}$ can be distinguished in a statistical experiment (12, 39). The dual of this norm reflects the experimental distinguish-ability of

CPTP maps and is defined by the equation $\|\mathcal{T}\|_\diamond = \sup_{s\in\mathbb{N}} \left\|\mathcal{T}^\dagger \otimes \mathcal{I}_s\right\|_{\infty\to\infty}$ (see e.g. (40)). $\|\mathcal{T}\|_\diamond$ is the so-called *diamond*-norm. Interestingly, the stabilized $\|\cdot\|_{\infty\to\infty}$-norm has been studied in operator theory for different reasons over the past four decades (31).

Definition 2 ((31, 40))**.** In the set-up of Definition 1 consider the linear map $\mathcal{T}_s : \mathcal{M}_s(\mathcal{A}) \to \mathcal{M}_s(\mathcal{B})$. We call $\mathcal{T}$ *completely bounded* if

$$\|\mathcal{T}\|_{CB} := \sup_{s\in\mathbb{N}} \|\mathcal{T}_s\| < \infty,$$

where $\|\mathcal{T}_s\| = \sup_{X\in\mathcal{M}_s(\mathcal{A})} \frac{\|\mathcal{T}_s(X)\|}{\|X\|}$. $\|\cdot\|_{CB}$ is called the norm of *complete boundedness*.

To each linear map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_D$ we can assign a spectrum $\sigma(\mathcal{T})$ via the usual eigenvalue equation: we have $\lambda \in \sigma(\mathcal{T})$ if and only if there is $X \neq 0$ with $\mathcal{T}(X) = \lambda X$. The largest magnitude of all eigenvalues is the spectral radius μ. The spectral radius of any element $\mathcal{M}$ of a Banach algebra with norm $\|\cdot\|$ can be computed via Gelfand's formula $\mu = \lim_{k\to\infty} \left\|\mathcal{M}^k\right\|^{1/k}$ (41). For a CPU map $\mathcal{T}$ we have from the Russo-Dye theorem (31) that

$$\|\mathcal{T}\|_{\infty\to\infty} = \|\mathcal{T}(\mathbb{1})\|_\infty = \|\mathbb{1}\|_\infty = 1.$$

In particular, the spectral radius of a CPU (or CPTP) map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_D$ is given by $\lim_{k\to\infty} \left\|\mathcal{T}^k\right\|^{1/k} = 1$. Furthermore the identity $\mathcal{T}(\mathbb{1}) = \mathbb{1}$ implies that 1 is an eigenvalue of any CPU (or CPTP) map.

We write $m_\mathcal{T}$ for the minimal polynomial associated with $\mathcal{T}$ (i.e., the minimal degree, monic polynomial that annihilates $\mathcal{T}$, $m_\mathcal{T}(\mathcal{T}) = 0$) and $|m_\mathcal{T}|$ for the number of linear factors in $m_\mathcal{T}$. In many cases a more natural annihilator for $\mathcal{T}$ is the *Blaschke product*

$$B(z) = \prod_{m_\mathcal{M}} \frac{z - \lambda_i}{1 - \bar{\lambda}_i z},$$

where the product is taken over all i such that the linear factor $z - \lambda_i$ occurs in $m_\mathcal{T}$ *respecting* multiplicities. Thus, the numerator of B as defined here is exactly the minimal polynomial $m_\mathcal{T}$.

2.5 The representation theorems of Stinespring and Kraus

The concept of a completely positive map originates from the study of operator algebras. A core observation for the application of such maps in theoretical physics is Stinespring's representation theorem. To demonstrate full respect to Stinespring's work we formulate the theorem for general completely positive maps.

Theorem 2 ((31, 42))**.** *Let $\mathcal{A}$ be a C^*-algebra with unit, let $\mathcal{H}$ be a Hilbert space and let $\mathcal{B}(\mathcal{H})$ denote the bounded operators on $\mathcal{H}$. Let $\mathcal{T} : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ be a completely positive linear map. Then there exists a Hilbert space $\mathcal{K}$ a bounded linear map $V : \mathcal{H} \to \mathcal{K}$ and a $*$-homomorphism (a representation) $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ such that*

$$\mathcal{T}(A) = V^\dagger \pi(A) V \qquad \forall A \in \mathcal{A}.$$

If $\mathcal{T}$ is unital then $V^\dagger V = \mathbb{1}$, i.e. V is an isometry.

In case that $\mathcal{A}$ has finite dimension it is not hard to see (28, 31) that any representation is given by $\pi(A) = A \otimes \mathbb{1}$. Hence, for matrix algebras we have that

$$\mathcal{T}(A) = V^\dagger (A \otimes \mathbb{1}) V \qquad \forall A \in \mathcal{M}_D.$$

Taking adjoints in the above formula shows (see (43)) that any CPTP map can be written as

$$\mathcal{T}(\rho) = \mathrm{tr}_2 \left(U(\rho \otimes \phi) U^\dagger \right),$$

where U is a unitary, ϕ is a quantum state and $\mathrm{tr}_2(\cdot)$ denotes the partial trace over the second system. This equation has an obvious interpretation as a description of the evolution of a state of an *open system*: The action of the channel $\mathcal{T}$ consists of a joint unitary evolution of a system and its environment followed by the partial trace, which discards the environment. Although this interpretation is appealing some care has to be taken. The environment and in particular ϕ are not specified by Stinespring's representation theorem. Hence, in order to have a consistent interpretation it is necessary to add ad hoc experimental knowledge about the state of the environment. Note that only the *mathematical* existence of the unitary is guaranteed. This however must not correspond to a an actual *physical* evolution of a certain system. An important property of the Stinespring representation is that it is continuous. Two quantum channels are close if and only if one can find Stinespring representation that are close. Theorem 3 can be stated for general completely positive maps on C^*-algebras (44).

Theorem 3 ((45))**.** *Let $\mathcal{T}_1, \mathcal{T}_2$ be completely positive unital maps*

$$\mathcal{T}_1, \mathcal{T}_2 : \mathcal{M}_D \to \mathcal{M}_{D'}$$

with Stinespring isometries $V_1, V_2 : \mathbb{C}^{D'} \to \mathbb{C}^D \otimes \mathbb{C}^{D''}$. We have that

$$\inf_U \|(\mathbb{1} \otimes U) V_1 - V_2\|_\infty^2 \leq \|\mathcal{T}_1 - \mathcal{T}_2\|_{CB} \leq 2 \inf_U \|(\mathbb{1} \otimes U) V_1 - V_2\|_\infty ,$$

where the minimization is with respect to all unitaries on the extension space $\mathbb{C}^{D''}$.

Another frequently used representation theorem for completely positive maps is due to Kraus. We state the theorem for matrix algebras as the generalization to C^*-algebras is too technical for this exposition.

Theorem 4 ((46)). *The map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_{D'}$ is completely positive and linear if and only if there is a set of operators $\{V_i\}_{i=1,\dots,n}$ such that*

$$\mathcal{T}(X) = \sum_{i=1}^{n} V_i^\dagger X V_i \quad \forall X \in \mathcal{M}_D.$$

$\mathcal{T}$ is unital iff $\sum_{i=1}^n V_i^\dagger V_i = \mathbb{1}$. Two sets $\{V_i\}_{i=1,\dots,n}$ and $\{\tilde{V}_i\}_{i=1,\dots,n}$ represent the same map iff there is unitary U such that $V_i = \sum_k U_{ik}\tilde{V}_k$.

The operators V_i are commonly called *Kraus operators*. A classical stochastic matrix[1] $S \in \mathcal{M}_D(\mathbb{R}^{\geq})$ can be embedded into the framework of completely positive maps by fixing an orthonormal basis $|i\rangle_{i=1,\dots,D}$ and setting $\mathcal{T}(\cdot) = \sum_{ij}^D S_{ij}|i\rangle\langle j| \cdot |j\rangle\langle i|$. However, as is obvious from the above theorem not every quantum channel stems from a stochastic matrix. Hence, a quantum channel constitutes a concept strictly more general than that of a classical stochastic matrix.

2.6 Quantum Markov processes and the evolution of open quantum systems

One of the most fundamental and fruitful concepts in classical probability theory is that of a Markov chain. Whenever a probabilistic process does not depend on the current state or its past i.e. it has the *Markov property* it can be appropriately described as a Markov chain. Such processes occur in ubiquitous real-world scenarios ranging from the description of algorithms for Internet search over the description of stock markets to computational biology and theoretical physics to name just a few. The universality of the concept is accompanied by a significant effort to understand its mathematical structure, which led to an almost endless body of literature on the topic. We list a few standard references (47, 48, 49, 50, 51).
Quantum Markov processes constitute a natural analogue of the classical concept (28). They occur for example in the natural sciences, whenever the dynamics of a large and

[1] $S \in \mathcal{M}_D(\mathbb{R}^{\geq})$ is called a *stochastic* matrix iff $\sum_j S_{ij} = 1 \forall i$.

fast-mixing environment can be neglected; often times, the equilibration and thermalization of physical systems is being described in a Markovian way. In this work we will only be concerned with finite and homogeneous Markov chains. We formulate the definition for the Heisenberg picture. The Schrödinger picture analogue is obvious.

Definition 3. Let $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_D$ be a CPU map. A homogeneous quantum Markov chain is a semigroup $\{\mathcal{T}_t\}_{t\in I}$ of CPU maps with $\mathcal{T}_0 = \mathcal{I}$[1], where the parameter set I is either $\mathbb{R}^{\geq}$ or $\mathbb{N}$. If A_0 is the initial observable then the observable at time $t \in I$ is $A_t = \mathcal{T}_t(A_0)$.

If $I = \mathbb{N}$ we call $\{\mathcal{T}_t\}_{t\in I}$ a Markov chain in discrete time, while $I = \mathbb{R}^{\geq}$ corresponds to a Markov chain in continuous time. In discrete time $\mathcal{T}$ is referred to as the transition map of the Markov chain. The Markov property is reflected in the semigroup structure in Definition 3. For any times $s, t \in I$ we have that $\mathcal{T}_{s+t} = \mathcal{T}_s \cdot \mathcal{T}_t$, which implies that the state of the systems at time $t + s$ only depends on the state of the system at time s and the evolution $\mathcal{T}^t$. From a physical point of view it is natural to add the assumption that a Markov chain in continuous time $\mathcal{T}_t$ is *continuous* in t[2]. In this case a detailed characterization is as follows. (We focus on the Heisenberg picture but the analogous equation in Schrödinger picture is very common, too.)

Proposition 5 ((52))**.** *Let $\mathcal{T}$ be a CPU map and let $\{\mathcal{T}_t\}_{t\in\mathbb{R}^{\geq}}$ be a continuous semigroup with $\mathcal{T}_0 = \mathcal{I}$. Then there is a set of operators $\{L_j\} \subset \mathcal{M}_D$ and a Hermitian operator $H \in \mathcal{M}_D$ such that*

$$\mathcal{T}_t = e^{t\mathcal{L}}$$

with

$$\mathcal{L}(X) = i[H, X] + \sum_j L_j X L_j^\dagger - \frac{1}{2}\{L_j L_j^\dagger, X\} \quad \forall X \in \mathcal{M}_D,$$

where $[X, Y] = XY - YX$ and $\{X, Y\} = XY + YX$. In particular, the semigroup is differentiable.

Hence, the dynamical equation governing a Markovian evolution in the Heisenberg picture is

$$\frac{\mathrm{d}}{\mathrm{d}t}A = \mathcal{L}(A).$$

[1] Recall that $\mathcal{I}$ denotes the operator identity.

[2] In the infinite dimensional context this step is tricky as there are various inequivalent definitions for continuity.

Setting $L_j = 0$ we obtain the famous *Heisenberg equation* (26, p. 129)

$$\frac{\mathrm{d}}{\mathrm{d}t}A = i[H, A],$$

which describes the evolution of a *closed* quantum system in the Heisenberg picture. It is the quantum mechanical analogue of the classical Liouville equation (53, p. 344). H is called the *Hamiltonian* of the evolution. Eigenvalues of H are called energy levels and eigenvectors of H are called energy-eigenstates. The eigenstate corresponding to the smallest energy is called the ground-state. The total evolution of the system is *unitary* in the sense that

$$A(t) = e^{iHt}Ae^{-iHt}$$

solves the Heisenberg equation and e^{iHt} is a unitary operator (26). The remaining terms of $\mathcal{L}$ in Proposition 2.1 determine the *dissipative* part of the evolution. The physical interpretation of this decomposition is that a Markovian evolution of an open system can be seen as a "superposition" of a unitary and a dissipative constituent. The unitary part reflects a "coherent" evolution of a closed system while the dissipative term reflects the interaction with the environment. Again certain care has to be taken with this interpretation since the above decomposition is *not unique*. However, the choice can be made unique via a gauge transformation setting $\mathrm{tr}(L_j) = 0$.

2.7 Dissipative quantum state preparation and quantum computation

In this section we discuss specific quantum information theory tasks and shift to the Schrödinger picture as this is the natural language for the discussion. Under certain conditions (28, Proposition 7.5) a purely dissipative quantum evolution drives an arbitrary initial state to a steady state that only depends on the evolution. In fact the evolution can be engineered to prepare important classes of quantum states or to perform an arbitrary quantum computation (6). We consider a quantum system composed of a number of subsystems (e.g. qubits[1]) interacting with local environments giving rise to memoryless and time-independent dissipation processes. (See Chapter 4 for a rigorous

[1] A qubit is a quantum system, whose observable algebra is $\mathcal{M}_2(\mathbb{C})$.

discussion of quantum many particle systems.) The dynamical equation is $\frac{d}{dt}\rho = \mathcal{L}(\rho)$ with generator in Lindblad form

$$\mathcal{L}(X) = \sum_j L_j X L_j^\dagger - \frac{1}{2}\{L_j^\dagger L_j, X\} \quad \forall X \in \mathcal{M}_D. \tag{2.1}$$

In the following schemes the evolution is engineered to have a unique steady state, which shall encode the outcome of the computation or state preparation.

1. *Dissipative Quantum Computation (DQC):* A quantum circuit is a set of wires on which gates are applied. Each wire corresponds to a quantum system (e.g. qubit) and each gate corresponds to some unitary operation being applied to neighbouring wires. Hence, the computation can be modelled by a sequence of unitary gates $\{U_t\}_{t=1}^T$ acting locally on neighbouring sites. The outcome of the computation is

   $$\rho_T := U_T \cdot .... \cdot U_1 \rho U_1^\dagger \cdot .... \cdot U_T^\dagger.$$

   In (6) explicit operators L_j (Equation 2.1) are constructed with the following properties: The operators act locally (on neighbouring wires) and the evolution has a steady state, (1) that is unique; (2) that can be reached in a time $poly(T)$; (3) such that ρ_T can be extracted from it in a time $poly(T)$. This method defies some of the DiVincenzo-criteria for quantum computation as it neither requires state preparation, nor unitary dynamics (54). However, in the above sense it is nevertheless as powerful as standard quantum computation.

2. *Dissipative State Preparation (DSE):* We consider a quantum system with N particles on a lattice Λ in any dimension. (Chapter 4 contains in introduction to the topic.) We are interested in ground states of Hamiltonians H that are local and frustration-free. "Local" here means that H consists of a sum of Hamiltonians h_λ that act non-trivially only on a small set λ of sites (for example, nearest neighbours), $H = \sum_{\lambda \subset \Lambda} h_\lambda$. A ground state of H is called frustration-free if minimizes the energy of each h_λ individually, that is if ρ minimizes the energy of H then it minimizes the energy of h_λ for any λ. Ground states of such Hamiltonians constitute an extremely wide class. They include matrix product states (MPS) (10) and projected entangled-pair states (PEPS) (55) as for instance Kitaev's topological code (56). In (6) it is shown how to engineer a dissipative

processes (Equation 2.1), which prepares any ground state of a frustration free Hamiltonian as its steady state.

Both DQC and DSE are robust in the following sense: The dissipative process drives the system towards a steady state independent of perturbations of the system along the evolution. For a practical implementation of the described schemes it is however crucial to analyse

1. whether the evolution is stable with respect to perturbations *in the generator* $\mathcal{L}$ and

2. whether the computation or state preparation can be performed on a *reasonable time scale*.

The first point is crucial since in any practical implementation the generators will be furnished with errors that might accumulate along the way. The second point is crucial since the dimension of the state space in quantum mechanics is exponential in the number of underlying sites (qu-bits). For the schemes to be feasible one requires that the computation time should be polynomial in the number of underlying sites. So far such estimates only exist in extremely restricted special case and the general study is a hard task. At the moment of appearance of our contributions the first topic was essentially untouched territory. Motivated by the mentioned issues we study spectral stability and convergence estimates for quantum Markov processes. We develop a new theory for spectral estimates of Markov processes and approach the above points, although admittedly we cannot present a satisfactory treatment of DQC and DSE. The following chapter lays down the mathematics required.

3

Spectral estimates from Nevanlinna-Pick interpolation and Model Theory

In the articles I, II and III we employ recent[1] mathematical methods to derive norm estimates for functions of quantum channels. The techniques used are new to the quantum information theory community. The purpose of this chapter is to facilitate the access of the interested reader to those articles. The presented exposition mostly relies on the textbooks (24, 57, 58) and does not contain any scientific discoveries made by the author (although the selection and collocation of results did not occur in the literature so far). It contains a very brief introduction to a model theoretic approach to spectral estimates. The latter constitutes a powerful framework that enables us to obtain *spectral* estimates for functions of wide classes of operators, a particularly interesting instance of which constitute CPTP maps. We are looking for estimates of the type

$$\|f(X)\| \leq \Phi(\sigma(X), n),$$

where X denotes an operator, $\sigma(X)$ is its spectrum and n the dimension of the underlying space. For simplicity we restrict our attention to X acting on spaces of finite dimension, although under certain constraints the theory can be extended to infinitely

[1] The Nagy-Foiaş invariant subspace theory and their commutant lifting approach to interpolation theory originate in the 1960s.

dimensional spaces[1]. At the core of the presented approach lies the observation that the task of finding a spectral bound on $\|f(X)\|$ is related to a Nevanlinna-Pick interpolation problem in a certain class of functions. The latter problem was solved by G. Pick (for the space H^∞, see below) already in 1916 but until today the theory experienced a vivid development. The modern approach we take here is based on the theory of certain function spaces and operators thereon and was pioneered by B. Sz.-Nagy, C. Foiaş and D. Sarason.

3.1 The Nevanlinna-Pick interpolation problem

The classical Nevanlinna-Pick interpolation problem is the following task. Let $\mathbb{D} = \{z \in \mathbb{C} || z| < 1\}$ denote the open unit disk and let $Hol(\mathbb{D})$ be the set of holomorphic functions on $\mathbb{D}$. Suppose we are given distinct points $\{\lambda_i\}_{i=1,\dots,n} \subset \mathbb{D}$ and let $\{w_i\}_{i=1,\dots,n} \subset \mathbb{D}$. The problem is to find $f \in Hol(\mathbb{D})$ that interpolates the data i.e. satisfies

$$f(\lambda_i) = w_i \qquad 1 \leq i \leq n$$

such that

$$\sup_{z\in\mathbb{D}} |f(z)| \leq 1.$$

The second requirement only serves as a normalization of the problem and could be scaled to any constant other than 1. A necessary and sufficient condition to solve this problem was found by G. Pick in 1916 (59).

Proposition 6 (Pick (59))**.** *There exists $f \in Hol(\mathbb{D})$ such that $f(\lambda_i) = w_i$ for $1 \leq i \leq n$ and $\sup_{z\in\mathbb{D}} |f(z)| \leq 1$ if and only if the Pick matrix*

$$(P)_{ij} = \frac{1 - w_i\bar{w}_j}{1 - \lambda_i\bar{\lambda}_j}$$

is positive semi-definite $P \geq 0$. The function is unique if and only if the Pick matrix has not maximal rank.

Intuitively, one might suspect that our initial problem of estimating $\|f(X)\|$ in terms of the spectrum is related to a Nevanlinna-Pick problem. The reason is as follows. Suppose that X can be diagonalized. A simple way to define $f(X)$ is by diagonalizing X and by letting f act on its eigenvalues λ_i. Consequently, any function g that coincides

[1] In fact the extension to algebraic operators is trivial.

with f on the spectrum of X satisfies $f(X) = g(X)$. In order to compute $\|f(X)\|$ one can *minimize the norm* $\|g(X)\|$ *under the constraint that* $g(\lambda_i) = f(\lambda_i)$. This resembles "an operator version of the original Nevanlinna-Pick problem". We will make this precise in Section 3.10.

3.2 Function spaces

In our consecutive discussion we will be concerned with certain subspaces of $Hol(\mathbb{D})$, an important class of which constitute the Hardy spaces (60). For $p > 0$ they are defined as

$$H^p := \left\{ f \in Hol(\mathbb{D}) | \, \|f\|_{H^p}^p := \sup_{0 \leq r < 1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\phi})|^p \mathrm{d}\phi < \infty \right\},$$

and

$$H^\infty := \left\{ f \in Hol(\mathbb{D}) | \, \|f\|_{H^\infty} := \sup_{z \in \mathbb{D}} |f(z)| < \infty \right\}.$$

It is immediate from the definition that the spaces H^p are vector spaces, that the mapping $f \mapsto \|f\|_{H^p}$ is a norm for $p \geq 1$ and that $H^p \subset H^q$ for $p \geq q$.

Elements of Hardy spaces can be characterized by their boundary behaviour. We denote by $L^p(\mathbb{T})$ the usual complex L^p-space of the unit circle $\mathbb{T}$. For any $f \in H^p$ the limit

$$f_b(e^{i\phi}) := \lim_{r \to 1} f(re^{i\phi}) \tag{3.1}$$

exists for almost every ϕ (61). Furthermore we have the following.

Proposition 7 ((24, 61))**.** *Let f_b be as in Equation 3.1 and let $p \geq 1$. Then $f \in H^p$ is equivalent to $f_b \in L^p(\mathbb{T})$ and the Fourier coefficients $\hat{f}_b(n)$ of f_b satisfy $\hat{f}_b(n) = 0$ for $n < 0$. Moreover, $\|f_b\|_{L^p} = \|f\|_{H^p}$.*

Proposition 7 yields a canonical identification for $p \geq 1$ of the Hardy space with the subspace of L^p with vanishing Fourier coefficients for negative n,

$$H^p \equiv \left\{ f \in L^p \mid \hat{f}(n) = 0, \, n < 0 \right\}.$$

Under this identification the Taylor coefficients and the Fourier coefficients of $f \in H^p$ are treated equivalently. Finally, we will encounter the Wiener algebra, which is the subset of $Hol(\mathbb{D})$ of absolutely convergent Taylor series,

$$W := \{f = \sum_{k\geq 0} \hat{f}(k) z^k \mid \|f\|_W := \sum_{k\geq 0} |\hat{f}(k)| < \infty\}.$$

3.3 Shift-invariant subspaces of $L^2(\mathbb{T})$

In this section we introduce the multiplication operator on $L^2(\mathbb{T})$ and identify its invariant subspaces. The structure of these spaces is described by the famous Beurling-Helson Theorem. We denote by $L^2 = L^2(\mathbb{T})$ the Hilbert space of square-integrable functions with respect to normalized Lebesgue measure $\mathrm{d}m$

$$L^2 = \left\{ f : \mathbb{T} \to \mathbb{C} \; measurable \mid \|f\|_2^2 = \int_{\mathbb{T}} |f|^2 \mathrm{d}m < \infty \right\}.^1$$

If φ is such that $\varphi f \in L^2$ for any $f \in L^2$ we denote by $Mult_\varphi$ the multiplication operator

$$\begin{aligned} Mult_\varphi : L^2 &\to L^2 \\ f &\mapsto Mult_\varphi(f) := \varphi f. \end{aligned}$$

The multiplication operator by z is of particular importance. We reserve the notation

$$S := Mult_z$$

for this operator.

Recall that the list $\mathcal{B} = \{z^n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for L^2. Any $f \in L^2$ can be written with respect to $\mathcal{B}$ as

$$\begin{aligned} f &= \sum_{n\in\mathbb{Z}} \hat{f}(n) z^n, \\ \hat{f}(n) &= \int_{\mathbb{T}} f \bar{z}^n \mathrm{d}m. \end{aligned}$$

[1] In this section we slightly overload the notation for L^2. To achieve a clearer exposition we do not factor out the kernel of the L^2-seminorm. It would be too much to introduce a new notation to emphasize this.

This expansion yields a natural isomorphism $\mathcal{F}$ of L^2 with the space of square summable sequences $l^2(\mathbb{Z}) = \{(a_n)_{n\in\mathbb{Z}} \mid \sum_{n\in\mathbb{Z}} |a_n|^2 < \infty\}$

$$\begin{aligned} \mathcal{F} : L^2 &\to l^2(\mathbb{Z}) \\ f &\mapsto (\hat{f}(n))_{n\in\mathbb{Z}}. \end{aligned}$$

The action of S is realized on $l^2(\mathbb{Z})$ by a bilateral right shift of coefficients. For this reason S is simply called the *shift operator*.
We call a subspace $E \subset L^2$ invariant with respect to S if

$$SE \subset E.$$

We always will assume that E is closed. An important example of an invariant subspace is the Hardy space:

$$H^2 = \{f \in L^2 \mid \hat{f}(n) = 0,\ n < 0\} = \text{span}\{z^n \mid n \geq 0\},$$

where span$\{\cdot\}$ denotes the closed linear span of a list. The following theorem classifies the invariant subspaces of L^2 and reveals the extraordinary role played by H^2.

Theorem 8 (Beurling-Helson (62, 63))**.** *Let E be an invariant subspace of L^2 with respect to the multiplication operator S. Then there are only two possibilities.*

1. *If $SE = E$ then $E = \chi_\sigma L^2$, where σ is a measurable subset of $\mathbb{T}$ and χ_σ is its characteristic function.*
2. *If $SE \neq E$ then there exists a measurable function θ on $\mathbb{T}$ with $|\theta| = 1$ almost everywhere and $E = \theta H^2$.*

Hence, any invariant subspace of L^2, which is not mapped onto itself by S is equivalent to H^2 up to a function θ with $|\theta| = 1$ almost everywhere. We can use the above to classify the invariant subspaces of H^2.

Corollary 9 (Beurling)**.** *Let $E \neq \{0\}$ be an invariant subspace of H^2 with respect to the multiplication operator. Then there exists $\theta \in H^2$ with $|\theta| = 1$ almost everywhere on $\mathbb{T}$ with $E = \theta H^2$.*

A function $\theta \in H^2$ with the property $|\theta| = 1$ almost everywhere on $\mathbb{T}$ is called an *inner function.*

Proof. In view of Theorem 8 we exclude the possibility $SE = E$. For the sake of contradiction suppose that $z^{-1}E \subset E$. Choose $f \in E$, $f \neq 0$ and note that the latter implies that for some n we have $\hat{f}(n) \neq 0$. It follows that $z^{-(n+1)}f \in E$ by assumption. On the other hand $z^{-(n+1)}f \notin H^2$, a contradiction. It follows that $SE \neq E$ and by Theorem 8 there is θ with $|\theta| = 1$ almost everywhere and $E = \theta H^2$. Since $E \subset H^2$ it follows that $\theta \in H^2$. □

3.4 Operator dilation theory

We introduce the concept of a *dilation* of a Hilbert space operator and show a fundamental observation going back to D. Sarason, which characterizes when a certain operator is a dilation of another. This observation is at the core of a Hilbert space theoretic approach to Nevanlinna-Pick interpolation and plays a fundamental role in model theory more generally. We start by recalling some basic concepts from the theory of operators on Hilbert space. Let $X : \mathcal{H} \to \mathcal{H}$ be a linear operator on a Hilbert space $\mathcal{H}$. We denote the operator norm of X by

$$\|X\|_\infty := \sup_{\substack{v \in \mathcal{H} \\ v \neq 0}} \frac{\|Xv\|_2}{\|v\|_2},$$

where $\|v\|_2 = \sqrt{\langle v|v\rangle}$. We call X a *Hilbert space contraction* iff $\|X\|_\infty \leq 1$. If X is bounded (i.e. $\|X\|_\infty < \infty$) the *adjoint* of X is the unique operator $X^\dagger : \mathcal{H} \to \mathcal{H}$ with

$$\langle Av|w\rangle = \langle v|A^\dagger w\rangle \quad \forall v, w \in \mathcal{H}.$$

An operator $V : \mathcal{H} \to \mathcal{H}$ is called an *isometry* iff

$$\langle Vv|Vw\rangle = \langle v|w\rangle \quad \forall v, w \in \mathcal{H}.$$

V is called a *co-isometry*, iff $V^\dagger$ is an isometry. If an operator U is both an isometry and a co-isometry we call it *unitary*. With these notions defined we can proceed to the main topic of this section. We study dilations of operators on Hilbert space.

Definition 4. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces and let $\mathcal{H} \subset \mathcal{K}$. Let $X : \mathcal{H} \to \mathcal{H}$ and $Y : \mathcal{K} \to \mathcal{K}$ be bounded linear operators. We call Y a *dilation*[1] of X and equivalently X a *compression* of Y iff for every polynomial p we have $p(X) = P_\mathcal{H}\, p(Y)|_\mathcal{H}$, where $P_\mathcal{H}$ denotes the orthogonal projector in $\mathcal{K}$ on $\mathcal{H}$.

[1] Sometimes the Stinespring representation theorem is referred to as "Stinespring dilation theorem". In view of this (standard) definition the terminology is wrong.

Suppose we are given an operator $X : \mathcal{H} \to \mathcal{H}$ and let $Y : \mathcal{K} \to \mathcal{K}$ be an operator on the larger space $\mathcal{K}$. Clearly Y can only be a dilation of X if $X = P_{\mathcal{H}} Y|_{\mathcal{H}}$. However the latter condition is insufficient in general and a certain relation of $\mathcal{H}$ to the invariant subspaces of Y is required to ensure that Y is a dilation for X. Suppose we could decompose $\mathcal{K}$ such that

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

then X would be a compression of Y. Here X acts on a space which is the orthogonal difference of two invariant subspaces for Y. It is natural to ask if this is the most general structure such that Y is a dilation for X.

Theorem 10 (Sarason (64))**.** *Let $Y : \mathcal{K} \to \mathcal{K}$, let $\mathcal{H} \subset \mathcal{K}$ and let $X := P_{\mathcal{H}} Y|_{\mathcal{H}}$. Y is a dilation of X if and only if there exist two invariant subspace $\mathcal{H}_1, \mathcal{H}_2$ for Y such that $\mathcal{H}_1 \subset \mathcal{H}_2$ and $\mathcal{H} = \mathcal{H}_2 \ominus \mathcal{H}_1$*[1]*.*

Observe that X does not necessarily have to act on an invariant subspace of Y in order for Y to be a dilation of X. The latter situation would correspond to $\mathcal{H} = \mathcal{H}_2$ in Theorem 10. A stronger property of Y than being a dilation is being an extension.

Definition 5. Let Y be a dilation of X and suppose that X acts on an invariant subspace of Y. Then we call Y and *extension* of X or equivalently X a *part* of Y.

A paradigm of the study of dilations is the following theorem, which must be stated in this context.

Theorem 11 (Sz.-Nagy (65))**.** *Every contraction has a unitary dilation and a co-isometric extension.*

A powerful extension of this result is Andô's theorem, which generalizes Theorem 11 to the situation, where one is given a pair of commuting contractions.

Theorem 12 (Andô (66))**.** *Any two commuting contractions have commuting unitary dilations and commuting co-isometric extensions.*

Perhaps surprisingly there is no obvious generalization of Andô's theorem to three or more contractions. Parrott (67) constructed three commuting contractions that do not

[1] The difference $\mathcal{H}_2 \ominus \mathcal{H}_1$ is defined with respect to the scalar product from $\mathcal{K}$.

have commuting unitary dilations. The operator-theoretic significance of Theorem 12 together with Parrott's counterexample triggered further research in this direction. We list some references for the interested reader (68, 69, 70, 71).

The following theorem is the famous commutant lifting theorem by B. Sz.-Nagy and C. Foiaş. It will constitute the core building block for our solution of the Nevanilinna-Pick interpolation problem presented in Sections 3.7 and 3.9.

Theorem 13 (Commutant lifting theorem (72, 73))**.** *Let $T : \mathcal{H} \to \mathcal{H}$ be a contraction. Let $W : \mathcal{K} \to \mathcal{K}$ be a co-isometric extension of T. Let $X : \mathcal{H} \to \mathcal{H}$ be an operator that commutes with T. Then there is an operator Y such that*

1. $P_{\mathcal{H}} Y|_{\mathcal{H}} = X$
2. $YW = WY$
3. $\|Y\|_\infty = \|X\|_\infty$.

Proof. The commutant lifting theorem can be seen as a corollary of Andô theorem. It is not hard to see that any co-isometric extension W on $\mathcal{K}$ of a contraction T on $\mathcal{H}$ can be decomposed into a direct sum $W = W_{min} \oplus W'$, where W_{min} has the property

$$\mathcal{K} = \text{span}_n\{(W^*_{min})^n \mathcal{H} \mid n \in \mathbb{N}\}$$

and W' is another co-isometry (58). W_{min} is called the *minimal* co-isometric extension. Let us come back to the statement of Theorem 13. Without loss of generality we can assume that $\|X\|_\infty \leq 1$. (Otherwise rescale X dividing it through $\|X\|_\infty$.) By Andô's theorem there exist commuting co-isometric extensions say A and B for T and X, respectively. Suppose for the moment that W (as in Theorem 13) is a minimal co-isometric extension for X. Then $A = W \oplus A'$ for some co-isometry A'. Set $Y := P_{\mathcal{K}} B|_{\mathcal{K}}$ and observe that $P_{\mathcal{H}} Y|_{\mathcal{H}} = X$ that $YW = WY$ and that $\|Y\|_\infty = 1$. Hence, we have proven the theorem in the special case that W is minimal. In general $W = W_{min} \oplus W'$ and we set $Y = P_{\mathcal{H}_{min}} B|_{\mathcal{H}_{min}} \oplus \mathbf{0}$, where $\mathcal{H}_{min}$ is the subspace on which W_{min} acts. □

3.5 Kernels and multiplication operators for H^2

For any $\zeta \in \mathbb{D}$ the evaluation map

$$\begin{aligned} eval_\zeta : \; & H^2 \to \mathbb{C} \\ & f \mapsto eval_\zeta(f) := f(\zeta) \end{aligned}$$

is a continuous, linear functional on H^2. Hence, by the Riesz representation theorem (33) there is $k_\zeta \in H^2$ such that

$$f(\zeta) = \langle f | k_\zeta \rangle. \tag{3.2}$$

In the context of general Hilbert function spaces[1] Equation 3.2 is referred to as the *reproducing property* of the *reproducing kernel* k_ζ. In the particular case of H^2 we call k_ζ the Cauchy kernel. It is not hard to verify that $k_\zeta(z) = \frac{1}{1-\bar{\zeta}z}$. We have

$$\langle f | k_\zeta \rangle = \int_{\partial\mathbb{D}} f(z)\overline{k_\zeta}\,\frac{|\mathrm{d}z|}{2\pi} = \int_{\partial\mathbb{D}} f(z)\frac{z}{z-\zeta}\,\frac{|\mathrm{d}z|}{2\pi} = \frac{1}{2\pi i}\int_{\partial\mathbb{D}} \frac{f(z)}{z-\zeta}\mathrm{d}z = f(\zeta)$$

by the Residue theorem.

The rest of this section is devoted to the study of restrictions of multiplication operators (as in Section 3.3) to H^2. We slightly overload the notation and also write $Mult_\varphi$ for the multiplication operator on H^2. In this case the shift operator acts on a sequence-space $l^2(\mathbb{N})$ as an unilateral right shift of coefficients. A first question to answer in this context is for which functions φ is $Mult_\varphi : H^2 \to H^2$ an endomorphism. The *multiplier algebra* of H^2 by definition is the set of functions φ on $\mathbb{D}$ given by

$$\{\varphi \mid \varphi f \in H^2, \quad \forall f \in H^2\}.$$

To begin with, for $f \in H^2$ and $\varphi \in H^\infty$ the estimate

$$\|\varphi f\|_{H^2} \leq \left(\int_{\partial\mathbb{D}} |\varphi|^2 |f|^2 \frac{|\mathrm{d}z|}{2\pi}\right)^{1/2} \leq \|\varphi\|_{H^\infty} \|f\|_{H^2}$$

implies that any $\varphi \in H^\infty$ is contained in the multiplier algebra of H^2. This allows one to study for $\varphi \in H^\infty$ the action of the adjoint multiplication operator $Mult^*_\varphi$ on the reproducing kernels, which is of particular interest. Observe that for any f we have

$$\langle f | Mult^*_\varphi k_\zeta \rangle = \langle Mult_\varphi f | k_\zeta \rangle = \langle \varphi f | k_\zeta \rangle = \varphi(\zeta) f(\zeta) = \varphi(\zeta)\langle f | k_\zeta \rangle = \langle f | \overline{\varphi(\zeta)} k_\zeta \rangle.$$

It follows that k_ζ is eigenvector of $Mult^*_\varphi$ with eigenvalue $\overline{\varphi(\zeta)}$

$$Mult^*_\varphi k_\zeta = \overline{\varphi(\zeta)} k_\zeta.$$

[1] A Hilbert function space is a Hilbert space that consists of functions, such that the evaluation map is a non-zero, continuous functional. In particular Riesz' representation theorem applies.

For the operator norm this entails the following estimate

$$\|Mult_\varphi\|_\infty \geq \sup_{\zeta\in\mathbb{D}} |\varphi(\zeta)| = \|\varphi\|_{H^\infty}$$

The latter holds for multipliers in general Hilbert functions spaces, with $\mathbb{D}$ replaced with the appropriate domain of functions. But in H^2 we can say more, namely that (as above)

$$\|Mult_\varphi\|_\infty = \sup_{\substack{f\in H^2 \\ \|f\|_{H^2}=1}} \|Mult_\varphi f\|_{H^2} = \sup_{\substack{f\in H^2 \\ \|f\|_{H^2}=1}} \left(\int_{\partial\mathbb{D}} |\varphi|^2|f|^2 \frac{|\mathrm{d}z|}{2\pi}\right)^{1/2} \leq \|\varphi\|_{H^\infty}\,.$$

In total we have found

$$\|Mult_\varphi\|_\infty = \|\varphi\|_{H^\infty}\,. \tag{3.3}$$

On the other hand if φ is in the multiplier algebra of H^2 then clearly φ must itself be in H^2. The above implies that φ is bounded and therefore in H^∞.

Proposition 14 ((58))**.** *The algebra H^∞ is isometrically isomorphic to the multiplier algebra of H^2.*

3.6 The compressed shift operator

We use Theorem 10 in combination with Beurling's Theorem 9 to study compressions of the multiplication operator on H^2.

Definition 6. Let θ be an inner function. The *model space* is defined as

$$\mathcal{K}_\theta := H^2 \ominus \theta H^2.$$

The *model operator* is defined as

$$\begin{aligned} M_\theta : \mathcal{K}_\theta &\to \mathcal{K}_\theta \\ f &\mapsto M_\theta f := P_{\mathcal{K}_\theta}(zf), \end{aligned}$$

where $P_{\mathcal{K}_\theta}$ denotes the orthogonal projection onto $\mathcal{K}_\theta$.

Theorem 15. *Let θ be an inner function and let p be a polynomial. Then for any $f \in \mathcal{K}_\theta$ we have $p(M_\theta)f = P_{\mathcal{K}_\theta}(p(S)f) = P_{\mathcal{K}_\theta}(pf)$. In addition $\|p(M_\theta)\|_\infty = \|P_{\mathcal{K}_\theta}p(S)|_{\mathcal{K}_\theta}\|_\infty \leq \|p\|_{H_\infty}$.*

Proof. By Theorem 8 H^2 and θH^2 are invariant subspaces of L^2. By definition $M_\theta = P_{\mathcal{K}_\theta} S|_{\mathcal{K}_\theta}$. Thus it follows from Theorem 10 that S is a dilation of M_θ to the whole space L^2 i.e. $p(M_\theta) = P_{\mathcal{K}_\theta} p(S)|_{\mathcal{K}_\theta}$. □

Theorem 15 is important for several reasons. First, the theorem gives a meaning to the model operator a as compression of the shift operator onto the model subspace. Second, the property $p(M_\theta) = P_{\mathcal{K}_\theta} p(S)|_{\mathcal{K}_\theta}$ greatly facilitates the computation of matrix entries of $p(M_\theta)$ as for any orthonormal vectors $\{e_k\}_k$ in $\mathcal{K}_\theta$

$$(p(M_\theta))_{ij} = \langle e_i | p(M_\theta) e_j \rangle = \langle e_i | p e_j \rangle.$$

Lastly, we have already pointed out in Proposition 14 that the multiplier algebra of H^2 is H^∞. The commutant of the model operator will play a fundamental role in our approach. For this reason we need to give a meaning to $\varphi(M_\theta)$ for arbitrary $\varphi \in H^\infty$. This can be achieved via Theorem 15.

Definition 7 (Sz.-Nagy and Foias (65))**.** Let $\varphi \in H^\infty$. We define $\varphi(M_\theta)$ by the relation

$$\varphi(M_\theta) f := P_{\mathcal{K}_\theta}(\varphi f) \qquad \forall f \in \mathcal{K}_\theta.$$

3.7 Lifting the commutant of M_θ

The *commutant* of M_θ is the set of all linear operators $X : \mathcal{K}_\theta \to \mathcal{K}_\theta$ with

$$X M_\theta = M_\theta X. \tag{3.4}$$

It is clear that any operator $\varphi(M_\theta)$ with $\varphi \in H^\infty$ commutes with M_θ. It turns out that the converse is also true i.e. if Equation 3.4 holds then $X = \varphi(M_\theta)$. The reason is that one can *lift* the commutation relation 3.4 to H^2 using the Commutant Lifting Theorem 13 and exploit that the multiplier algebra of H^2 is H^∞.
To be more concrete let us consider the adjoint commutation relation

$$X^\dagger M_\theta^\dagger = M_\theta^\dagger X^\dagger, \tag{3.5}$$

which is equivalent to Equation 3.4. We note that $\left\| M_\theta^\dagger \right\|_\infty \leq 1$. The adjoint shift operator $S^\dagger$ is a co-isometry because S is an isometry on H^2 i.e. $S^\dagger S = \mathbb{1}_{H^2}$. Furthermore we have that K_θ is the difference of two invariant subspaces for S and M_θ is a compression of S to K_θ i.e.

$$S = \begin{pmatrix} * & * \\ \mathbf{0} & M_\theta \end{pmatrix}.$$

The equivalent statement

$$S^\dagger = \begin{pmatrix} * & \mathbf{0} \\ * & M_\theta^\dagger \end{pmatrix}$$

is nothing but the assertion that K_θ is an invariant subspace for $S^\dagger$. We conclude that $S^\dagger$ is a co-isometric extension for $M_\theta^\dagger$ to H^2. Hence, we can apply the Commutant Lifting Theorem 13 to the commutation relation 3.5. We conclude that there is an operator $Y^\dagger$ that commutes with $S^\dagger$ on H^2 such that $X^\dagger = P_{\mathcal{K}_\theta} Y^\dagger|_{\mathcal{K}_\theta}$ and $\left\|Y^\dagger\right\|_\infty = \left\|X^\dagger\right\|_\infty$. Observe that any bounded operator Z that commutes with the shift on H^2 can be written as

$$Z = Mult_\varphi \tag{3.6}$$

for some $\varphi \in H^\infty$. For suppose $Z(1) = \varphi \in H^\infty$ then $Z(z^n) = z^n\varphi$ and for any polynomial p it follows $Z(p) = \varphi p$. By continuity this extends to any function in H^2, see Proposition 14. We conclude that $Y = Mult_\varphi$ such that $X = \varphi(M_\theta)$. We can also write the norm of X in terms of φ. We note that $Y = Mult_\varphi$ is chosen such that $\|X\|_\infty = \|Y\|_\infty$. Hence,

$$\|X\|_\infty = \|\varphi\|_{H^\infty} .$$

Finally, observe that any operator Y' on H^2 that satisfies $X = P_{\mathcal{K}_\theta} Y'|_{\mathcal{K}_\theta}$ (and commutes with S, see below) has norm larger than or equal to the norm of X. The following proposition summarizes our discussion.

Proposition 16 ((24, 74))**.** *Let θ be an inner function. Then $XM_\theta = M_\theta X$ holds if and only if*

$$X = P_{\mathcal{K}_\theta} Mult_\varphi|_{\mathcal{K}_\theta} = \varphi(M_\theta)$$

for some $\varphi \in H^\infty$. Moreover, we have

$$\|X\|_\infty = \min\{\|\varphi\|_{H^\infty} \mid X = \varphi(M_\theta)\}.$$

3.8 Model spaces with respect to a Blaschke product

In view of the Nevanlinna-Pick interpolation problem a particular interest lies in model spaces where the inner function is a Blaschke product. These spaces also play an important role when one is looking for (optimal) spectral bounds for functions of operators.

We denote by

$$B(z) = \prod_{i=1}^{n} \frac{z - \lambda_i}{1 - \bar{\lambda}_i z}$$

the *Blaschke product* associated to a multi-set $\{\lambda_i\}_{i=1,\dots,n} \subset \mathbb{D}$, where to each λ_i we associate one factor in B *respecting* multiplicities. Observe that $|z| = 1$ implies that $|B(z)| = 1$ therefore B is an inner function. We start our analysis of the model space $\mathcal{K}_B = H^2 \ominus BH^2$ by providing one natural (though not orthogonal) basis. With the Cauchy kernels $k_{\lambda_i} = \frac{1}{1-\bar{\lambda}_i z}$ (see Section 3.5) we can write

$$\langle f | k_{\lambda_i} \rangle = f(\lambda_i).$$

Since B has zeroes at λ_i it follows that any function $f \in BH^2$ satisfies $f(\lambda_i) = 0$ for $1 \leq i \leq n$. With the above relation we conclude that f must be orthogonal to all k_{λ_i}. Hence, if the zeros $\{\lambda_i\}_{i=1,\dots,n}$ of B are distinct $\mathcal{K}_B$ is spanned by the Cauchy kernels k_{λ_i} i.e.

$$\mathcal{K}_B = \operatorname{span}\left\{ \frac{1}{1 - \bar{\lambda}_i z} \right\}_{i=1,\dots,n}.$$

Thus $\mathcal{K}_B$ is a space of rational functions f of the form

$$f(z) = \frac{p(z)}{\prod_i (1 - \bar{\lambda}_i z)},$$

where $p(z)$ is a polynomial of degree at most $n - 1$. Note that if the zeros of B are not distinct the above argument remains valid but the Cauchy kernels have to be replaced by

$$\frac{z^{k_i - 1}}{(1 - \bar{\lambda}_i z)^{k_i}},$$

where k_i denotes the multiplicity of λ_i.

The action of the adjoint model operator M_B^* on the kernels k_{λ_i} is particularly interesting. For any $f \in \mathcal{K}_B$ we have

$$\langle M_B f | k_{\lambda_i} \rangle = \langle P_{\mathcal{K}_B}(zf) | k_{\lambda_i} \rangle = \langle zf | k_{\lambda_i} \rangle = \lambda_i f(\lambda_i) = \lambda_i \langle f | k_{\lambda_i} \rangle.$$

We conclude that for any $f \in \mathcal{K}_B$ it holds that $\langle f | M_B^* k_{\lambda_i} \rangle = \langle f | \bar{\lambda}_i k_{\lambda_i} \rangle$ and hence

$$M_B^* k_{\lambda_i} = \bar{\lambda}_i k_{\lambda_i}. \tag{3.7}$$

The kernel k_{λ_i} is an eigenvector of M_B^* with eigenvalue $\bar{\lambda}_i$. This implies that *the eigenvalues of the model operator* M_B *are* $\{\lambda_i\}_{i=1,\dots,n}$. For any given matrix with distinct eigenvalues we can consider the Blaschke product corresponding to the spectrum and construct M_B, which will have the same spectrum as the original matrix.

3.9 A solution to the Nevanlinna-Pick problem

To demonstrate the power of the framework introduced above we present a solution to the Nevanlinna-Pick interpolation problem proving Proposition 6. Recall that we are given a set of distinct[1] points $\{\lambda_i\}_{i=1,\dots,n} \subset \mathbb{D}$ and a set $\{w_i\}_{i=1,\dots,n} \subset \mathbb{D}$. We are looking for $f \in H^\infty$ with $f(\lambda_i) = w_i$ and $\|f\|_{H^\infty} \leq 1$.

For any $\varphi \in H^\infty$ we have by Definition 7 and Equation 3.7 that

$$\varphi(M_B)^* k_{\lambda_i} = \overline{\varphi(\lambda_i)} k_{\lambda_i} \qquad 1 \leq i \leq n. \tag{3.8}$$

Let now W be an operator on $\mathcal{K}_B$ defined by the relation

$$W^* k_{\lambda_i} = \bar{w}_{\lambda_i} k_{\lambda_i}.$$

Clearly, $M_B W = W M_B$ and we conclude from Proposition 16 that $W = \varphi(M_B)$ for some $\varphi \in H^\infty$ and that

$$\|W\|_\infty = \min\{\|\varphi\|_{H^\infty} \mid W = \varphi(M_B)\} = \min\{\|\varphi\|_{H^\infty} \mid w_i = \varphi(\lambda_i)\}.$$

For the second equality we used Equation 3.8. The requirement $\|\varphi\|_{H^\infty} \leq 1$ is equivalent to $\|W\|_\infty \leq 1$ i.e. for any $f \in \mathcal{K}_B$ we have

$$\langle Wf | Wf \rangle \leq \langle f | f \rangle.$$

Equivalently

$$\begin{aligned} 0 \leq \langle f | \left(\mathbb{1} - WW^*\right) f \rangle &= \sum_{ij} a_i \bar{a}_j \langle k_{\lambda_i} | \left(\mathbb{1} - WW^*\right) k_{\lambda_j} \rangle \\ &= \sum_{ij} a_i \bar{a}_j \left(\frac{1}{1 - \bar{\lambda}_i \lambda_j} - \frac{\bar{w}_i w_j}{1 - \bar{\lambda}_i \lambda_j} \right), \end{aligned}$$

which proves Proposition 6.

[1] It is not crucial that the λ_i are actually distinct. However, the situation where multiplicities occur would corresponds to a mixture of the Schur- and Nevanlinna-Pick interpolation problems (see (24, 75)). This generalized problem can also be studied within the framework of this chapter.

3.10 Bounding the norm of a function of an operator

In this section we return to our original problem of bounding the norm of an operator acting on a (finite-dimensional) Hilbert or Banach space. The main idea is to associate to a given class of linear operators Γ a certain Banach algebra $\mathcal{A}$ of functions, which "mirrors the way those operators are bounded". Instead of working with operators directly we switch to the function algebra and estimate the norm of a representative function in the function algebra. More precisely, our discussion is based on inequalities of the type

$$\|f(X)\| \leq C \|f\|_{\mathcal{A}}\,, \tag{3.9}$$

which relate for a given $X \in \Gamma$ the norm $\|f(X)\|$ to the norm of f in $\mathcal{A}$. Let us make the notions described above more precise.

Definition 8 (Function algebra)**.** A unital Banach algebra $\mathcal{A}$ with elements in $Hol(\mathbb{D})$ will be called a function algebra, if

1. A contains all polynomials and $\lim_{n\to\infty} \|z^n\|_A^{1/n} = 1$.
2. $(a \in \mathcal{A},\ \lambda \in \mathbb{D},\ a(\lambda) = 0) \Rightarrow \frac{a}{z-\lambda} \in \mathcal{A}$.

Definition 9 (Functional calculus)**.** Let $X : \mathcal{B} \to \mathcal{B}$ be an operator on a Banach space $\mathcal{B}$. A bounded algebra homomorphism from a function algebra $\mathcal{A}$ into the set of linear operators on $\mathcal{B}$,

$$\jmath_X :\ \mathcal{A} \to L(\mathcal{B}),$$

will be called a functional calculus for X, if it satisfies $\jmath_X(z) = X$ and $\jmath_X(1) = \mathbb{1}$.

Intuitively $\jmath_X$ captures the notion of "plugging X into f ", that is for $f \in \mathcal{A}$ we set $f(X) = \jmath_X(f)$ and by the boundedness property there is a constant C_X such that $\|f(X)\| \leq C_X \|f\|_{\mathcal{A}}$. Clearly this is only possible if the spectrum of X is contained in the closed unit disk $\sigma(X) \subset \bar{\mathbb{D}}$. Given a family Γ of operators we say that this family obeys a functional calculus with constant C if each $X \in \Gamma$ admits a functional calculus with $C_X \leq C$. Thus, one approach to the problem of bounding the norm $\|f(X)\|$ for $X \in \Gamma$ is by constructing a functional calculus for the family Γ and then bounding the norm of f in the function algebra.

3. SPECTRAL ESTIMATES FROM NEVANLINNA-PICK INTERPOLATION AND MODEL THEORY

At first glance the outlined method might appear to be of little use since the right hand side of Inequality 3.9 no longer depends on X. However, it is possible to exploit spectral properties of X to significantly strengthen the inequality. Let m_X be the minimal polynomial of X. For any $f, g \in \mathcal{A}$ we have then that $\|(f + m_X g)(X)\|_{\mathcal{A}} = \|f(X)\|_{\mathcal{A}}$ and an application of (3.9) reveals that for all $g \in \mathcal{A}$ we have $\|f(X)\| \leq C \|f + m_X g\|_{\mathcal{A}}$. This reduces the problem of bounding $\|f(X)\|$ to an interpolation problem in the associate function algebra, i.e. we are looking for the least norm function $f + m_X g$ in $\mathcal{A}$. The following simple but crucial lemma summarizes this point:

Lemma 17 ((9), Lemma 3.1)**.** *Let $m \neq 0$ be a polynomial and let Γ be a set of matrices that obey an $\mathcal{A}$ functional calculus with constant C and that satisfy $m(X) = 0 \; \forall X \in \Gamma$. Then*

$$\|f(X)\| \leq C \|f\|_{\mathcal{A}/m\mathcal{A}}, \; \forall X \in \Gamma,$$

where $\|f\|_{\mathcal{A}/m\mathcal{A}} = \inf \{\|h\|_{\mathcal{A}} | \; h = f + mg, \; g \in \mathcal{A}\}$.

Proof. For any $g \in \mathcal{A}$ we have that $\|f(X)\| := \|(f + mg)(X)\| \leq C \|f + mg\|_{\mathcal{A}}$. □

For convenience we shall assume in the sequel that $\sigma(X) \subset \mathbb{D}$ and that X can be diagonalized. This does not affect the generality of our discussion since a spectral bound obtained for $\|f(X)\|$ under these assumptions extends by continuity to hold even if X cannot be diagonalized and $\sigma(X) \subset \bar{\mathbb{D}}$. It follows straight from the definition of the function algebra, Definition 8 (see also (9), Section 3.1 (iii) or (75), Section 1.2 P4) that we can rewrite

$$\|f\|_{\mathcal{A}/m\mathcal{A}} = \inf\{\|g\|_{\mathcal{A}} \; | \; g \in \mathcal{A}, \; g(\lambda_i) = f(\lambda_i) \; \forall \lambda_i \in \sigma(A)\}. \tag{3.10}$$

We make a crucial observation. *The right hand side of Equation 3.10 is a Nevanlinna-Pick interpolation problem in the function algebra $\mathcal{A}$.* This is the link to the model theory developed in the previous sections.

Since for $\sigma(A) \subset \mathbb{D}$ the Blaschke product is holomorphic on a set containing $\bar{\mathbb{D}}$ we can define $\|f\|_{\mathcal{A}/B\mathcal{A}}$ as above and note ((75), Lemma 3.1) that as before

$$\|f\|_{\mathcal{A}/B\mathcal{A}} = \inf\{\|g\|_{\mathcal{A}} \; | \; g \in \mathcal{A}, \; g(\lambda_i) = f(\lambda_i) \; \forall \lambda_i \in \sigma(A)\}.$$

In the special case $\mathcal{A} = H^\infty$ we can directly evaluate $\|f\|_{H_\infty/BH^\infty}$ using our solution to the Nevanlinna-Pick problem in Section 3.9 and in particular Proposition 16.

Theorem 18 ((65, 74),Thm. 3.12 (9),Thm. 3.1.11 (24))**.** *For any $\varphi \in H^\infty$ it holds that*

$$\|\varphi\|_{H^\infty/BH^\infty} = \|\varphi(M_B)\|_\infty .$$

We conclude this section with two examples. In the first example we consider power-bounded Banach spaces operators, while the second one treats Hilbert space contractions.

i) Consider a family $\Gamma = \{X \in L(\mathcal{B}) | \, \|X^n\| \leq C \ \forall n \in \mathbb{N}\}$ of Banach space operators that are power bounded by some constant C. This family admits a Wiener algebra functional calculus since for any $f \in W$ and $X \in \Gamma$

$$\|f(X)\| = \left\| \sum_{k\geq 0} \hat{f}(k) X^k \right\| \leq \sum_{k\geq 0} |\hat{f}(k)| \left\| X^k \right\| \leq C \sum_{k\geq 0} |\hat{f}(k)| = C \, \|f\|_W \tag{3.11}$$

holds. We use Lemma 17 to trace back the norm estimate to a Nevanilinna-Pick interpolation problem in W, $\|f(X)\| \leq C \, \|f\|_{W/BW}$.

ii) We consider the semigroup of Hilbert space contractions $\Gamma = \{X \in L(\mathcal{H}) | \, \|X\|_\infty \leq 1\}$. This family allows for an H^∞ functional calculus (with constant $C = 1$), since by von Neumann's inequality (31, 65) we have for any $\varphi \in H^\infty$ and $X \in \Gamma$

$$\|\varphi(X)\|_\infty \leq \|\varphi\|_{H^\infty} .$$

From Lemma 17 we conclude that $\|\varphi(X)\|_\infty \leq \|\varphi\|_{H^\infty/BH^\infty}$ and Lemma 18 provides a solution to the interpolation problem

$$\|\varphi(X)\|_\infty \leq \|\varphi(M_B)\|_\infty . \tag{3.12}$$

Equation 3.12 is a remarkable result. For any contraction X on a finite dimensional Hilbert space we have found a contraction M_B, which has the *same spectrum* as X with equal or larger operator norm. Hence, to obtain a spectral estimate for $\varphi(X)$ it is sufficient to consider $\varphi(M_B)$. This is the reason why we took an operator-theoretic approach to the Nevanlinna-Pick problem in this chapter. This method allowed us to compute the norm of an optimal function for the Nevanlinna-Pick problem (in H^∞) in terms of the norm of an operator, M_B. In this section we have shown how we can bound the norm of an operator with given spectrum provided that we can solve a related Nevanlinna-Pick problem. Both points taken together yield that we can bound the norm of any sufficiently regular function of any contractive operator in terms of the same function of M_B, which has the same spectrum.

3. SPECTRAL ESTIMATES FROM NEVANLINNA-PICK INTERPOLATION AND MODEL THEORY

4

Quantum Spin Chains

The most simple model of a quantum condensed-matter system is provided by a so-called *spin chain*. The latter is a rigid one-dimensional lattice with quantum particles located at its vertices. We demand that the algebra of observables for each quantum particle has finite dimension. However, when the number of sites in the lattice grows to infinity a general C^*-algebraic framework is often inevitable (see for example (29)). The particles exhibit a certain interaction, which should be confined within the system. Under this assumption the spin chain constitutes a closed quantum system, whose evolution is governed by a Hamilton operator (see Section 2.6). Although spin chains are extremely well studied in the literature, important questions remain unanswered. Perhaps most prominently, the Haldane conjecture (76) is open despite considerable effort. One core difficulty lies in the peculiar complexity of the structure of quantum states. The *matrix product state* (MPS) representation provides a framework in which the structure of quantum states on spin chains can be described and analysed efficiently. In fact, the MPS formalism lies at the heart of the ubiquitous "density matrix renormalization group" methods (77, 78) and constitutes the basis for a large number of developments in quantum information and condensed-matter theory. In the literature one also encounters more sophisticated models for condensed-matter systems such as higher dimensional lattices of quantum particles and dissipative evolutions of many-particle systems. For simplicity we will not consider such systems in this exposition.

4.1 Classical vs. quantum spin chains

We model quantum spin chains using subsets $\Lambda \subset \mathbb{Z}$, where each site $x \in \Lambda$ is equipped with a d-dimensional, complex Hilbert space $\mathcal{H}_x \cong \mathbb{C}^d$. The total Hilbert space associated with a finite subset Λ will be $\mathcal{H}_\Lambda = \bigotimes_{x\in\Lambda} \mathcal{H}_x$ [1] For finite Λ any pure state of the spin system is given by a vector $|\psi\rangle \in \mathcal{H}_\Lambda$. When convenient we will impose *periodic boundary conditions* on $|\psi\rangle$ by identifying the last and first sites of Λ.

Let us briefly compare the structures of pure states of a lattice system in the case that the particles at each site constitute a two-level system that is *a)* subject to the laws of classical physics, or *b)* is governed by quantum mechanics. In the classical case *a)* each particle can occupy either one or the other state. We can view the classical lattice as a sequence of coins, which can either show "head" or "tail". If we have full information about the system i.e. the system's state is pure, this means that it can be described completely by a sequence of heads and tails. If our knowledge was incomplete we would have to assign a probability to each coin to show head or tail resulting in a mixed state for the system. The number of pure states of the system is simply 2^N, where N is the number of particles in the lattice. The situation is entirely different for the quantum spin chain *b)*. In this case each local two-level system is described by a normalized vector in $\mathbb{C}^2$ such that the description of the whole system requires a vector in $\mathbb{C}^{2^N} \cong \left(\mathbb{C}^2\right)^{\otimes N}$. Even if we posses full knowledge about the system the number of states is infinite and the dimension of the Hilbert space grows exponentially. The complexity of the state space of quantum many-particle systems makes them hardly tractable with classical methods. A key observation at this point is that physical interactions are often local such that states arising from such interactions are not uniformly distributed in Hilbert space and potentially have a much simpler structure. For this reason it is necessary to have an efficient representation of such states. Despite the fact that it is hard to make this point rigorous the so-called *matrix product state representation* (MPS) intuitively captures the idea of states, whose correlations arise from local interaction. Ultimately, the use of this formulation is justified by its success.

[1] It is a *postulate* of quantum mechanics that the representative Hilbert space of a composite system is given by the tensor product of the Hilbert spaces of constituent systems.

4.2 Matrix product state representation

We consider a finite subset $\Lambda \subset \mathbb{Z}$ consisting of N sites with representative Hilbert space $\mathcal{H}_\Lambda \cong \left(\mathbb{C}^d\right)^{\otimes N}$. Clearly, every pure state of the spin system of Λ can be written as

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} c_{i_1 i_2 \ldots i_N} |i_1 \ldots i_N\rangle,$$

where $\{|i_k\rangle\}_{i_k=1,\ldots,d}$ denotes an orthonormal basis for the Hilbert space at site k and $|i_k i_l\rangle = |i_k\rangle \otimes |i_l\rangle$. The matrix product state representation corresponds to a particular way of writing the coefficients $c_{i_1 i_2 \ldots i_N}$. $|\Psi\rangle$ is referred to as a matrix product state if

$$c_{i_1 i_2 \ldots i_N} = \mathrm{tr}(A^{[1]}_{i_1} \cdot A^{[2]}_{i_2} \cdot \ldots \cdot A^{[N]}_{i_N}),$$

where $A^{[k]}_{i_k}$ denotes a $D_k \times D_{k+1}$ matrix "at site k"[1]. In fact, *every* pure state can be written in this way. Hence, the matrix product state representation is a true representation of quantum states rather than the characterization of a specific class. However, when speaking about MPS one typically has in mind that the dimension of the matrices $A^{[k]}_{i_k}$ (the so-called *bond* dimension) is "small". This property in fact distinguishes a class of states, which has a local structure and bears local correlations.

Theorem 19 ((79))**.** *Every pure state of the spin system of Λ can be written as*

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A^{[1]}_{i_1} \cdot A^{[2]}_{i_2} \cdot \ldots \cdot A^{[N]}_{i_N}) |i_1 \ldots i_N\rangle$$

with site dependent $D_k \times D_{k+1}$ matrices $A^{[k]}_{i_k}$.

For completeness we provide the simple proof.

Proof of Theorem 19. We perform a successive singular value decomposition (80) of the coefficients $c_{i_1 i_2 \ldots i_N}$. In the first step we group the coefficients $i_2 i_3 \ldots i_N$ into one single coefficient $(i_2 i_3 \ldots i_N)$ running from 1 to d^{N-1} such that $c_{i_1 (i_2 \ldots i_N)}$ denote the entries of a $d \times d^{N-1}$ matrix. Singular value decomposition of the latter yields

$$c_{i_1 (i_2 \ldots i_N)} = \sum_{\alpha_1} U_{i_1 \alpha_1} \sigma_{\alpha_1} V_{\alpha_1 (i_2 \ldots i_N)} = \sum_{\alpha_1} \left(A^{[1]}_{i_1}\right)_{\alpha_1} V_{\alpha_1 (i_2 \ldots i_N)}$$

[1] Mostly, subscript indices will enumerate different mathematical objects, while superscripts identify the physical system to which the objects belong.

where U and V are unitary and σ_{α_1} denote singular numbers. In the second step we introduced an appropriate $1 \times d$ matrix $A^{[1]}_{i_1}$. We proceed with a singular value decomposition of $V_{\alpha_1 i_2 i_3 \ldots i_N}$. As above there is unitary W and appropriate $A^{[2]}_{i_2}$ with

$$V_{(\alpha_1 i_2)(i_3 \ldots i_N)} = \sum_{\alpha_2} \left(A^{[2]}_{i_2}\right)_{\alpha_1 \alpha_2} W_{\alpha_2 (i_3 \ldots i_N)}.$$

We conclude by induction that

$$c_{i_1 i_2 \ldots i_N} = \sum_{\alpha_1 \alpha_2 \ldots \alpha_N} \left(A^{[1]}_{i_1}\right)_{\alpha_1} \left(A^{[2]}_{i_2}\right)_{\alpha_1 \alpha_2} \left(A^{[3]}_{i_3}\right)_{\alpha_2 \alpha_3} \cdots \left(A^{[N]}_{i_N}\right)_{\alpha_N}$$

for appropriate $A^{[1]}_{i_1}, \ldots, A^{[N]}_{i_N}$. We observe that the first and last matrix in this representation are in fact vectors and that the bond dimension $\max_k D_k$ is bounded by $\max_k D_k \leq d^{\frac{N}{2}}$. □

A particularly instructive way of looking at MPS is provided by the so-called *valence bond picture*. The idea here is to imagine that each site of the spin chain is subdivided into two "virtual" sites. Virtual sites that belong to different nearby physical sites share a maximally entangled state of the form $|I\rangle = \sum_{\alpha=1}^{D} |\alpha\alpha\rangle$, see Picture 4.1. Suppose now that we identify the last and first site of the chain (i.e. we impose periodic boundary conditions on states of the spin system) and we apply maps of the form

$$\mathcal{A}^{[k]} = \sum_{i_k, \alpha, \beta} A^{[k]}_{i_k, \alpha, \beta} |i_k\rangle\langle\alpha\beta|$$

to the k-th site of the chain. The resulting state is a MPS of the form of Theorem 19.

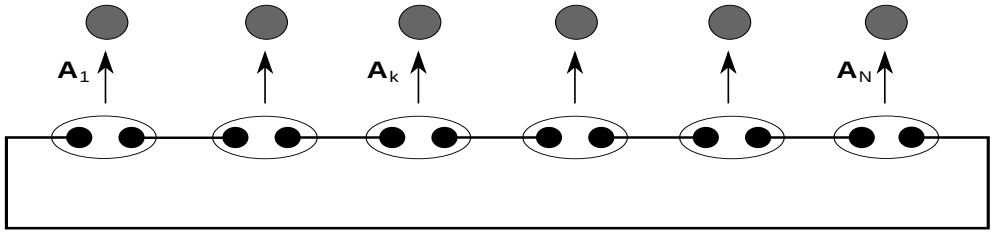

Figure 4.1: Valence bond picture for MPS (with periodic boundary conditions). Ovals depict physical sites. Each physical site is subdivided into two virtual sites (black circles). Solid lines depict entangled states of the form $|I\rangle = \sum_\alpha |\alpha\alpha\rangle$.

4.3 Translational invariance and canonical form

In many cases it is convenient to impose periodic boundary conditions on the states of a spin chain, that is to identify the last with the first site of the spin chain. Often the physical structure[1] of the spin chain motivates the assumption that the state of the chain should be invariant under translations. Any MPS

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_N})|i_1 \ldots i_N\rangle$$

with site-independent $D \times D$ matrices $\{A_i\}_{i=1,\ldots d}$ has both properties. It turns out that the converse statement is also true.

Proposition 20 ((11))**.** *Let $|\Psi\rangle$ be a translationally invariant MPS with periodic boundary conditions on $\Lambda \subset \mathbb{Z}$ then there are $D \times D$ matrices $\{A_i\}_{i=1,\ldots d}$ such that*

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_N})|i_1 \ldots i_N\rangle.$$

Proof. Let

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A^{[1]}_{i_1} \cdot A^{[2]}_{i_2} \cdot \ldots \cdot A^{[N]}_{i_N})|i_1 \ldots i_N\rangle$$

be a MPS representation of $|\Psi\rangle$ with site-dependent matrices. Consider the matrices

$$B_i = \left(\frac{1}{N}\right)^{1/N} \begin{pmatrix} 0 & A_i^{[1]} & 0 & \cdots & 0 \\ 0 & 0 & A_i^{[2]} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & A_i^{[N-1]} \\ A_i^{[N]} & 0 & \cdots & 0 & 0 \end{pmatrix}$$

and compute

$$\begin{aligned} &\sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(B_{i_1} \cdot B_{i_2} \cdot \ldots \cdot B_{i_N})|i_1 \ldots i_N\rangle \\ &= \frac{1}{N} \sum_{j=1}^{N} \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A^{[1]}_{i_{1+j-1}} \cdot A^{[2]}_{i_{2+j-1}} \cdot \ldots \cdot A^{[N]}_{i_{N+j-1}})|i_1 \ldots i_N\rangle \\ &= \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A^{[1]}_{i_1} \cdot A^{[2]}_{i_2} \cdot \ldots \cdot A^{[N]}_{i_N})|i_1 \ldots i_N\rangle. \end{aligned}$$

[1] Think for instance of a large number of atomic spins located in a cavity.

In the last step we used that by assumption $|\psi\rangle$ is translationally invariant and has periodic boundary conditions. Note that if D was the bond dimension in the original MPS representation, then the bond dimension is bounded by ND in the representation with site-independent matrices. □

It follows that a set of matrices $\{A_i\}_{i=1,\dots d}$ provides complete description of translationally invariant MPS with periodic boundary conditions. It is an important conceptual step to associate a CP map

$$\mathcal{E}(X) = \sum_i A_i X A_i^\dagger \tag{4.1}$$

to such a MPS. The study of the structure of $\mathcal{E}$ reveals many important properties of the state $|\Psi\rangle$. We will use $\mathcal{E}$ to study the large scale behaviour of $|\psi\rangle$ in Section 4.5. For now we restrict our discussion to the following observations. The Kraus operators of $\mathcal{E}$ are uniquely determined up to unitary summation, see Theorem 4. It follows that $\mathcal{E}$ determines $|\Psi\rangle$ up to local unitary rotations of each site. The converse question, which CP maps belong to the same MPS is much harder to answer. Obviously the correspondence between sets $\{A_i\}_{i=1,\dots d}$ and MPS is not bijective; for example the set $\{UA_iU^\dagger\}_{i=1,\dots,d}$ with unitary U belongs to the same $|\Psi\rangle$ as above. The following theorem provides a canonical choice of $\{A_i\}_{i=1,\dots d}$ for translationally invariant MPS with periodic boundary conditions.

Theorem 21 ((11))**.** *Let $|\Psi\rangle$ be a translationally invariant MPS with periodic boundary conditions. The matrices A_i can be decomposed as*

$$A_i = \begin{pmatrix} \lambda_1 A_i^{(1)} & 0 & \cdots & 0 \\ 0 & \lambda_2 A_i^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_l A_i^{(l)} \end{pmatrix},$$

where $\lambda_i \in (0,1]$ and the matrices $A_i^{(j)}$ in the j-th block satisfy:

1) *The map $\mathcal{E}_j(X) = \sum_i A_i^{(j)} X \left(A_i^{(j)}\right)^\dagger$ has a unique fixed-point, which is $\mathbb{1}$.*
2) *There are positive and diagonal $\Lambda^{(j)}$ such that $\sum_i \left(A_i^{(j)}\right)^\dagger \Lambda^{(j)} A_i^{(j)} = \Lambda^{(j)}$.*

Proof. Let $|\Psi\rangle = \sum_{i_1,\dots,i_N}^d \mathrm{tr}(A_{i_1} \cdot \dots \cdot A_{i_N}) |i_1 \dots i_N\rangle$ be given and consider the map $\mathcal{E}$ as in 4.1. By rescaling $\mathcal{E}$ with a factor $\lambda > 0$ we can assume that the spectral radius of $\mathcal{E}$ is one. A modification of Brouwer's fixed-point theorem (81, Theorem 2.5) yields that

$\mathcal{E}$ has a positive-semidefinite fixed-point. (Note, that the positive-semidefinite matrices constitute a convex cone, which is preserved under the action of $\mathcal{E}$.) So let $Y \geq 0$ be such that $\mathcal{E}(Y) = Y$ and suppose for now that Y is invertible (that is Y is strictly positive). We can then consider the matrices $B_i = Y^{-1/2} A_i Y^{1/2}$ and observe that

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A_{i_1} \cdot \ldots \cdot A_{i_N}) |i_1 \ldots i_N\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(B_{i_1} \cdot \ldots \cdot B_{i_N}) |i_1 \ldots i_N\rangle$$

and that the map $\tilde{\mathcal{E}}(X) = \sum_i B_i X B_i^\dagger$ satisfies

$$\tilde{\mathcal{E}}(\mathbb{1}) = \mathbb{1},$$

which is the first assertion of *1)*. (We check the other conditions later.) Now suppose that Y has not maximal rank and define P to be the projector onto the support of Y. It follows that

$$A_i P = P A_i P \qquad \forall i. \tag{4.2}$$

To see this suppose for the sake of contradiction that $|\mu\rangle$ is an eigenvector of Y with non-zero eigenvalue λ_μ such that $A_i|\mu\rangle$ is not in the support of Y. Since $\mathcal{E}(Y) = Y$ it follows that the matrix given by

$$\mathcal{E}(Y) - \lambda_\mu A_i |\mu\rangle\langle\mu| A_i^\dagger$$

has a negative eigenvalue. But this is impossible since any map of the form $X \mapsto A_j X A_j^\dagger$ is CP and the matrix is a sum of positive-semidefinite matrices and 4.2 follows. Let $Q = \mathbb{1} - P$. We can decompose

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(P A_{i_1} \cdot \ldots \cdot A_{i_N}) |i_1 \ldots i_N\rangle + \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(Q A_{i_1} \cdot \ldots \cdot A_{i_N}) |i_1 \ldots i_N\rangle$$

and observe that due to 4.2

$$\mathrm{tr}(P A_{i_1} \cdot \ldots \cdot A_{i_N}) = \mathrm{tr}(P A_{i_1} P \cdot \ldots \cdot P A_{i_N} P)$$

and

$$\mathrm{tr}(Q A_{i_1} \cdot \ldots \cdot A_{i_N}) = \mathrm{tr}(Q A_{i_1} Q \cdot \ldots \cdot Q A_{i_N} Q).$$

The above reasoning shows that we can write $|\Psi\rangle$ with matrices

$$\begin{pmatrix} B_i & 0 \\ 0 & C_i \end{pmatrix},$$

where $B_i = PA_iP$ and $C_i = QA_iQ$. Now the CP map with Kraus operators B_i has an invertible fixed-point and we can reason as above to conclude that it can be chosen unital without affecting the MPS. The fixed-point of the channel with Kraus operators C_i might have maximal rank. If this is the case then we argue as before. If not then we proceed iteratively and decompose the matrices C_i as in 4.2. In the end we obtain a block matrix as in Theorem 21, where the λ_i's stem from the scaling of the spectral radii of CP maps. To check the uniqueness of the fixed-point $\mathbb{1}$, let us assume that one of the maps $\mathcal{E}_j$ in Theorem 21 has a further fixed-point, Z. Since $\mathcal{E}_j$ preserves hermiticity we can assume that $Z = Z^\dagger$. Let λ_{max} denote the largest eigenvalue of Z. Clearly, $\mathbb{1} - \frac{1}{\lambda_{max}} Z$ is a positive fixed-point that has not full rank, which allows us to further decompose the matrices in this block until every block satisfies condition *1)*. In an identical way we can argue that any fixed-point of the dual maps $\mathcal{E}_j^\dagger$ has maximal rank. Finally, observe that we can diagonalize this fixed-point using some unitary U such that a final renaming of blocks ($A_i \mapsto UA_iU^\dagger$) achieves that the fixed-point is diagonal.

□

It is natural to ask for conditions under which there is only one block in the above canonical form. In our consecutive discussion two special classes of such MPS will be of particular importance. Both classes are characterized by a condition that is satisfied in generic cases.

4.4 Generic Matrix product states

In this section we consider two generic classes of MPS with translational invariance and periodic boundary conditions. The characterizing conditions (G1) and (G2) turn out to be *essentially* equivalent. For a more detailed discussion we refer to (10, 11).

Condition (G1):
There is a finite number L_0 such that for all $L \geq L_0$ the list of matrices

$$\{A_{i_1} \cdot \ldots \cdot A_{i_L}\}_{i_j \in \{1 \ldots d\}}$$

spans the entire algebra of $D \times D$ matrices.

Condition (G1) is generic in the sense that d matrices chosen randomly according

to some uniform measure will comply with this condition with probability one. It is not hard to see that (G1) holds iff the map

$$\Gamma_L : \; X \mapsto \sum_{i_1,\ldots,i_L}^{d} \mathrm{tr}(X A_{i_1} A_{i_2} \ldots A_{i_L}) |i_1 \ldots i_L\rangle$$

is injective for $L \geq L_0$.

Our second condition for generic MPS is related to the spectral properties of the map $\mathcal{E}$ associated to $|\Psi\rangle$. Without loss of generality we can choose the spectral radius of $\mathcal{E}$ to be one. Hence, all eigenvalues of $\mathcal{E}$ are confined within the closed unit disk in the complex plane.

Condition (G2):
The map $\mathcal{E}$ associated with $|\Psi\rangle$ via Equation 4.1 has a unique eigenvalue of magnitude one.

The (two-dimensional) Lebesgue measure of the boundary of the unit disk is zero. Hence, a randomly chosen spectrum satisfies condition (G2) with probability one.

Proposition 22 ((10, 11))**.** *Let $|\Psi\rangle$ be a translationally invariant MPS with periodic boundary conditions.*

i) If condition (G1) holds for some $L_0 < N$ then there is only one block in the canonical form of Theorem 21.

ii) If condition (G2) holds then there is only one block in the canonical form of Theorem 21.

It follows from Proposition 22 and Theorem 21 that without loss of generality $\mathcal{E}$ can be chosen to be CPU. Concerning the relation between (G1) and (G2) we have the following.

Proposition 23 ((10),(82))**.** *Let $\mathcal{E}$ be CPU. If (G1) holds for the Kraus operators of $\mathcal{E}$, then $\mathcal{E}$ has a unique eigenvalue of magnitude one (i.e. (G2) holds). If $\mathcal{E}$ has a unique eigenvalue of magnitude one (i.e. (G2) holds) and the corresponding eigenvector of $\mathcal{E}^\dagger$ is positive, then (G1) holds.*

We do not provide a proof of this statement in this exposition but refer to (82, Proposition 3), which contains a complete discussion. Proposition 23 is of conceptual

importance as it provides a link between the (a priori unrelated) *algebraic* and *spectral* structures of CP maps. Their interplay might be useful in the construction of an effective functional calculus for CP maps.

4.5 Generic large scale behaviour of MPS

If a MPS complies with the conditions (G1) and (G2) this has important consequences for the large scale behaviour of the state. From physical observation one expects that on sufficiently large scale i.e. under the assumption that a sufficient number of quantum particles is involved, the system behaves as if it was classical. To make this more rigorous let us analyse the asymptotic behaviour of the map $\mathcal{E}$ in Theorem 21 under the assumption (G2). Since the only eigenvalue of $\mathcal{E}$ of magnitude one is 1 we can compute

$$\lim_{n\to\infty} \mathcal{E}^n = \mathcal{E}^\infty,$$

where $\mathcal{E}^\infty$ denotes the part in the Jordan decomposition (83) of $\mathcal{E}$ corresponding to the eigenvalue 1. If we call Λ to the fixed-point of $\mathcal{E}^\dagger$ corresponding to the fixed-point $\mathbb{1}$ of $\mathcal{E}$, it follows that $\mathcal{E}^\infty(X) = \mathrm{tr}(\Lambda X)\mathbb{1}$. The following lemma summarizes the mentioned points.

Lemma 24 ((11, 84))**.** *Let $\mathcal{E}(X) = \sum_i A_i X A_i^\dagger$ be a CPU map such that 1 is the unique eigenvalue of magnitude one and suppose that $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_D)$ with $\lambda_i > 0$ is the corresponding fixed point of $\mathcal{E}^\dagger$. Then the limit $\lim_{n\to\infty} \mathcal{E}^n = \mathcal{E}^\infty$ exists and we can write $\mathcal{E}^\infty(X) = \sum_{i=1}^{D^2} A_i^{(\infty)} X (A_i^{(\infty)})^\dagger$ with matrices $A_{(pq)}^{(\infty)} = \sqrt{\lambda_q}|p\rangle\langle q|$ and $p, q \in \{1, ..., D\}$.*

The proof is clear. To analyse the large-scale behaviour of a MPS $|\Psi\rangle$ we perform a coarse-graining procedure. The latter consists of an iteration of the following two steps. First, we merge a number of neighbouring cites in the spin chain into a new one. This is reflected by a simple renaming of the corresponding matrices in the MPS:

$$\begin{aligned} |\Psi\rangle &= \sum_{i_1,...,i_N}^{d} \mathrm{tr}(A_{i_1} \cdot ... \cdot \underbrace{A_{i_k} \cdot ... \cdot A_{i_{k+L}}}_{\hat{A}_{j_k}} \cdot ... \cdot A_{i_N}) |i_1 ... \underbrace{i_k ....i_{k+L}}_{j_k} ...i_N\rangle \\ &= \sum_{j_1,...,j_M}^{d^{L+1}} \mathrm{tr}(\hat{A}_{j_1} \cdot ... \cdot \hat{A}_{j_M}) |j_1 \cdot ... \cdot j_M\rangle \end{aligned}$$

Second, we perform a suitable unitary rotation of the new sites. The aim of this rotation is to reduce the number of matrices at each site, see Lemma 25.

$$|\Psi'\rangle = U^{\otimes M}|\Psi\rangle = \sum_{j_1,\ldots,j_M}^{d^{L+1}} \mathrm{tr}(\hat{A}_{j_1}\cdot\ldots\cdot\hat{A}_{j_M})U^{\otimes M}|j_1\cdot\ldots\cdot j_M\rangle$$
$$= \sum_{m_1,\ldots,m_M} \mathrm{tr}(A^{(L)}_{m_1}\cdot\ldots\cdot A^{(L)}_{m_M})|m_1\cdot\ldots\cdot m_M\rangle$$

Picture 4.2 contains a depiction of the evolution of MPS under the coarse-graining procedure.

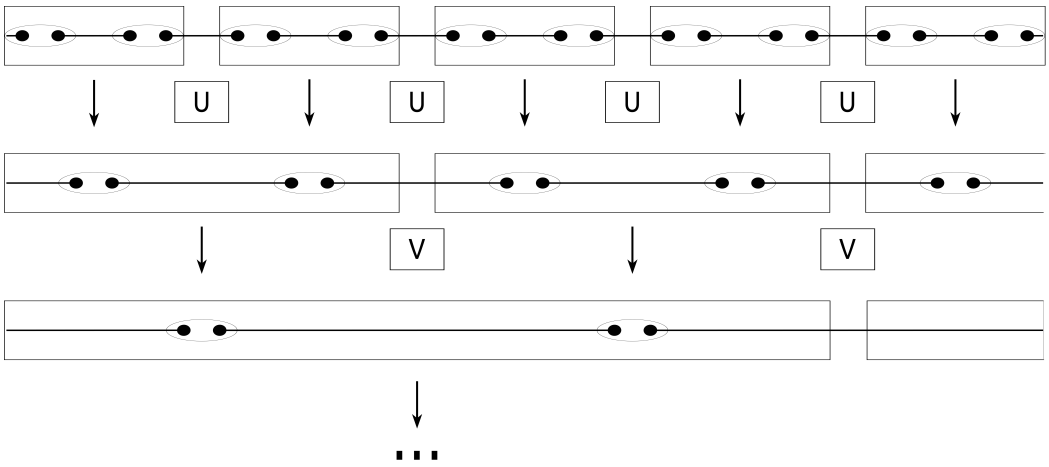

Figure 4.2: Depiction of course-graining procedure. First line shows MPS, where the rectangles group spins into blocks. A suitable unitary is applied in the second line. It follows successive blocking and unitary transformation.

The following Lemma 25 is the result of (84). It summarizes the mathematical observations that constitute the fundamentals of the described coarse-graining procedure. The formulation is taken from (85).

Lemma 25 ((84, 85))**.** *Let $\{A_i\}_{i=1,\ldots,d}$ be a set of $D\times D$ matrices and consider the set $\{A_{i_1}\cdot\ldots\cdot A_{i_L}\}_{i_j=1,\ldots,d}$ of all matrix products formed by matrices from $\{A_i\}_{i=1,\ldots,d}$. There is a $d^L\times d^L$ unitary matrix U and matrices $A^{(L)}_m$ with*

$$\sum_{i_1,\ldots,i_L}^{d} U_{m(i_1\ldots i_L)}A_{i_1}\cdot\ldots\cdot A_{i_L} = A^{(L)}_m \tag{4.3}$$

such that $A^{(L)}_m = 0$ for all $m > \min\{D^2, d^L\}$. Moreover, it holds that $\mathcal{E}^L = \mathcal{E}^{(L)}$, where $\mathcal{E}^{(L)}$ denotes the CP map with Kraus operators $A^{(L)}_m$.

Proof of Lemma 25. We write $(A_{i_1} \cdot ... \cdot A_{i_L})_{\alpha,\beta}$ with $\alpha, \beta \in \{1, ..., D\}$ for the entry of the matrix $A_{i_1} \cdot ... \cdot A_{i_L}$ in row α and column β. Let $\tilde{A}$ be the $d^L \times D^2$ matrix which has the entry $(A_{i_1} \cdot ... \cdot A_{i_L})_{\alpha,\beta}$ in its $(i_1...i_L)$-th row and (α, β)-th column. We perform a singular value decomposition of $\tilde{A}$ writing

$$\tilde{A}_{(i_1...i_L),(\alpha\beta)} = \sum_{l=1}^{\min(D^2,d^L)} (U^\dagger)_{(i_1...i_L),l}\, \rho_l\, V_{l,(\alpha\beta)}.$$

For the m-th row of $U\tilde{A}$, $(U\tilde{A})^{(m)}$, then

$$(U\tilde{A})^{(m)} = \begin{cases} \rho_m V^{(m)} & ;\ m \leq \min\{d^L, D^2\} \\ 0 & ;\ m > \min\{d^L, D^2\} \end{cases}$$

holds. The rows of the matrix $U\tilde{A}$ now correspond to the matrices $A_i^{(L)}$ and thus the first assertion of the lemma follows.
For the second assertion simply observe that for any X the quantity

$$\mathcal{E}^L(X) = \sum_{i_i,...,i_L}^{d} A_{i_1} \cdot ... \cdot A_{i_L} X A_{i_L}^\dagger \cdot ... \cdot A_{i_1}^\dagger$$

is invariant under unitary summations i.e.

$$\mathcal{E}^L(X) = \sum_m A_m^{(L)} X (A_m^{(L)})^\dagger = \mathcal{E}^{(L)}(X).$$

□

There are two core observations in this lemma. First, there is the assertion that within the coarse-graining procedure the number of matrices at each site *remains bounded by* D^2. That is, once the dimension of the local sites has reached D^2 it does not have to be increased. In this sense a MPS carries only finite-range correlations. Second, we observe that the description of coarse-graining via $\mathcal{E}$ is particularly simple. Each coarse-graining step is reflected by taking an appropriate power of the map $\mathcal{E}$. Hence, the asymptotic behaviour of the described procedure is intrinsically connected to the structure of $\mathcal{E}^\infty$. The speed of convergence of $\mathcal{E}^N$ towards $\mathcal{E}^\infty$ is obviously exponential. It follows that the same is true for the coarse-graining procedure. See (85) for a rigorous discussion of the latter point. The matrix representation of Kraus operators of $\mathcal{E}^\infty$ given in Lemma 24 allows one to compute the state arising from coarse-graining in the asymptotic regime. For this we assume that the number of sites in the chain is

much larger than the number of iterations that are required to bring $\mathcal{E}^N$ "close" to $\mathcal{E}^\infty$. The asymptotic state corresponding to K sites is given by

$$\begin{aligned}|\Psi^{(\infty)}\rangle &= \sum_{i_1,\ldots,i_K}^{D^2} \mathrm{tr}(A_{i_1}^{(\infty)} \cdot \ldots \cdot A_{i_K}^{(\infty)})|i_1 \cdot \ldots \cdot i_K\rangle \\ &= (|\varphi\rangle)^{\otimes K}\,.\end{aligned}$$

Here, the state $|\varphi\rangle = \sum_{i=1}^{D} \sqrt{\lambda_i}|ii\rangle$ is shared by neighbouring virtual sites corresponding to different physical sites, see Picture 4.3. Thus, at the fixed point of the aforementioned iteration, the "scale-invariant" state can be described in terms of two virtual spins at each site. Note also that the state $(|\varphi\rangle)^{\otimes K}$ is *classical* with respect to a basis of half-shifted spins containing $|\varphi\rangle$ (see (12, 86)).

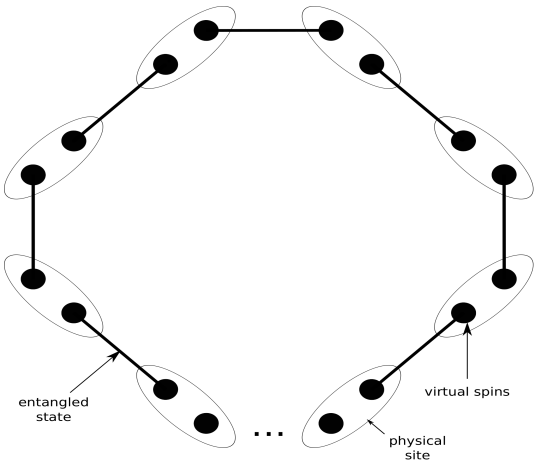

Figure 4.3: Depiction of asymptotic state after course-graining. Ovals depict physical sites. Black circles depict virtual sites. Solid lines between virtual sites correspond to entangled state $|\varphi\rangle = \sum_{i=1}^{D} \sqrt{\lambda_i}|ii\rangle$.

4.6 Frustration free Hamiltonians

A "frustration free" Hamiltonian describes a certain type of evolution of a quantum spin chain due to local interaction. We fix a spin system $\Lambda \subset \mathbb{Z}$ and an interaction range $\Lambda_0 \subset \Lambda$. As before, it is convenient to identify the first and last sites of Λ. Consider an interaction Hamiltonian h_x acting (non-trivially) on $\mathcal{H}_{\Lambda_0+x}$, where $\Lambda_0 + x$ denotes

a translate of Λ_0 by $x \in \Lambda$. The total Hamiltonian can be written as a formal sum

$$H_\Lambda = \sum_{x \in \Lambda} h_x$$

where, h_x is extended to $\mathcal{H}_\Lambda$ by tensoring with an implicit identity on $\mathcal{H}_{\Lambda - \Lambda_0}$. This is the typical set-up for the description of the evolution of a closed quantum spin system with local interaction. What distinguishes a *frustration free* quantum spin system are the following assumptions.

1. The operator h_x is positive-semidefinite for any x.
2. We have $\dim(\mathrm{Kern}(H_\Lambda)) > 0$ (if Λ is not empty).

If these assumptions are satisfied $\mathrm{Kern}(H_\Lambda)$ is referred to as the ground-state subspace of H_Λ and the ground-states are called frustration-free. We observe that since all h_x are positive-semidefinite,

$$\mathrm{Kern}(H_\Lambda) = \bigcap_{x \in \Lambda} \mathrm{Kern}(h_x).$$

This means that if $|\Omega\rangle_\Lambda$ is a ground state of H_Λ then it is also a ground-state of each local interaction term, $h_x|\Omega\rangle_\Lambda = 0$ for any x. We say that H_Λ is gapped if it has a gap $\gamma > 0$ above the ground state energy

$$H_\Lambda|_{\mathcal{H}_\Lambda \ominus |\Omega\rangle_\Lambda} \geq \gamma \mathbb{1}.$$

Any gapped Hamiltonian can be approximated by frustration-free Hamiltonians if one allows to increase the interaction range of the local terms to $\mathcal{O}(\log(N))$ (87). To mention a few examples of frustration free systems, there are ferromagnetic systems: the isotropic Heisenberg ferromagnet has a frustration free ground state as well as the anisotropic XXZ ferromagnetic models. A frustration free antiferromagnetic model is the AKLT model (88). It is of particular importance for the understanding of the role of spin-dimension in antiferromagnetic chains. Haldane (76) conjectured that the isotropic Heisenberg antiferromagnet has continuous excitations above the ground state energy if the local spins are half-integer and an energy gap if the spins take integer values.

4.7 Parent Hamiltonians of MPS

We consider a translationally invariant state with periodic boundary conditions $|\Psi\rangle = \sum_{i_1 \ldots i_N} \mathrm{tr}(A_{i_1} \cdot \ldots \cdot A_{i_N}) |i_1 \ldots i_N\rangle$ on a spin system Λ. For fixed $L \in \mathbb{N}$ we define $\mathcal{G}_L \subset$

$(\mathbb{C}^d)^{\otimes L}$ to be the subspace spanned by the vectors

$$|\Psi(X)\rangle = \sum_{i_1 \ldots i_L} \mathrm{tr}(X A_{i_1} \cdot \ldots \cdot A_{i_L}) |i_1 \ldots i_L\rangle,$$

where X are complex $D \times D$ matrices. Note that if condition (G1) holds A_i then for $L \geq L_0$ the space spanned by $|\Psi(X)\rangle$ has dimension D^2. We write $h_{\mathcal{G}_L}$ for the projector onto the orthogonal complement of $\mathcal{G}_L$ in $(\mathbb{C}^d)^{\otimes L}$. The *canonical parent Hamiltonian* for $|\Psi\rangle$ (and fixed L) is defined as the formal expression

$$H_\Lambda = \sum_i^N \tau^i(h_{\mathcal{G}_L}), \tag{4.4}$$

where τ denotes the translation operation by one site (10, 11). It is clear from the definition that $H_\Lambda |\Psi\rangle = 0$ and that H_Λ is frustration free.

Proposition 26 ((10, 11))**.** *Let $|\Psi\rangle$ be a translationally invariant MPS with periodic boundary conditions such that condition (G1) is satisfied. Suppose that $N \geq 2L_0$ and $L > L_0$ and let H_Λ be the canonical parent Hamiltonian for $|\psi\rangle$, see 4.4. Then $|\Psi\rangle$ is the unique ground state of H_Λ.*

Proof. Suppose $|\Phi\rangle$ is such that $H_\Lambda |\Phi\rangle = 0$, we show that then $|\Psi\rangle = |\Phi\rangle$. The proof is carried out in two steps. First, we show that $|\Phi\rangle \in \mathcal{G}_N$. Second, we use this fact together with the cyclicity of H_Λ to conclude our proof.

Clearly, $H_\Lambda |\Phi\rangle = 0$ implies $\tau^i(h_{\mathcal{G}_L})|\Phi\rangle = 0$ for any i. We can compute for example for $i = 0$

$$\begin{aligned} 0 = (h_{\mathcal{G}_L} \otimes \mathbb{1})|\Phi\rangle &= (h_{\mathcal{G}_L} \otimes \mathbb{1}) \sum_{i_1,\ldots,i_N} c_{i_1 \ldots i_N} |i_1 \ldots i_N\rangle \\ &= \sum_{i_{L+1},\ldots,i_N} h_{\mathcal{G}_L} \left(\sum_{i_1,\ldots,i_L} c_{i_1 \ldots i_N} |i_1 \ldots i_L\rangle \right) \otimes |i_{L+1} \ldots i_N\rangle. \end{aligned}$$

Therefore $c_{i_1 \ldots i_N} = \mathrm{tr}(X_{i_{L+1} \ldots i_N} A_{i_1} \cdot \ldots \cdot A_{i_L})$. But we also have that $(\tau(h_{\mathcal{G}_L}) \otimes \mathbb{1})|\Phi\rangle = 0$ and the same computation reveals that $c_{i_1 \ldots i_N} = \mathrm{tr}(X_{i_1 i_{L+2} \ldots i_N} A_{i_2} \cdot \ldots \cdot A_{i_{L+1}})$. By assumption $L > L_0$ such that condition (G1) entails

$$X_{i_{L+1} \ldots i_N} A_{i_1} = A_{i_{L+1}} X_{i_1 i_{L+2} \ldots i_N}.$$

We have without loss of generality (see Theorem 21) $\sum_i A_i A_i^\dagger = \mathbb{1}$, such that $X_{i_{L+1} \ldots i_N} = A_{i_{L+1}} Y_{i_{L+2} \ldots i_N}$ with $Y_{i_{L+2} \ldots i_N} = \sum_{i_1} X_{i_1 i_{L+1} \ldots i_N} A_{i_1}^\dagger$. We conclude that

$$c_{i_1 \ldots i_N} = \mathrm{tr}(Y_{i_{L+2} \ldots i_N} A_{i_1} \cdot \ldots \cdot A_{i_L} A_{i_{L+1}}).$$

One can now use that for all i, $\tau^i(h_{\mathcal{G}_L})|\Phi\rangle = 0$ and repeat the depicted steps to find that

$$|\Phi\rangle = \sum_{i_1,\ldots,i_N} \mathrm{tr}(XA_{i_1} \cdot \ldots \cdot A_{i_N})|i_1 \ldots i_N\rangle.$$

This proves that $|\Phi\rangle \in \mathcal{G}_N$ and concludes the first step. But we can also write

$$|\Phi\rangle = \sum_{i_1,\ldots,i_N} \mathrm{tr}(A_{i_1} \cdot \ldots \cdot A_{i_{L_0}} Y A_{i_{L_0+1}} \cdot \ldots \cdot A_{i_N})|i_1 \ldots i_N\rangle$$

since H_Λ is invariant under translation. Using the assumption $N \geq 2L_0$ it follows that for every $i_1, \ldots, i_{L_0}$ we have

$$XA_{i_1} \cdot \ldots \cdot A_{i_{L_0}} = A_{i_1} \cdot \ldots \cdot A_{i_{L_0}} Y.$$

Since the products $A_{i_1} \cdot \ldots \cdot A_{i_{L_0}}$ span the entire algebra, which in particular contains $\mathbb{1}$, it follows $X = Y$. The latter is a matrix, which commutes with any matrix in the algebra, which by Schur's Lemma yields $X = c\mathbb{1}$ for some $c \in \mathbb{C}$, which by normalization must be 1. □

More generally, under condition (G1) H_Λ can be shown to have a spectral gap $\gamma > 0$ above the ground state energy (10) even in the limit of an *infinite* chain. For us it will also be important to study the spectral gap of a restriction of H_Λ to a certain subset $\Lambda_1 \subset \Lambda$. Let $\Lambda_1 \subset \Lambda$ and let G_{Λ_1} denote the projector onto the kernel of $H_{\Lambda_1} = \sum_{i:\{i+1,\ldots,i+L\}\subset\Lambda_1} \tau^i(h_{\mathcal{G}_L})$. Observe, that H_{Λ_1} does not have periodic boundary conditions. The *local gap* is defined to be the largest number γ_{Λ_1} such that

$$H_{\Lambda_1} \geq \gamma_{\Lambda_1} (\mathbb{1} - G_{\Lambda_1}).$$

The local gap does not depend on Λ but only on the number of sites in Λ_1. The "Local-Gap condition" of (89) refers to the property that the spectral gap of a frustration-free Hamiltonian decays at most polynomially in the number of lattice sites. In (90, 91) a constant lower bound on the local gap of a (one-dimensional) frustration-free Hamiltonians is derived. In particular, parent Hamiltonians satisfy the Local-Gap condition.

Bibliography

[1] P.J. Schweitzer. **Perturbation Theory and finite Markov chains**. *J. Appl. Prob.*, **5**:401–413, 1968. 3, 4

[2] E.G Cho and C.D Meyer. **Comparison of perturbations bounds for the stationary distributions of a Markov chain**. *Lin. Alg. and Appl.*, **335:1**:137–150, 2001. 3

[3] A.Yu Mitrophanov. **Sensitivity and convergence of uniformly ergodic Markov chains**. *J. Appl. Prob.*, **42**:1003–1014, 2005. 3, 4

[4] C.D Meyer. **Sensitivity of the stationary distribution of a Markov chain**. *SIAM J. Matrix Anal. Appl.*, **15:3**:715–728, 1994. 3

[5] E Seneta. **Perturbation of the Stationary Distribution measured by Ergodicity Coefficients**. *Adv. Appl. Prob.*, **20**:228–230, 1988. 3

[6] F. Versraete, M. Wolf, and I. Cirac. **Quantum computation and quantum-state engineering driven by dissipation**. *Nat. Phys.*, **5 No.9**:633–636, 2009. 3, 19, 20

[7] A.Yu Mitrophanov. **Stability and exponential convergence of continuous-time Markov chains**. *J. Appl. Prob.*, **40**:970–979, 2003. 4

[8] B. Davies and B. Simon. **Eigenvalue estimates for non-normal matrices and the zeros of random orthogonal polynomials on the unit circle**. *Jour. Approx. Theo.*, **141:2**:189–213, 2006. ArXiv:0603098v1. 4

[9] N.K. Nikolski. **Condition Numbers Of Large Matrices and Analytic Capacities**. *St. Petersburg Math. J.*, **17**:641–682, 2006. 4, 38, 39

[10] M. Fannes, B. Nachtergaele, and R.F. Werner. **Finitely correlated states on quantum spin chains**. *Comm. in Math. Phys.*, **144(3)**:443–490, 1992. 4, 20, 48, 49, 55, 56

[11] D. Perez-Garcia, F. Verstaete, M.M. Wolf, and J.I. Cirac. **Matrix product state representations**. *Quantum Inf. Comput.*, **7**:401–430, 2007. 4, 5, 45, 46, 48, 49, 50, 55

[12] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. 5, 13, 14, 53

[13] Frédéric Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, 2009. 6

[14] Alexander S. Holevo. **The capacity of the quantum channel with general signal states**. *IEEE Transactions on Information Theory*, **44**:269–273, 1998. 6

[15] Benjamin Schumacher and M. D. Westmoreland. **Sending classical information via noisy quantum channels**. *Physical Review A*, **56**:131–138, 1997. 6

[16] A. Rényi. **On Measures of Information and Entropy**. In *Proc. Symp. on Math., Stat. and Probability*, pages 547–561, Berkeley, 1961. University of California Press. 6

[17] Renato Renner. *Security of quantum key distribution*. PhD thesis, ETH Zürich, 2005. arXiv: 0403133 [quant-ph]. 6

[18] Marco Tomamichel. *A framework for non-asymptotic quantum information theory*. PhD thesis, ETH Zürich, 2012. 6

[19] Hans Maassen and J. Uffink. **Generalized Entropic Uncertainty Relations**. *Phys. Rev. Lett.*, **60**(12):1103–1106, March 1988. 6

[20] Rupert L. Frank and Elliott H. Lieb. **Monotonicity of a relative Rényi entropy**. June 2013. 6

[21] Salman Beigi. **Sandwiched Rényi Divergence Satisfies Data Processing Inequality**. 2013. ArXiv:1306.5920(1). 7

[22] Mark M. Wilde, Andreas Winter, and Dong Yang. **Strong converse for the classical capacity of entanglement-breaking channels**. June 2013. 7

[23] M. Mosonyi and T. Ogawa. **Quantum hypothesis testing and the operational interpretation of the quantum Renyi relative entropies**. 2013. ArXiv:1309.3228. 7

[24] N.K. Nikolski. *Operators, Functions and Systems: An Easy Reading*. AMS: Mathematical Surveys and Monographs: 93, 2009. 7, 23, 25, 34, 36, 39

[25] E.P. Wigner. **The Unreasonable Effectiveness of Mathematics in the Natural Sciences**. *Comm. on pure and appl. Math.*, **XIII**:001–14, 1960. 10

[26] N. Straumann. *Quantenmechanik: Nichtrelativistische Quantentheorie*. Springer-Lehrbuch, 2002. 10, 19

[27] J. Froehlich. **Quantum Mechanics: Lecture notes**. CTAN: `http://www.itp.phys.ethz.ch/research/mathphys/froehlich/lecture_notes`, 2009. 10

[28] Michael Wolf. **Quantum Channels and Operations: Lecture notes**. CTAN: `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf`, 2011. 10, 13, 14, 16, 17, 19

[29] O. Bratteli and B. Robinson. *Operator Algebras and Quantum Statistical Mechanics 1*. Springer, 2010. 10, 11, 12, 41

[30] W.B. Arveson. *An invitation to C^*-algebras*. Springer, New York, 1976. 10, 11

BIBLIOGRAPHY

[31] V. I. Paulsen. *Completely bounded maps and operator algebras*. Cambridge University Press, 2002. 10, 13, 15, 16, 39

[32] I.M. Gelfand and M.A. Naimark. **On the imbedding of normed rings into the ring of operators on a Hilbert space**. *Math. Sbornik*, **12(2)**:197–217, 1943. 11

[33] M. Hazewinkel. *Riesz representation theorem*. Springer, 2001. 12, 31

[34] C. Simon, V. Bžek, and N. Gisin. **The no-signaling condition and quantum dynamics**. 2001. ArXiv:quant-ph/0102125v1. 13

[35] M. Ferrero, D. Salgado, and Sánchez-Gómez. **Nonlinear quantum evolution does not imply supraluminal communication**. *Physical Review A*, **70**:014101, 2004. 13

[36] J.F. Jordan. **Assumptions that imply that quantum dynamics is linear**. 2006. ArXiv:quant-ph/0508092v4. 13

[37] W. Arveson. **Subalgebras of C^*-algebras**. *Acta. Math.*, **123**:141–224, 1969. 13

[38] J. Watrous. **Notes on super-operator norms induced by Schatten norms**. arXiv: 0411077v1, November 2004. 14

[39] A. Kitaev. **Quantum computations: algorithms and error correction**. *Russian Mathematical Surveys*, **52**(6):1191–1249, 1997. 14

[40] N. Johnston, D.W. Kribs, and V.I. Paulsen. **Computing stabilized norms for quantum operations via the theory of completely bounded maps**, 2007. arXiv:0711.3636. 15

[41] D.E. Evans. **Spectral properties of positive maps on C*-algebras**. *J. of London Math. Soc.*, **17(2)**:345–355, 1978. 15

[42] W.F. Stinespring. **Positive functions on C^*-algebras**. *Proc. Amer. Math. Soc.*, **6**:211–216, 1955. 16

[43] G. Lindblad. **Completely Positive Maps and Entropy Inequalities**. *Comm. Math. Phys*, **40**:147–151, 1975. 16

[44] D. Kretschmann, D. Schlingemann, and R.F. Werner. **A continuity theorem for Stinespring's dilation**. *J. of Funct. Analysis*, **255:8**:1889–1904, 2007. 16

[45] D. Kretschmann, D. Schlingemann, M.M. Wolf, and R.F. Werner. **The Information-Disturbance Tradeoff and the Continuity of Stinespring's Theorem**. *IEEE Transactions on Information Theory*, **54(4)**:1708–1717, 2006. 16

[46] K. Kraus. **General state changes in quantum theory**. *Ann. Phys.*, **64**:311–335, 1970. 17

[47] E. Seneta. *Non-negative matrices and Markov chains*. Springer Ser. in Stat., 2006. 17

[48] A.D. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times: With a Chapter on Coupling from the Past*. AMS Publishing, 2000. 17

[49] D. Revuz. *Markov Chains*. Elsevier Sc. Pub., 1991. 17

[50] D. Freedman. *Markov Chains*. Springer, 1983. 17

[51] P. Diaconis. *Group representations in probability and statistics*. IMS Lecture Notes, 1988. 17

[52] G. Lindblad. **On the generators of quantum dynamical semigroups**. *Comm. Math. Phys.*, **48**:119, 1976. 18

[53] F. Kuypers. *Klassische Mechanik*. Wiley-VCH, 2005. 19

[54] D. DiVincenzo. **The Physical Implementation of Quantum Computation**. 2000. ArXiv:quant-ph/0002077. 20

[55] F. Verstaete and I.J. Cirac. **Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions**. 2004. ArXiv:cond-mat/0407066v1. 20

[56] A. Yu. Kitaev. **Fault-tolerant quantum computation by anyons**. *Ann. Phys.*, **303:2**, 2003. 20

[57] N.K. Nikolski. *Treatise on the Shift Operator*. Springer: Grundlehren der mathematischen Wissenschaft, 1986. 23

[58] J. Agler and J.E McCarthy. *Pick Interpolation in Hilbert function spaces*. AMS: Graduate studies in mathematics: 44, 2002. 23, 30, 32

[59] G. Pick. **Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden können.** *Math. Ann.*, **77**:7–23, 1916. 24

[60] F. Riesz. **Über die Randwerte einer analytischen Funktion**. *Math. Z.*, **18**:87–95, 1923. 25

[61] Y. Katznelson. *An introduction to Harmonic Analysis*. Dover, 1976. 25

[62] A. Beurling. **On two problems concerning linear transformations in Hilbert space**. *Acta Math.*, **81**:239–255, 1949. 27

[63] H. Helson. *Lectures on invariant subspaces*. Academic press, New York, 1964. 27

[64] D. Sarason. **On spectral sets having connected complement**. *Acta Sci. Math.*, **26**:289–299, 1965. 29

[65] Sz. B. Nagy, C. Foias, H. Bercovici, and L. Kerchy. *Harmonic analysis of operators on Hilbert spaces*. Springer, 2010. 29, 33, 39

[66] T. Andô. **On a pair of commutative contractions**. *Acta Sci. Math. (Szeged)*, **24**:88–90, 1963. 29

[67] S. Parrott. **Unitary dilations for commuting contractions**. *Pacific Math.*, **34**:481–490, 1978. 29

[68] M.J. Crabb and A.M. Davie. **Von Neumann's inequality for Hilbert space operators**. *Bull. London Math. Soc.*, **7**:49–50, 1975. 30

[69] D. Opelă. **A generalization of Andô's theorem and Parrott's example**. 1978. arXiv:0505154v1. 30

[70] Th.N. Varopoulos. **On an inequality of von Neumann and an application of the metric theory of tensor products to operators theory**. *J. Func. Anal.*, **16**:83–100, 1974. 30

[71] B.A. Lotto and T. Steger. **Von Neumann's inequality for commuting diagonalizable contractions**. *Proc. Amer. Math. Soc.*, **120:3**:897–901, 1994. 30

[72] Sz. Nagy and C. Foiaş. **Commutants de certains opérateurs**. *Act. Sci. Math.*, **29**:1–17, 1968. 30

[73] F. Foiaş and A. Frazho. **The commutant lifting approach to interpolation problems**. *Birkhäuser, Basel*, 1990. 30

[74] D. Sarason. **Generalized interpolation in H^∞**. *Trans. Amer. Math. Soc.*, **127**:289–299, 1967. 34, 39

[75] R. Zarouf. **Effective H^∞-interpolation**. *Houston J. Math.*, **39:2**:487–514, 2013. 36, 38

[76] F.D.M. Haldane. **Continuum dynamics of the 1-d Heisenberg antiferromagnet: Identification with the O(3) nonlinear sigma model**. *Phys. Lett. A*, **93**:464–468, 1983. 41, 54

[77] K.G. Wilson. **The renormalization group: Critical phenomena and the Kondo problem**. *Rev.Mod.Phys*, **47**:773–840, 1975. 41

[78] S. White. **Density matrix formulation for quantum renormalization groups**. *Phys. Rev. Lett.*, **69**:2863, 1992. 41

[79] G. Vidal. **Efficient Classical Simulation of Slightly Entangled Quantum Computations**. *Phys. Rev. Lett.*, **91**:147902, 2003. 43

[80] A. Quarteroni, R. Sacco, and F. Saleri. *Numerische Mathematik 1*. Springer Lehrbuch, 2002. 43

[81] D. Evans and R. Hoegh-Krohn. **Spectral properties of positive maps on C^*-algebras**. *Jour. of Lon. Math. Soc.*, **17:2**:345–355, 1977. 46

[82] M. Sanz, D. Péres-García, M. Wolf, and Cirac J.I. **A Quantum Version of Wielandt's Inequality**. 2010. 49

[83] G. Fischer. *Lineare Algebra*. Vieweg Studium- Grundkurs Mathematik, 2002. 50

[84] F. Verstraete, J. I. Cirac, J. I. Latorre, E. Rico, and M. M. Wolf. **Renormalization-Group Transformations on Quantum States**. *Phys. Rev. Lett.*, **94**:140601, Apr 2005. 50, 51

[85] O. Szehr and M. Wolf. **Perturbation theory for parent Hamiltonians of Matrix Product States**. 2013. forthcoming, Presented at DPG Tagung, March 2012. 51, 52

[86] D. A. Yarotsky. **Ground states in relatively bounded quantum perturbations of classical lattice systems**. *Comm. in Math. Phys.*, **261(3)**:799–819, 2004. 53

[87] M. B. Hastings. **Solving gapped Hamiltonians locally**. *Phys. Rev. B*, **73**:085115, Feb 2006. 54

[88] I. Affleck, T. Kennedy, E. Lieb, and H. Tasaki. **Rigorous results on valence-bond ground states in antiferromagnets**. *Phys. Rev. Lett.*, **59(7)**:799–802, 1987. 54

[89] S. Michalakis and J. Pytel. **Stability of Frustration-Free Hamiltonians**. *Comm. in Math. Phys*, **322(2)**:277–302, 2013. 56

[90] B. Nachtergaele. **The spectral gap for some quantum spin chains with discrete symmetry breaking**. *Comm. in Math. Phys*, **175**:565–606, 1996. 56

[91] W.L. Spitzer and S. Starr. **Improved bounds on the spectral gap above frustration free ground states of quantum spin chains**. *Lett. Math. Phys.*, **63**:165–177, 2002. 56

Spectral convergence bounds for classical and quantum Markov processes

O. Szehr, D. Reeb and M. Wolf December 17, 2013

We introduce a new framework that yields spectral bounds on norms of functions of transition maps of homogeneous Markov chains. The employed techniques work for classical as well as for quantum Markov chains and we emphasize that our method is new even to the extremely well developed classical theory. Our convergence estimates improve significantly upon the known spectral bounds and are more general in the sense that they do not require additional assumptions like detailed balance, irreducibility or aperiodicity.

1 Spectrum and convergence

Convergence estimates for Markov chains are a classical field of research. Probably the most basic insight in this area is a relation between the asymptotic behaviour of a homogeneous Markov chain and the spectrum of the transition map. If $\mathcal{T}$ and $\mathcal{T}^\infty$ are the transition map and its asymptotic part, respectively, then

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq K\mu^{n-k}n^k,$$

after n time steps. Here μ is the largest in magnitude eigenvalue of $\mathcal{T}$ inside the open unit disc and $k+1$ is the size of the largest corresponding Jordan block. K is constant w.r.t. n, but it depends on $\mathcal{T}$. Estimates of this type occur in innumerable introductory books on Markov chains, see for instance [2, Thm. 1.2], [3, Chp. 3], [1, Fct. 3]. However, the constant K is not specified in the literature such that the mentioned bound does not fully illuminate the relation between spectrum and convergence. In fact, this estimate is merely a qualitative statement about the asymptotic behaviour of the chain, which is unsatisfactory.

2 The method

We consider the following basic task. Given an arbitrary norm $\|\cdot\|$ and a holomorphic function f, obtain an upper bound for $\|f(\mathcal{T})\|$ as a function of the spectrum of $\mathcal{T}$. A simple but crucial observation on this way is that transition maps are *power bounded operators*, meaning that there is a $C \in \mathbb{R}$ so that for all $\mathcal{T}$ and $n \in \mathbb{N}$ we have $\|\mathcal{T}^n\| \leq C$. This property can be exploited in order shift our problem from spaces of operators to function spaces, which offer a plethora of powerful tools to conduct the analysis. In our context, it is natural to consider the *Wiener algebra* of absolutely convergent

holomorphic functions $W := \left\{ f = \sum_{k\geq 0} \hat{f}(k) z^k \mid \|f\|_W = \sum_{k\geq 0} |\hat{f}(k)| < \infty \right\}$, where the $\hat{f}(k)$'s are the Taylor coefficient of f. The reason is that for $f \in W$ we can bound

$$\|f(\mathcal{T})\| \leq \sum_{k\geq 0} |\hat{f}(k)| \left\|\mathcal{T}^k\right\| \leq \sum_{k\geq 0} C|\hat{f}(k)| = C \|f\|_W . \tag{1}$$

For any $f, g \in W$ and $m_{\mathcal{T}}$ the minimal polynomial of $\mathcal{T}$ we have that $\|(f + m_{\mathcal{T}} g)(\mathcal{T})\| = \|f(\mathcal{T}) + m_{\mathcal{T}}(\mathcal{T}) g(\mathcal{T})\| = \|f(\mathcal{T})\|$ and an application of (1) reveals that for all $g \in W$ we have $\|f(\mathcal{T})\| \leq \|f + m_{\mathcal{T}} g\|_W$.

Lemma 1. *Classical and quantum Markovian maps obey a Wiener algebra functional calculus: Let $\|\cdot\|$ be any norm such that for every transition map $\mathcal{T} \in \mathfrak{T}$ we have that $\|\mathcal{T}\| \leq C$. Then for $f \in W$ it holds that*

$$\|f(\mathcal{T})\| \leq C \inf\{\|f + m_{\mathcal{T}} g\|_W \mid g \in W\}.$$

3 A purely spectral bound on the speed of convergence

We use the above observation in order to derive a spectral bound on $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|$.

Theorem 2. *Let $\mathcal{T} \in \mathfrak{T}$ be the transition map of a classical or quantum Markov chain and let $\mathcal{T}_\infty$ be the map describing its limit behaviour. We write $m = m_{\mathcal{T}-\mathcal{T}_\infty}$ for the minimal polynomial, $\sigma(\mathcal{T} - \mathcal{T}_\infty) = \{\lambda_1, ..., \lambda_D\}$ for the spectrum and $\mu = |\lambda_D|$ for the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. Finally, let $\|\cdot\|$ be any norm such that $\|\mathcal{T}\| \leq C$ for all $\mathcal{T} \in \mathfrak{T}$. Then for $n > \frac{\mu}{1-\mu}$ we have*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq 4Ce^2 \sqrt{|m|}(|m|+1) \, \frac{\mu^n}{\left(1 - (1+\frac{1}{n})\mu\right)^{3/2}} \, \mathcal{B}(m,n),$$

$$\mathcal{B}(m,n) \; := \prod_{m/(z-\lambda_D)} \frac{1 - (1+\frac{1}{n})\mu|\lambda_i|}{\mu - |\lambda_i| + \frac{\mu}{n}}.$$

Here, the product is taken over all i such that the corresponding linear factor $(z - \lambda_i)$ occurs in a prime factorization of $m/(z - \lambda_D)$ respecting *multiplicities.*

In contrast to previous estimates Theorem 2 specifies the constant K in Section 1 and provides a quantitative statement. Theorem 2 is the strongest purely spectral convergence estimate even for classical Markov chains.

4 Legal statement

The project was assigned by Michael Wolf. The novel method to obtain spectral estimates for Markov processes as well as the derivation of Theorem 2 are the work of the first author.

References

[1] J.S. Rosenthal. Convergence rates of Markov chains. *SIAM*, 37(1995):387–405, 1993.

[2] E. Seneta. *Non-negative matrices and Markov chains.* Springer Ser. in Stat., 2006.

[3] W. Woess. *Denumerable Markov Chains.* EMS Textbooks in Mathematics, 2009.

Spectral convergence bounds for classical and quantum Markov processes

Oleg Szehr,[1,*] David Reeb,[1,†] and Michael M. Wolf[1,‡]

[1] *Department of Mathematics, Technische Universität München, 85748 Garching, Germany*

(Dated: January 21, 2013)

We introduce a new framework that yields spectral bounds on norms of functions of transition maps for finite, homogeneous Markov chains. The techniques employed work for bounded semigroups, in particular for classical as well as for quantum Markov chains and they do not require additional assumptions like detailed balance, irreducibility or aperiodicity. We use the method in order to derive convergence bounds that improve significantly upon known spectral bounds. The core technical observation is that power-boundedness of transition maps of Markov chains enables a Wiener algebra functional calculus in order to upper bound any norm of any holomorphic function of the transition map. Finally, we discuss how general detailed balance conditions for quantum Markov processes lead to spectral convergence bounds.

Contents

[*] Electronic address: o.dim.qit@googlemail.com
[†] Electronic address: david.reeb@tum.de
[‡] Electronic address: m.wolf@tum.de

I. INTRODUCTION

Across scientific disciplines, Markov chains are ubiquitous in algorithms as well as in models for time evolutions. In many cases one is interested in when their limit behavior is setting in. For algorithms this is often necessary in order to extract the right information and for time evolutions of physical systems this is the time scale on which relaxation or equilibration takes place. Some of the most widespread tools for bounding this time scale are based on the spectrum of the transition map. For time-homogeneous Markov chains with finite state space, the transition map is a stochastic matrix in the context of classical probability distributions and a completely positive trace-preserving map in the quantum case. Since these maps have spectral radius equal to 1, it is somehow clear that only eigenvalues of magnitude 1 survive the limit, that the largest subdominant eigenvalue governs the speed of convergence, and that the rest of the spectrum only matters on shorter time scales. Let $\mathcal{T}$ and $\mathcal{T}_\infty$ be the transition map and its asymptotic part, respectively. We seek convergence estimates of the form

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq K\mu^n \tag{1}$$

after n time steps, where μ is the magnitude of the largest eigenvalue of $\mathcal{T}$ inside the open unit disc and K depends on the spectrum of $\mathcal{T}$, on n and on the dimension of the underlying space. We demand that the dependence of K on n is not exponential, capturing the intuition that the convergence is determined by an exponential decay as μ^n at larger timescales, while for smaller n the whole spectral data is relevant. Such bounds are of general interest for the theory of Markov chains, and they are especially important for stochastic algorithms, which are widely used in statistics and computer science. They are related to the sensitivity of the chain to perturbations [9, 10, 25], are used to study "cut-off" phenomena [2] and random walks on groups [19].

Before describing our main innovation, we mention two traditional, linear algebraic, approaches to bounding convergence times of classical Markov chains as in (1). A Jordan decomposition of the difference $\mathcal{T} - \mathcal{T}_\infty$ yields a bound of the form Equation (1) with $K = k\,\mu^{-d_\mu+1}n^{d_\mu-1}$, where d_μ is the size of the largest Jordan block corresponding to any eigenvalue of magnitude μ and k is constant with respect to n but depends on $\mathcal{T}$ as it is essentially the condition number of the similarity transformation to Jordan normal form. Unfortunately, there is no a priori bound on this condition number. An alternative way is to use Schur's instead of Jordan's normal form. This leads indeed to an expression as in Equation (1) where K can be bounded independent of $\mathcal{T}$, albeit not of n, and we obtain roughly $K \sim \mu^{-D+1}(Dn)^D$, where D is the dimension of the underlying vector space. (See Section III B for details.) Needless to say, this "constant" seems to be far from optimal, especially it does not capture the (correct) asymptotic n-dependence of the Jordan bound.

When proving bounds of the form of Equation (1), one typically employs additional properties of the Markov chain such as detailed balance, irreducibility, aperiodicity, uniqueness of the fixed point, Gibbs distribution of the stationary state, etc.. Clearly, these assumptions are not always fulfilled—in particular in the quantum context detailed balance seems to be a less natural assumption and, furthermore, especially in the area of dissipative quantum computing [28] and dissipative state preparation [3, 6, 28], one aims at preparing rank deficient states.

For classical Markov chains convergence estimates have been widely studied [7, 22] and estimates based on the Jordan and Schur decompositions have been known for many years. Although the latter are generally referred to as spectral convergence bounds, they do not provide a satisfactory spectral desciption of the convergence of a Markov chain. While in case of the Jordan bound it is not possible to compute K in terms of the eigenvalues of $\mathcal{T}$, the Schur bound cannot provide the correct asympotic behavior and does not reflect the full spectral structure of $\mathcal{T}$. So far there is no a priori estimate as in Inequality (1) such that K can *simply be inferred from the localization of the eigenvalues* of $\mathcal{T}$ and such that one obtains the correct asymptotic behavior of the chain. One goal of the present work is to close this gap and to understand what information the spectrum of the transition map of a classical or quantum Markov chain carries about the speed at which it approaches its stationary behavior, i.e., to determine K in terms of the spectrum of $\mathcal{T}$.

Our primary interest lies in the study of classical and quantum Markovian evolutions. However, to obtain a unified picture, in this article we will state our results more generally for bounded semigroups of linear maps. Any endomorphism $\mathcal{T}$ of a vector space $\mathcal{V}$ naturally generates a semigroup consisting of all n-fold concatenations $\mathcal{T}^n$, $n \in \mathbb{N}$. In our analysis we shall assume that the vector space of endomorphisms of $\mathcal{V}$ carries a norm and that the map $\mathcal{T}$ is *power-bounded* with respect to that norm. That means there is a constant C such that, for any n, $\|\mathcal{T}^n\|$ is bounded by C. This is equivalent to saying that the semigroup $(\mathcal{T}^n)_{n\geq 0}$ generated by $\mathcal{T}$ is bounded. The framework of bounded semigroups naturally incorporates both classical and quantum Markov chains (see Section II B).

We start our discussion by analyzing the asymptotic behavior of a bounded semigroup $(\mathcal{T}^n)_{n\geq 0}$. We discuss spectrum related properties of $\mathcal{T}$ that generate a bounded semigroup and define the asymptotic part of the evolution $\mathcal{T}_\infty$ in Section III A. In Section III B we extend the known convergence estimates based on the Jordan and Schur decompositions to cope with bounded semigroups. Implicitly, the analysis covers quantum Markov processes, where we state new convergence estimates. Section IV contains our main result, a convergence estimate with the form of Equation (1), where K is fully determined by n and the spectrum of $\mathcal{T}$. We start Section IV with a mathematical primer, Section IV A, containing an introduction to an entirely new mathematical toolbox in the context of Markov processes. We proceed by analyzing what information can be inferred from the spectrum of $\mathcal{T}$ about the speed at which $(\mathcal{T}^n)_{n\geq 0}$ approaches its asymptotic behavior, Section IV B. The methods, which we employ, enable us in principle to derive spectral bounds on norms of arbitrary functions of transition maps. When applied to power functions, we basically obtain the sought convergence bounds. We discuss in Subsection IV C how our new bound outperforms the convergence estimates based on the Jordan and Schur decompositions.

Nevertheless, it turns out that for many application, such as dissipative quantum com-

putation and state preparation, the convergence estimates obtained still are insufficient to prove the efficiency of a possible implementation. The problem is that the convergence time grows with D, which in turn is exponential in the number of constituent particles (Section IV C). We discuss aspects related to the optimality of our new estimate as well as the convergence speed of contractive Hilbert space semigroups in Section IV D. We prove that stronger estimates, i.e. estimates such that roughly $\log(D)$ time steps bring the chain close to stationarity cannot rely on the spectrum of the transition map alone, the latter simply does not contain sufficient information.

As an approach to better convergence estimates in Section V we extend the detailed balance condition for classical Markov chains and define this property in the context of bounded semigroups, which then includes quantum evolutions. The core theorem of this section is an extension of a convergence estimate that is frequently used to prove cut-off behavior for classical Markov chains (Section V B).

Our discussion focuses on general bounded semigroups but the corresponding statements about classical and quantum Markov processes are implicit, and we will frequently use these for illustration. In what follows one can think of $\mathcal{T}$ either as a quantum channel or an ordinary stochastic matrix.

II. PRELIMINARIES

A. Bounded Semigroups

Throughout this paper $\mathcal{V}$ will be a real or complex vector space of finite dimension D. The set of linear endomorphisms of $\mathcal{V}$ will be denoted by $\mathcal{L}(\mathcal{V})$, which shall be endowed with a norm $\|\cdot\|$. For a given $\mathcal{T} \in \mathcal{L}(\mathcal{V})$ we consider the semigroup $(\mathcal{T}^n)_{n\geq 0} = \{\mathcal{T}^n \,|n \in \mathbb{N}\}$ of linear maps on $\mathcal{V}$ generated by $\mathcal{T}$. Throughout, we assume that $(\mathcal{T}^n)_{n\geq 0}$ is bounded, i.e., there is a constant $C > 0$ such that $\sup_{n\geq 0} \|\mathcal{T}^n\| \leq C$.

Our main approach applies for $(\mathcal{T}^n)_{n\geq 0}$ with a general norm. Nevertheless, for certain results concerning convergence of classical and quantum Markov chains it will be convenient to endow $\mathcal{V}$ with a scalar product $\langle\cdot|\cdot\rangle$. We will consider the induced Hilbert space norm (shortly, 2-norm) $\|v\|_2 = \sqrt{\langle v|v\rangle}$ and the operator norm (shortly, ∞-norm) on $\mathcal{L}(\mathcal{V})$ defined by $\|\mathcal{T}\|_\infty = \sup_{v\neq 0} \frac{\|\mathcal{T}(v)\|_2}{\|v\|_2}$. In some of our examples (e.g., classical Markov chains) it is useful to fix an orthonormal basis $\{e_i\}_{i=1,\ldots,D}$ for $\mathcal{V}$. In this case we write T for the matrix representation of $\mathcal{T}$ with respect to $\{e_i\}_i$, i.e., $T_{ij} = \langle e_i|\mathcal{T}(e_j)\rangle$. We will emphasize whether or not $\mathcal{V}$ has such additional structure in the corresponding sections.

B. Classical and Quantum Markov chains

We briefly review the definitions of classical and quantum Markov chains and discuss certain related concepts.

A classical, finite and time homogeneous Markov process is characterized by a semigroup generated by a classical stochastic matrix T. More precisely, in this scenario $\mathcal{V} \cong \mathbb{R}^D$ equipped with the canonical basis $\{e_i\}_i$ and standard scalar product. The assertion that

T is stochastic is equivalent to $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$. The latter is equivalent to saying that the vector $e = \sum_{i=1}^{D} e_i \in \mathcal{V}$ is fixed by the adjoint map, $T^*(e) = e$. In the context of classical Markov chains the 1-norm plays an exceptional role. For $v \in \mathcal{V}$ we write $v_i = \langle e_i|v\rangle$ and define $\|v\|_1 = \sum_{i=1}^{D} |v_i|$. The induced norm on the set of matrices M acting on $\mathcal{V}$ is called the 1-to-1 norm,

$$\|M\|_{1\to 1} = \sup_{v\neq 0} \frac{\|Mv\|_1}{\|v\|_1}.$$

The 1-to-1 norm and the ∞-norm (i.e. the 2-to-2 norm) are equivalent with

$$D^{-1/2} \|M\|_\infty \leq \|M\|_{1\to 1} \leq D^{1/2} \|M\|_\infty . \tag{2}$$

It is easily seen that $\|T\|_{1\to 1} = 1$ for any stochastic matrix T. We note that if $\|\cdot\|$ is any norm such that $\|T\| \leq C$ holds for all stochastic matrices T, then $\|T^n\| \leq C \ \forall n \in \mathbb{N}$; that is the Markov chain constitutes a bounded semigroup with constant C. Since we are working in finite dimensions, such a semigroup is bounded with respect to any norm.

A time homogeneous quantum Markov chain is also characterized by a semigroup. In the context of quantum evolutions, however, the space $\mathcal{V}$ has different and additional structure. In this article we think of $\mathcal{V}$ as the real vector space consisting of hermitian matrices acting on a complex Hilbert space of dimension d, i.e., $D = d^2$. A matrix $\rho \in \mathcal{V}$ that is positive semidefinite ($\rho \geq 0$) and has unit trace ($\mathrm{tr}[\rho] = 1$) is referred to as a quantum state. An element $\mathcal{T} \in \mathcal{L}(\mathcal{V})$ is called positive iff $X \geq 0$ implies $\mathcal{T}(X) \geq 0$ for any $X \in \mathcal{V}$, and trace-preserving iff $\mathrm{tr}[\mathcal{T}(X)] = \mathrm{tr}[X]\ \forall X \in \mathcal{V}$. $\mathcal{T}$ is trace-preserving iff the adjoint $\mathcal{T}^*$ of $\mathcal{T}$ with respect to the Hilbert-Schmidt inner product $\langle X|Y\rangle = \mathrm{tr}(XY)$ on $\mathcal{V}$ preserves the identity matrix, $\mathcal{T}^*(\mathbb{1}) = \mathbb{1}$. If $\mathcal{T}\otimes\mathcal{I}$, with $\mathcal{I}$ being the operator identity, acts as a positive map on $\mathcal{V}\otimes\mathcal{V}$, then $\mathcal{T}$ is called completely positive [12, 16]. We denote by $\mathfrak{T}$ the subset of $\mathcal{L}(\mathcal{V})$ containing trace preserving and positive maps (TPPMs) and by $\mathfrak{T}_+ \subset \mathfrak{T}$ the set of completely positive maps in $\mathfrak{T}$ (TPCPMs). The latter describe the dynamics of a quantum system, whenever the evolution of the system is independent of its history, and they are called quantum channels in the realm of quantum information theory.

For $X \in \mathcal{V}$ we denote by $\|X\|_1$ the Schatten 1-norm of X. The induced distance $\|\rho - \sigma\|_1$ of two quantum states ρ and σ corresponds to the maximum probability to detect a difference between ρ and σ in an experiment, i.e.

$$\|\rho - \sigma\|_1 = \sup_{\|O\|_\infty \leq 1} |\mathrm{tr}(O(\rho - \sigma))|,$$

where $\|O\|_\infty$ stands for the largest singular value of $O \in \mathcal{V}$. For linear maps $\mathcal{M} \in \mathcal{L}(\mathcal{V})$ we define[1] the induced 1-to-1-norm via

$$\|\mathcal{M}\|_{1\to 1} = \sup_{X\neq 0} \frac{\|\mathcal{M}(X)\|_1}{\|X\|_1}.$$

[1] Note that in this article we define the 1-to-1 norm with the supremum taken over Hermitian matrices. Alternatively, the supremum could be taken over all matrices. The resulting norms are different, but the latter can be upper bounded in terms of 2 times the former.

The diamond norm is the "stabilized version" of the 1-to-1 norm,

$$\|\mathcal{M}\|_\diamond = \|\mathcal{M} \otimes \mathcal{I}\|_{1\to 1}\,,$$

where $\mathcal{I}$ denotes the operator identity in $\mathcal{L}(\mathcal{V})$. It is the dual of the norm of complete boundedness (CB-norm), i.e., we have $\|\mathcal{M}\|_\diamond = \|\mathcal{M}^*\|_{CB}$. The diamond norm and the 1-to-1 norm are equivalent with [23]

$$\|\mathcal{M}\|_{1\to 1} \leq \|\mathcal{M}\|_\diamond \leq D^{1/2}\,\|\mathcal{M}\|_{1\to 1}\,. \tag{3}$$

For any quantum channel $\mathcal{T}$ we have $\|\mathcal{T}\|_{1\to 1} = \|\mathcal{T}\|_\diamond = 1$ [16]. The distance $\|\mathcal{E}-\mathcal{T}\|_\diamond$ of two channels $\mathcal{E}, \mathcal{T}$ measures how well these channels can be distinguished by any quantum experiment. In the quantum context the 1-to-1 norm and the ∞-norm (i.e., the 2-to-2 norm) are equivalent with

$$D^{-1/4}\,\|\mathcal{T}\|_\infty \leq \|\mathcal{T}\|_{1\to 1} \leq D^{1/4}\,\|\mathcal{T}\|_\infty\,. \tag{4}$$

Note that due to the different structure of the underlying space $\mathcal{V}$ in case of quantum Markov chains, the above differs from the Inequalities (2).

If we are given a norm such that $\|\mathcal{T}\| \leq C\ \forall \mathcal{T} \in \mathfrak{T}$ (or $\forall \mathcal{T} \in \mathfrak{T}_+$) the quantum Markov chain generated by $\mathcal{T}$ constitutes a semigroup bounded by C. Again, due to $D < \infty$, this implies that the semigroup is bounded with respect to any norm.

C. Spectral properties

To each linear map $\mathcal{M} \in \mathcal{L}(\mathcal{V})$ we can assign a spectrum $\sigma(\mathcal{M})$ via the usual eigenvalue equation: we have $\lambda \in \sigma(\mathcal{M})$ if and only if there is $X \neq 0$ with $\mathcal{M}(X) = \lambda X$. We write $m_\mathcal{M}$ for the minimal polynomial associated with $\mathcal{M}$ (i.e., the minimal degree, monic polynomial that annihilates $\mathcal{M}$, $m_\mathcal{M}(\mathcal{M}) = 0$) and $|m_\mathcal{M}|$ for the number of linear factors in $m_\mathcal{M}$. Another important object is the Blaschke product associated with $m_\mathcal{M}$,

$$B(z) = \prod_{m_\mathcal{M}} \frac{z-\lambda_i}{1-\bar{\lambda}_i z}, \tag{5}$$

where the product is taken over all i such that the linear factor $z-\lambda_i$ occurs in $m_\mathcal{M}$ *respecting* multiplicities. Thus, the numerator of B as defined here is exactly the corresponding minimal polynomial, $m_\mathcal{M}$.

For convenience, we shall always assume that the eigenvalues in $\sigma(\mathcal{M})$ are arranged such that their magnitudes are non-decreasing. (This ordering is not unique when several eigenvalues have the same magnitude. This ambiguity will, however, be irrelevant in the following. Whenever the situation occurs that we pick an eigenvalue of a certain magnitude $|\lambda|$ we mean that we can take *any* eigenvalue that has this property.)

For any $\mathcal{M} \in \mathcal{L}(\mathcal{V})$ the largest magnitude of all eigenvalues is the spectral radius, which we denote as μ. It follows from Gelfand's formula $\mu = \lim_{k\to\infty} \left\|\mathcal{M}^k\right\|^{1/k}$ [4] that the spectral radius of any element of a bounded semigroup is at most 1. For stochastic matrices and TPPM it is clear that 1 is an eigenvalue of $\mathcal{T}$.

III. LIMITING BEHAVIOR AND CLASSICAL CONVERGENCE ESTIMATES

A. Limiting behavior of $(\mathcal{T}^n)_{n\geq 0}$

In this section we begin our discussion of spectral convergence bounds for semigroups. Based on a spectral decomposition of $\mathcal{T} \in (\mathcal{T}^n)_{n\geq 0}$ we introduce a map $\mathcal{T}_\infty$ and show that this map reflects the behavior of $(\mathcal{T}^n)_{n\geq 0}$ for large n. In the following we extend known spectral convergence bounds for classical Markov chains to the more general semigroup setup. We consider the classical derivations based on the Jordan and Schur decomposition (Section III B). For this reason in this section we assume that $\mathcal{V}$ carries a scalar product.

Our main result Theorem IV.3 will later outperform the bounds proven in this section in terms of convergence speed even in the context of classical Markov chains. Moreover, the techniques introduced there will allow us to consider general norms, which are not induced by a scalar product.

To formalize our intuition that the spectrum of $\mathcal{T}$ determines the convergence properties of $(\mathcal{T}^n)_{n\geq 0}$ let us consider a Jordan decomposition of $\mathcal{T}$,

$$\mathcal{T} = \sum_i \left(\lambda_i \mathcal{P}_i + \mathcal{N}_i\right), \quad \text{with} \tag{6}$$

$$\mathcal{N}_i\mathcal{P}_i = \mathcal{P}_i\mathcal{N}_i = \mathcal{N}_i, \quad \mathcal{P}_i\mathcal{P}_j = \delta_{i,j}\mathcal{P}_i \ \forall i,j. \tag{7}$$

Here, the summation is taken over all distinct eigenvalues of $\mathcal{T}$, the $\mathcal{P}_i$ are projectors whose rank equals the algebraic multiplicity of λ_i and the $\mathcal{N}_i$ are the corresponding nilpotent blocks. All contributions to $\mathcal{T}^n$ that stem from eigenvalues of $\mathcal{T}$ with magnitude smaller than 1 will vanish with increasing n. Hence, we expect the image of $\mathcal{T}^n$ to converge to a subspace of $\mathcal{V}$ spanned by all eigenvectors of $\mathcal{T}$ whose eigenvalues are of magnitude one. We therefore define the linear map $\mathcal{T}_\infty$ whose range is this subspace by

$$\mathcal{T}_\infty := \sum_{|\lambda_i|=1} \lambda_i \mathcal{P}_i, \tag{8}$$

where the $\mathcal{P}_i$ are spectral projectors corresponding to the eigenvalues of $\mathcal{T}$ of magnitude 1. In cases where the spectral radius of $\mathcal{T}$ is strictly smaller than 1, $\mathcal{T}_\infty$ is simply zero. If $\mathcal{T}$ has only one eigenvalue of magnitude one and this eigenvalue is equals 1, then the sequence $\mathcal{T}^n$ converges to $\mathcal{T}_\infty$, which is the unique rank one projection onto the stationary eigenspace of $\mathcal{T}$. In the following lemma we shall prove that $\mathcal{T}_\infty$ mirrors the limit behavior of $(\mathcal{T}^n)_{n\geq 0}$ also in the more general case. More precisely, as n is increasing $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|$ approaches 0 and, for every $k \in \mathbb{N}$, $\mathcal{T}_\infty^k$ indeed is an accumulation point of $(\mathcal{T}^n)_{n\geq 0}$. The latter assertion is relevant especially in the case of classical and quantum Markov chains: the set of stochastic matrices (or quantum channels) constitutes a closed set in the corresponding space, which implies that $\mathcal{T}_\infty$ is again a bona fide stochastic matrix (or quantum channel).

Lemma III.1 (Limiting behavior of $\mathcal{T}^n$)**.** *Let $(\mathcal{T}^n)_{n\geq 0}$ be a semigroup within $\mathcal{L}(\mathcal{V})$ such that $\|\mathcal{T}^n\| \leq C \ \forall n \in \mathbb{N}$ and let $\mathcal{T}_\infty$ be as in Equation* (8)*. Then we have that*
i) all eigenvalues of $\mathcal{T}$ with magnitude 1 have trivial Jordan blocks (i.e., $|\lambda_i| = 1 \Rightarrow \mathcal{N}_i = 0$),
ii) $(\mathcal{T}^n - \mathcal{T}_\infty^n) = (\mathcal{T} - \mathcal{T}_\infty)^n \ \forall \ n \in \mathbb{N}\backslash\{0\}$,

iii) $\lim_{n\to\infty} \|\mathcal{T}^n - \mathcal{T}_\infty^n\| = 0$,
iv) for any $k \in \mathbb{N}$, $\mathcal{T}_\infty^k$ *is contained in the closure of* $(\mathcal{T}^n)_{n\geq 0}$ *in* $\mathcal{L}(\mathcal{V})$,
v) $\left\|\mathcal{T}_\infty^k\right\| \leq C \ \forall k \in \mathbb{N}$.

Proof. i) We proceed by contradiction and consider $\|\mathcal{T}^n\|_\infty$. Since $(\mathcal{T}^n)_{n\geq 0}$ is bounded and in finite dimensions all norms are equivalent there is $0 < K_1 < \infty$ with $\|\mathcal{T}^n\|_\infty \leq K_1$. On the other hand there is $K_2 > 0$ with $\|\mathcal{T}^n\|_\infty \geq K_2 \left\|\sum_j (\lambda_j \mathcal{P}_j + \mathcal{N}_j)^n\right\|_\infty$. If λ_i has a non-trivial Jordan block the latter can be lower bounded by $\left\|\sum_j (\lambda_j \mathcal{P}_j + \mathcal{N}_j)^n\right\|_\infty \geq |\lambda_i|^{n-1} n$. It follows that if $|\lambda_i| = 1$ and λ_i has a non-trivial Jordan block then $\|\mathcal{T}^n\|_\infty$ grows unboundedly with n.[2]
ii) follows from the relations in (7) since $\mathcal{T}\mathcal{T}_\infty = \mathcal{T}_\infty \mathcal{T} = \mathcal{T}_\infty^2$. For $n > 2$ the statement follows by induction.
iii) By the previous assertion $\|T^n - T_\infty^n\| = \|(T - T_\infty)^n\|$ holds. The spectral radius μ of the map $\mathcal{T} - \mathcal{T}_\infty$ is strictly smaller than 1. We have from Gelfand's formula that $\lim_{n\to\infty} \|T^n - T_\infty^n\|^{1/n} = \mu < 1$ and hence for all n sufficiently large $\|T^n - T_\infty^n\| \leq \left(\frac{1+\mu}{2}\right)^n$. With increasing n the right hand side goes to 0 and the claim follows.
iv) We prove that for fixed k there is a subsequence $(T^{n_l})_l$ that converges to $\mathcal{T}_\infty^k$, that is $\lim_{l\to\infty} \left\|\mathcal{T}^{n_l} - \mathcal{T}_\infty^k\right\| = 0$. To achieve this we subdivide $\mathcal{V}$ into the invariant subspace of $\mathcal{T}$ corresponding to all eigenvalues of magnitude 1 and its complement. On the latter subspace we can directly invoke iii). On the former subspace, it is sufficient to find, for any $\epsilon > 0$, a subsequence $(T^{n_l})_l$ with the property that $|\lambda_i^{n_l} - \lambda_i^k| \leq \epsilon$ simultaneously for all i with $|\lambda_i| = 1$. The existence of such a subsequence follows from Dirichlet's Theorem on simultaneous Diophantine approximation, [21] Theorem 1B.
v) By iv) for any $k \in \mathbb{N}$ and any $\epsilon > 0$ there is n such that $\left\|\mathcal{T}^n - \mathcal{T}_\infty^k\right\| \leq \epsilon$. This implies that $\left\|\mathcal{T}_\infty^k\right\| \leq C + \epsilon \ \forall \epsilon > 0$ and hence the claimed inequality. □

B. Jordan and Schur convergence estimates

Our next aim is to understand qualitatively by how much for certain n the evolution $\mathcal{T}^n$ differs from its limit behavior, i.e., how small the quantity $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty = \|T^n - T_\infty^n\|_\infty$ is for any bounded semigroup. We shortly review two standard methods to obtain such estimates. Both methods rely on the fact that $\mathcal{T}^n - \mathcal{T}_\infty^n = (\mathcal{T} - \mathcal{T}_\infty)^n$ and perform a transformation of $T - T_\infty$ to upper triangular form. While the first approach is to choose the Jordan normal form for $T - T_\infty$, the second one is based on the Schur decomposition. Both decompositions involve a similarity transformation A that brings $T - T_\infty$ to upper triangular form, i.e., $T - T_\infty = A(\Lambda + N)A^{-1}$ with diagonal Λ and nilpotent N. While in case of Jordan decomposition $\Lambda + N$ has Jordan block structure, for the Schur decomposition A is unitary.

Theorem III.2. *Let* $(\mathcal{T}^n)_{n\geq 0}$ *be a bounded semigroup in* $\mathcal{L}(\mathcal{V})$, *let* $\mathcal{T}_\infty$ *be the map introduced in* (8) *and let* μ *be the spectral radius of* $\mathcal{T} - \mathcal{T}_\infty$. *Then there are constants* $C_1, C_2 > 0$

[2] See also the derivation of the lower bound in Theorem III.2.

such that, for all $n \geq 1$,

$$C_1 \mu^{n-d_\mu+1} n^{d_\mu-1} \leq \|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq C_2 \mu^{n-d_\mu+1} n^{d_\mu-1},$$

where d_μ *is the size of the largest Jordan block corresponding to any eigenvalue of* $\mathcal{T} - \mathcal{T}_\infty$ *of magnitude* μ.

Proof. We first state an upper bound on $\|(\Lambda + N)^n\|_\infty$ with diagonal Λ and nilpotent upper-triangular N. We note that any monomial in N and Λ vanishes if the total degree of N is larger than or equal to $jD - 1$. Using this together with the triangle inequality in the binomial expansion and exploiting the sub-multiplicativity of the ∞-norm we find

$$\|(\Lambda + N)^n\|_\infty \leq \sum_{k=0}^{\min\{n, D-1\}} \binom{n}{k} \|N\|_\infty^k \|\Lambda\|_\infty^{n-k}. \tag{9}$$

Let now $J(\lambda_i)$ be a Jordan block with diagonal part $\lambda_i \mathbb{1}$ and nilpotent part N_i. We consider the Jordan decomposition $T - T_\infty = A \left(\bigoplus_{i,\nu} J_\nu(\lambda_i) \right) A^{-1}$, where the summation goes over i, which labels the different eigenvalues of $T - T_\infty$, and over ν, which enumerates the Jordan blocks corresponding to an eigenvalue λ_i. We introduce the constant $\kappa = \inf \left(\|A\|_\infty \left\|A^{-1}\right\|_\infty \right)$ where the infimum is taken over all A that bring $T - T_\infty$ to Jordan form. It follows readily that

$$\kappa^{-1} \|J^n\|_\infty \leq \|T^n - T_\infty^n\|_\infty \leq \kappa \|J^n\|_\infty \tag{10}$$

with $J = \bigoplus_{i,\nu} J_\nu(\lambda_i)$. For any $J_\nu(\lambda_i)$ there is an n_0 such that for all $n \geq n_0$ one has $\|J_\nu(\lambda_i)^n\|_\infty \leq \|J_{max}(\lambda_{max})^n\|_\infty$, where $J_{max}(\lambda_{max})$ denotes the largest Jordan block corresponding to an eigenvalue λ_{max} of modulus μ. Therefore, to find an upper bound on the right hand side of (10) we can subdivide $J_{max}(\lambda_{max})$ in a nilpotent and a diagonal part and use Inequality (9). We note that for $k \leq d_\mu - 1$ we can bound $\binom{n}{k} \leq n^{d_\mu - 1}$ and taking everything together we obtain for large enough n

$$\|T^n - T_\infty^n\|_\infty \leq \kappa \sum_{k=0}^{d_\mu - 1} n^{d_\mu - 1} \mu^{n-d_\mu+1},$$

which proves the upper bound in Theorem III.2 since it can be extended to an upper bound valid for any $n \in \mathbb{N}$ by a rescaling of C_1. The lower bound is a consequence of the following inequalities for $n \geq d_\mu - 1$

$$\mu^{n-d_\mu+1} \binom{n}{d_\mu - 1} \leq \|J_{max}(\lambda_{max})^n\|_\infty \leq \|J^n\|_\infty.$$

□

One problem with the above proof is that n_0 and thus C_2 can get large if there is a sub-dominant eigenvalue close to the spectral radius. Another issue is that one cannot a priori bound κ for general T. Consequently, only little is known about C_1 and C_2. Most

awkward, C_1 and C_2 depend on the given channel $\mathcal{T}$, i.e. are not universal for all channels of a given dimension. For this reason Theorem III.2 is a qualitative statement about the asymptotic behavior of the semigroup. In contrast, the Schur decomposition allows us to state an upper bound on the rate of convergence that only depends on n, D and μ. This goes at the price of a rather pessimistic estimate.

Theorem III.3. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a bounded semigroup in $\mathcal{L}(\mathcal{V})$ such that $\|\mathcal{T}^n\|_\infty \leq C\ \forall n \in \mathbb{N}$ and let μ be the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. For any $n \in \mathbb{N}$ it holds that*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq 2\mu^{n-D+1} n^{D-1} (\mu + 2C)^{D-1}.$$

Proof. As already mentioned, this will be proven based on the Schur decomposition $T - T_\infty = U(\Lambda + N)U^\dagger$, where U is unitary. As before we can rely on the binomial expansion Inequality (9). We note that $\|U\|_\infty = 1$ and that for $n > 1$

$$\sum_{k=0}^{D-1} \binom{n}{k} \leq \sum_{k=0}^{D-1} n^k \leq 2n^{D-1}.$$

Thus, using the sub multiplicativity of the ∞-norm it follows from (9) that

$$\|T^n - T_\infty^n\|_\infty \leq 2n^{D-1}\mu^{n-D+1} \max\left(1, \|N\|_\infty^{D-1}\right).$$

In addition we have that $N = T - T_\infty - \Lambda$ and therefore $\|N\|_\infty \leq 2C + \mu$. □

To obtain a convergence estimate for Markov chains in 1-to-1 norm we can rely on the Inequalities (2). The corresponding statement of Theorem III.2 is immediate. Analogously, Theorem III.2 can be used to estimate the speed of convergence of TPPMs in 1-to-1 and diamond norm via the Inequalities (4), (3).

Due to the lower bound in (2) the singular values of stochastic matrices are bounded by $D^{1/2}$ from which we infer that $C \leq D^{1/2}$ in this case. For positive, trace preserving maps the singular values are bounded by $D^{1/4}$ ([17], or by the norm equivalence (4) and the fact that $\|\mathcal{T}\|_{1\to 1} = 1$). Thus, Theorem III.3 includes a convergence bound for both classical stochastic matrices and TPPMs. For a more detailed discussion of the resulting estimates in the quantum context see Subsection IV C.

IV. MAIN RESULT: SPECTRUM AND CONVERGENCE

The main contribution of this article is to introduce a new formalism that yields spectral bounds on norm of functions of transition maps of Markov processes and to apply this formalism to prove new estimates for the convergence of a such processes to stationarity. The core technical innovation will be to employ a Wiener algebra functional calculus in the context of bounded semigroups. To prove our estimates we will rely on the theory of function algebras, functional calculi and model spaces. To our knowledge these concepts have not found their way into the theory of classical or quantum Markov processes so far. For this reason at first we briefly introduce the mathematical framework in Subsection IV A.

A detailed introduction to the mathematics involved goes beyond the scope of this article and we refer to [13–15] for this. In Subsection IV B we employ the mathematical machinery to the context of bounded semigroups and derive the main theorem. In the subsequent subsection we discuss our main result and compare it to the convergence estimates from Jordan and Schur decompositions.

A. Function spaces and functional calculi

In this subsection we discuss the problem of bounding the norm of an operator acting on a Hilbert or Banach space in general. In the purely mathematical literature this problem is studied extensively [13, 15] and in this exposition we follow the theory developed in [14]. The main idea is to consider certain spaces of functions and to relate the operator under consideration with a certain function. The norm of the operator is then bounded by bounding the norm of its "function representative''.

We begin by defining the function spaces, which will be relevant in our discussion. The space of analytic functions on the open unit disc $\mathbb{D} = \{z \in \mathbb{C} || z| < 1\}$ is denoted by $Hol(\mathbb{D})$. We will be concerned with certain subspaces of $Hol(\mathbb{D})$, an important class of which constitute the Hardy spaces. For $p > 0$ those are defined as

$$H^p := \big\{ f \in Hol(\mathbb{D}) |\, \|f\|_{H^p}^p := \sup_{0\leq r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\phi})|^p \mathrm{d}\phi < \infty \big\},$$

and

$$H^\infty := \big\{ f \in Hol(\mathbb{D}) |\, \|f\|_{H^\infty} := \sup_{z\in\mathbb{D}} |f(z)| < \infty \big\}.$$

It is immediate from the definition that the spaces H^p are vector spaces, that the mapping $f \mapsto \|f\|_{H^p}$ is a norm for $p \geq 1$ and that $H^p \subset H^q$ for $p \geq q$. In the special case $p = 2$ the Hardy norm can be written using the Taylor coefficients of the analytic function f. More precisely, we write $f(z) = \sum_{k\geq 0} \hat{f}(k) z^k$ and use Parseval's identity to conclude that

$$\sup_{0\leq r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\phi})|^2 \mathrm{d}\phi = \sum_{k\geq 0} |\hat{f}(k)|^2.$$

Thus, $f \in Hol(\mathbb{D})$ is in H^2 if and only if $\sum_{k\geq 0} |\hat{f}(k)|^2 < \infty$ (see [15], p. 32). The Wiener algebra is defined as the subset of $Hol(\mathbb{D})$ of absolutely convergent Taylor series,

$$W := \{ f = \sum_{k\geq 0} \hat{f}(k) z^k | \sum_{k\geq 0} |\hat{f}(k)| < \infty \}.$$

For a given class of operators (for instance Hilbert space contractions or power bounded operators) the associated function algebra is a space of analytic functions that mirrors the "boundedness properties" of those operators. A functional calculus is a map that associates operators from the given class and elements of the function algebra and relates the norms of an operator and its representative in the function algebra. More precisely we have the following definitions [14]:

Definition IV.1 (Function algebra)**.** *A unital Banach algebra A with elements in $Hol(\mathbb{D})$ will be called a function algebra, if*

1. *A contains all polynomials and $\lim_{n\to\infty} \|z^n\|_A^{1/n} = 1$.*
2. *$(a \in A,\ \lambda \in \mathbb{D},\ a(\lambda) = 0) \Rightarrow \frac{a}{z-\lambda} \in A$.*

Definition IV.2 (Functional calculus)**.** *Let $X : \mathcal{B} \to \mathcal{B}$ be an operator on a Banach space $\mathcal{B}$. A bounded algebra homomorphism from a function algebra A into the set of linear operators on $\mathcal{B}$,*

$$\mathcal{J}_X :\ A \to L(\mathcal{B}),$$

will be called a functional calculus for X, if it satisfies $\mathcal{J}_X(z) = X$ and $\mathcal{J}_X(1) = \mathbb{1}$.

(In our case it is sufficient to assume that $\mathcal{B}$ has finite dimension.) Intuitively $\mathcal{J}_X$ captures the notion of "plugging an operator into a function", that is for $a \in A$ we have $a(X) = \mathcal{J}_X(a)$ and by the boundedness property there is a constant C_X such that

$$\|a(X)\| \leq C_X \|a\|_A .$$

Given a family Γ of operators we say that this family obeys a functional calculus with constant C if each $X \in \Gamma$ admits a functional calculus with $C_X \leq C$. Thus, one approach to the problem of bounding the norm $\|a(X)\|$ for $X \in \Gamma$ is by constructing a functional calculus for the family Γ and then bounding the norm of a in the function algebra. For us, two instances of functional calculi will be important. In the first example we consider power-bounded Banach spaces operators, while the second one treats Hilbert space contractions.
i) Consider a family $\Gamma = \{X \in L(\mathcal{B}) |\, \|X^n\| \leq C\ \forall n \in \mathbb{N}\}$ of Banach space operators that are power bounded by some constant C. This family admits a Wiener algebra functional calculus since for any $f \in W$ and $X \in \Gamma$

$$\|f(X)\| = \left\| \sum_{k\geq 0} \hat{f}(k) X^k \right\| \leq \sum_{k\geq 0} |\hat{f}(k)| \left\| X^k \right\| \leq C \sum_{k\geq 0} |\hat{f}(k)| = C \|f\|_W \tag{11}$$

holds.
ii) In Section IV D we discuss the semigroup of Hilbert space contractions $\Gamma = \{X \in L(\mathcal{H}) |\, \|X\|_\infty \leq 1\}$. This family allows for an H^∞ functional calculus (with constant $C = 1$), since by von Neumann's inequality [16] we have for any $f \in H^\infty$ and $X \in \Gamma$

$$\|f(X)\|_\infty \leq \|f\|_{H^\infty} . \tag{12}$$

At first glance, the outlined procedure seems to be of little use since the right hand sides of (11), (12) do not depend on X anymore. To obtain a better bound one can rely on the following insight. Recall that the minimal polynomial m_X annihilates the corresponding operator, i.e., $m_X(X) = 0$. Instead of considering the function a directly, we add multiples of $m = m_X$ (or any other annihilating polynomial) to this function and consider $c = a + mb$, $b \in A$ instead of a. It is immediate that $\|a(X)\| = \|c(X)\|$. The following simple but crucial lemma summarizes this point:

Lemma IV.1 ([14], Lemma 3.1)**.** *Let $m \neq 0$ be a polynomial and let Γ be a set of operators that obey an A functional calculus with constant C and that satisfy $m(X) = 0\ \forall X \in \Gamma$. Then*

$$\|a(X)\| \leq C \|a\|_{A/mA}, \ \forall X \in \Gamma,$$

where $\|a\|_{A/mA} = \inf \{\|c\|_A |\ c = f + mb,\ b \in A\}$.

Proof. For any $b \in A$ we have that $\|a(X)\| = \|(a + mb)(X)\| \leq C \|a + mb\|_A$. □

B. Spectral bounds for the convergence of Markovian processes to stationarity

Crucial for the main result Theorem IV.3 is that classical stochastic matrices and quantum channels both obey a power-boundedness condition. Given any norm $\|\cdot\|$ such that every $\mathcal{T} \in \mathfrak{T}$ satisfies $\|\mathcal{T}\| \leq C$, then for all $n \geq 0$, $\|\mathcal{T}^n\| \leq C$, i.e., $\mathcal{T}$ generates a bounded semigroup $(\mathcal{T}^n)_{n\geq 0}$. In view of Lemma IV.1 this entails that $(\mathcal{T}^n)_{n\geq 0}$ obeys a Wiener algebra functional calculus with $\|f(\mathcal{T})\| \leq C \|f\|_{W/mW}$. Although this observation is simple, we state it in a separate theorem to emphasize its importance.

Theorem IV.2. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a semigroup bounded with constant C and let m be the minimal polynomial of $\mathcal{T}$, $m(\mathcal{T}) = 0$. Then*

$$\|f(\mathcal{T})\| \leq C \|f\|_{W/mW}$$

holds for any function $f \in W$.

Theorem IV.2 can be used to bound various functions of transition maps of Markovian evolutions. For instance one might be interested in bounding the norm of the inverse of a transition map (if it exists). In [14] an estimate of X^{-1} is derived for an algebraic Banach space operator X by using Lemma IV.1 and bounding $\|z^{-1}\|_{W/mW}$. This estimate immediately carries over to Markov transition maps. In this article we seek bounds for the rate of convergence of a semigroup; we will use Theorem IV.2 to relate this problem to the one of bounding $\|z^n\|_{W/mW}$. In this way we obtain the following:

Theorem IV.3. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a semigroup bounded by C, and let $\mathcal{T}_\infty$ be its asymptotic evolution introduced in* (8)*. We write $m = m_{\mathcal{T}-\mathcal{T}_\infty}$ for the minimal polynomial and μ for the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$ and B for the Blaschke product* (5) *associated with m. Then, for $n > \frac{\mu}{1-\mu}$ we have*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq \mu^{n+1} \frac{4Ce^2\sqrt{|m|}(|m|+1)}{n\left(1 - (1+\frac{1}{n})\mu\right)^{3/2}} \sup_{|z|=\mu(1+1/n)} \left|\frac{1}{B(z)}\right|.$$

Before we proceed to the proof of Theorem IV.3 let us discuss some immediate consequences. First, note that the condition $n > \mu/(1-\mu)$ does not significantly restrict the range of n, where the theorem applies. For $n \leq \mu/(1-\mu)$ it holds that the exponentially

decaying factor $\mu^n \gtrapprox e^{-\mu}$ is still of order 1. In this range bounds of the form (1) only yield a trivial statement.

As compared to Theorem III.2 and Theorem III.3 the bound in Theorem IV.3 depends more explicitly on the spectral properties of $\mathcal{T} - \mathcal{T}_\infty$. The Jordan block structure of $\mathcal{T} - \mathcal{T}_\infty$ is reflected by the fact that the formula contains a certain factor for each factor of $m_{\mathcal{T}-\mathcal{T}_\infty}$. In contrast to Theorems III.2, III.3, Theorem IV.3 clarifies in which way the Jordan structure of $\mathcal{T} - \mathcal{T}_\infty$ influences the speed of convergence of a Markov process.

The upper bound in Theorem IV.3 can be made more explicit by taking the supremum over all factors in the Blaschke product individually. It is not difficult to see (see Appendix A) that for $|\lambda| < \mu(1 + 1/n) \leq 1$ one has

$$\sup_{|z|=\mu(1+1/n)} \left| \frac{1 - \bar{\lambda} z}{z - \lambda} \right| = \frac{1 - (1 + 1/n)\mu|\lambda|}{\mu - |\lambda| + \mu/n}. \tag{13}$$

This leads to the following corollary:

Corollary IV.4. *Let $\sigma(\mathcal{T} - \mathcal{T}_\infty) = \{\lambda_1, ..., \lambda_D\}$ be the spectrum of $\mathcal{T} - \mathcal{T}_\infty$ so that the magnitudes are ordered non-decreasingly and let $\mu = |\lambda_D|$ be the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. Under the assumptions of Theorem IV.3 it holds that*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq \mu^n \frac{4Ce^2\sqrt{|m|}(|m| + 1)}{\big(1 - (1 + \frac{1}{n})\mu\big)^{3/2}} \prod_{m/(z-\lambda_D)} \frac{1 - (1 + \frac{1}{n})\mu|\lambda_i|}{\mu - |\lambda_i| + \frac{\mu}{n}},$$

where the product is taken over all i such that the corresponding linear factor $(z - \lambda_i)$ occurs in a prime factorization of $m/(z - \lambda_D)$, respecting *multiplicities and λ_D stands for* any *eigenvalue of magnitude μ.*

Every eigenvalue of magnitude μ contributes one factor proportional to n/μ in Equation (13). Whereas Theorem IV.3 contains an inverse Blaschke factor for each linear factor in the minimal polynomial m, in Corollary IV.4 we have canceled one of the factors corresponding to the spectral radius μ by the μ/n prefactor in Theorem IV.3.

A detailed discussion of Theorem IV.3 and Corollary IV.4 follows in Section IV C. Here, let us just mention some situations in which the above bound might be applied.

1. When $\mathcal{T}$ is the transition matrix of a classical time-homogenous Markov chain, Theorem IV.3 can be used to estimate the distance of $\mathcal{T}^n$ to stationarity. For the classical 1-to-1 norm it holds that $\|\mathcal{T}^n\|_{1\to 1} = 1$ for any stochastic matrix and any natural number n, such that Theorem IV.3 applies with $C = 1$.

2. For all $\mathcal{T} \in \mathfrak{T}_+$ and any n we have that $\|\mathcal{T}^n\|_\diamond = 1$. Thus, Theorem IV.3 provides a convergence bound for quantum Markov chains with $C = 1$.

3. Theorem IV.3 holds for general power bounded operators (in finite dimensions) whose spectrum is contained in the unit disc. Therefore our result applies to cone- and base-preserving maps with the corresponding norms, more general than transition matrices of classical Markov chains and TPPMs. An important class of such operations constitute LOCC maps [18].

4. In the context of classical and quantum Markov chains one is often interested in the quantity $\|\mathcal{T}^n(v) - \mathcal{T}_\infty^n(v)\|_1$, where, depending on the context, v is either a probability vector or a quantum state. If v is contained in an invariant subspace $\mathcal{V}_{inv}$ of $\mathcal{T}$ it is clear that one can improve the bound in Theorem IV.3. We then have that

$$\begin{aligned}\|\mathcal{T}^n(\rho) - \mathcal{T}_\infty^n(\rho)\|_1 &\leq \frac{4e^2\sqrt{|m|}(|m|+1)\mu^{n+1}}{n\left(1-(1+\frac{1}{n})\mu\right)^{3/2}} \sup_{|z|=\mu(1+1/n)} \left|\frac{1}{B(z)}\right| \\ &\leq \frac{4e^2\sqrt{|m|}(|m|+1)\mu^{n}}{\left(1-(1+\frac{1}{n})\mu\right)^{3/2}} \prod_{m/(z-\lambda_D)} \frac{1-(1+\frac{1}{n})\mu|\lambda_i|}{\mu-|\lambda_i|+\frac{\mu}{n}},\end{aligned}$$

where now $B = B_{(\mathcal{T}-\mathcal{T}_\infty)_{inv}}$ is the Blaschke product corresponding to the minimal polynomial $m = m_{(\mathcal{T}-\mathcal{T}_\infty)_{inv}}$ of $\mathcal{T} - \mathcal{T}_\infty$ restricted to $\mathcal{V}_{inv}$.

For the proof we present an upper bound on $\|\mathcal{A}^n\|$ for a general power bounded operator $\mathcal{A}$, whose spectrum is contained in $\mathbb{D}$ and we specialize to the case $\mathcal{A} = \mathcal{T} - \mathcal{T}_\infty$ only at the end. More precisely, we start with any $\mathcal{A} \in \mathcal{L}(\mathcal{V})$ whose spectrum is contained in the open unit disc and suppose that $\|\mathcal{A}^n\| \leq C$ for all $n \in \mathbb{N}$. For convenience we assume that the eigenvalues $\{\lambda_1, ..., \lambda_D\}$ of $\mathcal{A}$ are ordered with non-decreasing magnitude.

In what follows we shall assume that the map $\mathcal{A}$ is diagonalizable, i.e., its minimal polynomial decomposes into pairwise distinct linear factors. This assumption does not lead to any difficulties when it comes to finding upper bounds of the type of Theorem IV.3. To see this, assume that, for each fixed n, Theorem IV.3 holds true for any $\mathcal{A}$ such that the minimal polynomial $m_{\mathcal{A}}$ decomposes into pairwise distinct linear factors. To pass to the case when $\mathcal{A}$ has non-trivial Jordan structure one slightly perturbs the spectrum of $\mathcal{A}$ and obtains a diagonalizable map $\mathcal{A} + \epsilon$. Note that for sufficiently small ϵ the spectrum of $\mathcal{A} + \epsilon$ still is contained in the open unit disc, such that $\mathcal{A} + \epsilon$ is power bounded with some constant C_ϵ. In the limit of $\epsilon \to 0$, C_ϵ converges to C [14]. Thus, for each fixed n one can apply the theorem for diagonalizable matrices and pass to the limit $\epsilon \to 0$ on both sides of Theorem IV.3. By continuity of the norm this implies the claimed statement.

Proof of Theorem IV.3. We adapt techniques developed in [14] for general power bounded operators (see Theorem 3.20) and invoke Lemma IV.1 to transfer the problem of estimating $\|\mathcal{A}^n\|$ to the one of bounding $\|z^n\|_{W/mW}$. It follows from the definition of the function algebra, that [14]

$$\|z^n\|_{W/mW} = \inf\{\|g\|_W \mid g \in W,\, g(\lambda_i) = \lambda_i^n\}. \tag{14}$$

This means that the problem of bounding $\|z^n\|_{W/mW}$ is equivalent to finding a minimal norm function g that interpolates the data set $(\lambda_1, \lambda_1^n), ..., (\lambda_{|m|}, \lambda_{|m|}^n)$ in the sense that $g(\lambda_i) = \lambda_i^n$. (More generally, the problem of bounding a function f of a quantum channel is related to an interpolation problem in the Wiener algebra by replacing λ_i^n by $f(\lambda_i)$.) The strategy of our proof will be to consider one specific representative function g in (14) and bound its norm. To achieve this we employ the following method. Instead of considering g directly we choose a "smoothing parameter" r and pass to a "stretched" interpolation function.

Given any function $f \in H^2$ and $r \in (0,1)$, we write $f_r(z) := f(rz) = \sum_{k\geq 0} \hat{f}(k) r^k z^k$ and observe that by the Cauchy-Schwarz Inequality

$$\|f_r\|_W \leq \sqrt{\sum_{k\geq 0} |\hat{f}(k)|^2} \sqrt{\frac{1}{1-r^2}} \leq \|f\|_{H^\infty} \sqrt{\frac{1}{1-r^2}}. \tag{15}$$

This idea was used to obtain bounds to the inverse and resolvent of a power bounded operator in [14].
We use the Blaschke products $B(z) = \Pi_i \frac{z-\lambda_i}{1-\bar{\lambda}_i z}$ and $\tilde{B}(z) = \Pi_i \frac{z-r\lambda_i}{1-r\bar{\lambda}_i z}$, where in the latter product the spectrum is stretched by a factor of r. (The products are taken over all prime factors of $m_{\mathcal{A}}$, but to avoid cumbersome notation we do not write this explicitly.) Consider now the function g with

$$g(z) = \sum_k \lambda_k^n \frac{B(z)}{z-\lambda_k} (1-|\lambda_k|^2) \prod_{j\neq k} \frac{1-\bar{\lambda}_j \lambda_k}{\lambda_k - \lambda_j}.$$

g is analytic in the unit disc and $g(\lambda) = \lambda^n$ for all $\lambda \in \sigma(T)$. To be able to use the estimate (15) we perform the aforementioned smoothing. We define the modified function $\tilde{g}$ by

$$\tilde{g}(z) = \sum_k \lambda_k^n \frac{\tilde{B}(z)}{z-r\lambda_k} (1-r^2|\lambda_k|^2) \prod_{j\neq k} \frac{1-r^2\bar{\lambda}_j \lambda_k}{r\lambda_k - r\lambda_j}$$

and observe that $\tilde{g}_r$ enjoys the same basic properties as g, i.e., $\tilde{g}_r$ is analytic in $\mathbb{D}$ and $\tilde{g}_r(\lambda) = \lambda^n$ for any $\lambda \in \sigma(T)$. Thus, by Equation (14), we have that $\|z^n\|_{W/mW} \leq \|\tilde{g}_r\|_W$ and it follows from Inequality (15) that

$$\|\tilde{g}_r\|_W \leq \sqrt{\frac{1}{1-r^2}} \, \|\tilde{g}\|_{H^\infty}.$$

By the Maximum Principle for analytic functions $\|\tilde{g}\|_{H^\infty}$ is attained on the unit circle, that is $\|\tilde{g}\|_{H^\infty} = \sup_{|z|=1} |\tilde{g}(z)|$. Exploiting the fact that each elementary Blaschke factor preserves the unit circle, we conclude that

$$\|\tilde{g}\|_{H^\infty} = \sup_{|z|=1} \left| \sum_k \lambda_k^n \frac{1-r^2|\lambda_k|^2}{z-r\lambda_k} \prod_{j\neq k} \frac{1-r^2\bar{\lambda}_j\lambda_k}{r\lambda_k - r\lambda_j} \right|.$$

To bound this quantity we perform a contour integration along the circle $\gamma : \phi \mapsto se^{i\phi}$, where $s < 1$ is chosen in a way such that γ encircles all eigenvalues of $\mathcal{A}$. By the Residue Theorem (note that $|z| = 1$) we have that

$$\sum_k \lambda_k^n \frac{1-r^2|\lambda_k|^2}{z-r\lambda_k} \prod_{j\neq k} \frac{1-r^2\bar{\lambda}_j\lambda_k}{r\lambda_k - r\lambda_j} = \frac{1}{2\pi i} \int_\gamma \frac{\lambda^n}{\tilde{B}_r(\lambda)} \frac{1}{z-r\lambda} \mathrm{d}\lambda. \tag{16}$$

Integration by parts gives

$$\frac{1}{2\pi i}\int_\gamma \frac{\lambda^n}{\tilde{B}_r(\lambda)}\frac{1}{z-r\lambda}\mathrm{d}\lambda = -\frac{1}{2\pi i(n+1)}\int_\gamma \lambda^{n+1}\left[\frac{1}{\tilde{B}_r(\lambda)(z-r\lambda)}\right]'\mathrm{d}\lambda \tag{17}$$

and we arrive at

$$\|\tilde{g}\|_{H^\infty} \le \frac{s^{n+1}}{2\pi(n+1)}\sup_{|z|=1}\int_\gamma \left|\left[\frac{1}{\tilde{B}_r(\lambda)(z-r\lambda)}\right]'\right||\mathrm{d}\lambda|.$$

The right hand integral can be interpreted as the arc length of the image of γ under the rational function $\frac{1}{\tilde{B}_r(\lambda)(z-r\lambda)}$. For this quantity we have by Spijker's Lemma ([24], Equation (4))

$$\int_\gamma \left|\left[\frac{1}{\tilde{B}_r(\lambda)(z-r\lambda)}\right]'\right||\mathrm{d}\lambda| \le 2\pi(|m|+1)\sup_{|\lambda|=s}\left|\frac{1}{\tilde{B}_r(\lambda)(z-r\lambda)}\right|$$

and conclude that for $0<r<1$ and $\mu<s<1$ we have

$$\|\tilde{g}\|_{H^\infty} \le s^{n+1}\frac{(|m|+1)}{(n+1)}\frac{1}{1-rs}\sup_{|\lambda|=s}\left|\prod_i \frac{1-\bar{\lambda}_i r^2\lambda}{r\lambda - r\lambda_i}\right|.$$

In the above bound we choose $s=(1+1/n)\mu$ (where μ denotes the spectral radius of $\mathcal{A}$) and notice that

$$s^{n+1} = \mu^{n+1}\left(1+\frac{1}{n}\right)^{n+1} \le e(1+1/n)\mu^{n+1},$$

which entails

$$\|\tilde{g}\|_{H^\infty} \le \frac{\mu^{n+1}(|m|+1)e}{nr^{|m|}(1-r(1+1/n)\mu)}\sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\prod_i \frac{1-\bar{\lambda}_i r^2\lambda}{\lambda-\lambda_i}\right|$$

and

$$\|z^n\|_{W/mW} \le \sqrt{\frac{1}{1-r^2}}\frac{\mu^{n+1}(|m|+1)e}{nr^{|m|}(1-r(1+1/n)\mu)}\sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\prod_i \frac{1-\bar{\lambda}_i r^2\lambda}{\lambda-\lambda_i}\right|.$$

Finally, we observe that

$$\begin{aligned}\sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\prod_i \frac{1-\bar{\lambda}_i r^2\lambda}{\lambda-\lambda_i}\right| &= \sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\frac{1}{B(\lambda)}\right| \cdot \prod_i\left|1+\frac{\bar{\lambda}_i\lambda(1-r^2)}{1-\bar{\lambda}_i\lambda}\right| \\ &\le \sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\frac{1}{B(\lambda)}\right| \cdot \left(1+\frac{1-r^2}{1-\mu(1+1/n)}\right)^{|m|}.\end{aligned}$$

We can choose $1-r^2=\frac{1-\mu(1+1/n)}{|m|}$ and get

$$\|z^n\|_{W/mW}\leq \frac{2e^2\mu^{n+1}\sqrt{|m|}(|m|+1)}{n(1-(1+1/n)\mu)^{3/2}}\sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\frac{1}{B(\lambda)}\right|,$$

where we used the fact that $(1+1/|m|)^{|m|}\leq e$ and that, by the Bernoulli inequality for $|m|>1$, $r^{|m|}\geq(1-\frac{1-\mu(1+\frac{1}{n})}{2})\geq 1/2$.
We now specialize the above derivation to the case when $\mathcal{A}=\mathcal{T}-\mathcal{T}_\infty$. By assumption it holds for any n and $\mathcal{T}\in\mathfrak{T}$ that $\|\mathcal{T}^n\|\leq C$ and it follows that

$$\|(\mathcal{T}-\mathcal{T}_\infty)^n\|=\|\mathcal{T}^n-\mathcal{T}_\infty^n\|\leq\|\mathcal{T}^n\|+\|\mathcal{T}_\infty^n\|\leq 2C.$$

In total we can assert that

$$\|\mathcal{T}^n-\mathcal{T}_\infty^n\|=\|(\mathcal{T}-\mathcal{T}_\infty)^n\|\leq\frac{4Ce^2|m_{\mathcal{T}-\mathcal{T}_\infty}|^{1/2}(|m_{\mathcal{T}-\mathcal{T}_\infty}|+1)\cdot\mu^{n+1}}{n\left(1-(1+\frac{1h}{n})\mu\right)^{3/2}}\sup_{\substack{|\lambda|=\\(1+1/n)\mu}}\left|\frac{1}{B(\lambda)}\right|$$

completing the proof of Theorem IV.3. □

C. Comparison to the Schur and Jordan convergence bounds

Theorem IV.3 significantly improves upon both the Jordan and the Schur bounds, Theorems III.3, III.2. In this subsection we shall illustrate this comparing the different convergence estimates for a semigroup of quantum channels. Since for all $\mathcal{T}\in\mathfrak{T}_+$ we have that $\|\mathcal{T}\|_\diamond=1$, Theorem IV.3 gives a bound for the diamond norm. With the notation of Theorem IV.3 we have that

$$\|\mathcal{T}^n-\mathcal{T}_\infty^n\|_\diamond\leq\frac{4e^2\sqrt{|m|}(|m|+1)\cdot\mu^{n+1}}{n\left(1-(1+\frac{1}{n})\mu\right)^{3/2}}\sup_{\substack{|z|=\\\mu(1+1/n)}}\left|\frac{1}{B(z)}\right| \tag{18}$$

$$\leq\frac{4e^2\sqrt{|m|}(|m|+1)\cdot\mu^{n}}{\left(1-(1+\frac{1}{n})\mu\right)^{3/2}}\prod_{m/(z-\lambda_D)}\frac{1-(1+\frac{1}{n})\mu|\lambda_i|}{\mu-|\lambda_i|+\frac{\mu}{n}}. \tag{19}$$

For the inverse Blaschke product in (19) we can establish lower and upper bounds. The function $\frac{1-(1+1/n)\mu x}{(1+1/n)\mu-x}$ is monotonically increasing with $x\in[0,\mu]$ and we have that

$$\left(\frac{1}{(1+1/n)\mu}\right)^{|m|-1}\leq\prod_{m/(z-\lambda_D)}\frac{1-(1+\frac{1}{n})\mu|\lambda_i|}{\mu-|\lambda_i|+\frac{\mu}{n}}\leq\left(\frac{n}{\mu}(1-\mu^2)\right)^{|m|-1}. \tag{20}$$

In the following we compare Inequalities (18), (19) to the corresponding bounds resulting from the Jordan and Schur decompositions.

Comparison with the Jordan bound. To establish a convergence bound for quantum channels in diamond norm one can use Theorem III.2 together with the norm equivalence relations (3), (4). But as Theorem III.2 has a qualitative character only (i.e., it does not specify C_1, C_2), the constants coming from the norm equivalence are of no relevance. As expected, both Theorem III.2 and Inequality (19) include an exponential factor μ^n. Suppose that the largest Jordan block for λ_D has size d_μ and that there is no other eigenvalue of $\mathcal{T} - \mathcal{T}_\infty$ of magnitude μ. Then the minimal polynomial of $\mathcal{T} - \mathcal{T}_\infty$ contains a factor $(z - \lambda_D)^{d_\mu}$ and in (19) there are $d_\mu - 1$ factors for this eigenvalue. The denominator in Inequality (19) leads to a factor $(n/\mu)^{d_\mu - 1}$ in this estimate, which captures the same qualitative n-dependence as the upper bound of Theorem III.2. Due to the lower bound in Theorem III.2 the factor $(n/\mu)^{d_\mu - 1}$ is also necessary. But as compared to Theorem III.2 Inequality (19) bears the obvious advantage that it specifies C_2. On the other hand if there are several distinct eigenvalues of magnitude μ, Inequality (19) does not yield the correct asymptotic behavior from Theorem III.2, since *any* eigenvalue of magnitude μ occurring in m contributes a factor n/μ. The reason for this lies in the estimate (18), i.e., in bounding each Blaschke factor individually, which leads to Corollary IV.4. Roughly speaking, if there are distinct eigenvalues of magnitude μ then, for sufficiently large n, any z of magnitude $\mu(1 + 1/n)$ can be close at most to one of those eigenvalues. It is not difficult to make this intuition precise and prove the upper bound of Theorem III.2 based on Theorem IV.3 with the additional advantage of specifying C_2. Finally we note that the occurrence of the correct asymptotic n-dependence in Theorem IV.3 is linked to the integration by parts in (17) and our application of Spijker's Lemma. This procedure yields the $1/n$ prefactor in Theorem IV.3, which is canceled by one inverse Blaschke factor in Corollary IV.4. Had we bounded (16) directly by the supremum of the integrand on the circle, we would have obtained an estimate where one factor in the Blaschke product is proportional to n/μ even in case of only one eigenvalue of magnitude μ.

Comparison with the Schur bound. Taking into account the norm equivalence relations (3), (4) the Schur bound entails

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\diamond \leq 2D^{3/4}(\mu + 2D^{1/4})^{D-1} n^{D-1} \mu^{n-D+1}.$$

If one assumes that λ_D is D-fold degenerate with maximal Jordan block this results in a factor $(n/\mu)^{D-1}$ in Inequality (20). Hence, even in the case of the worst Jordan structure for $\mathcal{T} - \mathcal{T}_\infty$, Theorem IV.3 improves upon bounds obtained from Theorem III.3 *exponentially* in the D-dependend prefactor.

Finally, we discuss some implications of the lower bound in (20). We use that bound to estimate how good the upper bound of Corollary IV.4 can possibly be. Note that the left hand side of Inequality (20) contains a factor $(1/\mu)^{|m|-1}$. If all eigenvalues of $\mathcal{T}$ are distinct this factor grows with the dimension of the system. That is, for "generic'' $\mathcal{T}$ it needs D time steps until Corollary IV.4 can yield a nontrivial statement. This is unfortunate from the point of view of applications, where one is looking for estimates such that $poly(\log(D))$ steps are sufficient. It is natural to ask whether or not Theorem IV.3

is optimal and whether one might be able to dispense of the $(1/\mu)^{|m|-1}$ prefactor. The following subsection discusses aspects related to the optimality of Theorem IV.3. Even full information about the spectrum (alone) is never sufficient to prove $poly(\log D)$ convergence. To overcome this issue one may use properties of the semigroup beyond its spectrum. One important class of semigroups for which fast convergence can be proved under additional assumptions are detailed balanced semigroups (Definition V.1). We discuss the convergence of such semigroups in detail in Section V.

D. Semigroups of Hilbert space contractions

In this subsection we discuss semigroups of Hilbert space contractions. More precisely, suppose we are given a semigroup $(\mathcal{T}^n)_{n\geq 0}$ of linear operators acting on a finite-dimensional Hilbert space such that $\|\mathcal{T}\|_\infty \leq 1$. As before, our major interest lies in bounding the quantity $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty$ in terms of the spectrum of $\mathcal{T}$. Clearly, this setup is less general than our main setup in Section IV B and one can expect better bounds. In what follows we derive an analog of Theorem IV.3 for contractive semigroups and discuss the optimality of the obtained bounds.

Let us adopt the notation from Theorem IV.3. As before we write $\sigma(\mathcal{T} - \mathcal{T}_\infty)$ for the spectrum and $m = m_{\mathcal{T}-\mathcal{T}_\infty}$ for the minimal polynomial of $\mathcal{T} - \mathcal{T}_\infty$. $B(z) = \prod_i \frac{z-\lambda_i}{1-\lambda_i z}$ denotes the Blaschke product associated with m. To avoid cumbersome notation we shall again assume that m has simple zeros. The extension to the more general case does not result in any difficulties. Before we proceed with our main discussion we briefly introduce some notation and standard concepts from spectral operator theory. We define the $|m|$-dimensional model space

$$K_B := H^2 \ominus BH^2 := H^2 \cap (BH^2)^\perp,$$

where we employ the usual scalar product from the Hilbert space H^2. The model operator M_B acts on K_B as

$$\begin{aligned} M_B :\ & K_B \to K_B \\ & f \mapsto M_B(f) = P_B(zf), \end{aligned}$$

where P_B denotes the orthogonal projection on K_B. In other words, M_B is the compression of the multiplication operation by z to the model space K_B (see [13] for a detailed discussion of model operators and spaces). As multiplication by z has operator norm 1 it is clear that M_B is a Hilbert space contraction. More precisely, for any $\phi \in H^\infty$ the norm of $\phi(M_B)$ can be evaluated using Sarason's lifting Theorem [15, 20] as

$$\|\phi(M_B)\|_\infty = \|\phi\|_{H^\infty/mH^\infty}. \tag{21}$$

We can also write $\|\phi(M_B)\|_\infty$ as variational expression in the Hardy space H^1. From [5] we get that

$$\|\phi(M_B)\|_\infty = \sup_{\substack{F\in H^1 \\ \|F\|_1\leq 1}} \left| \frac{1}{2\pi i} \int_{|z|=1} \frac{\phi}{B} F \,\mathrm{d}z \right|. \tag{22}$$

Note that this trivially implies

$$\left|\frac{1}{2\pi i}\int_{|z|=1}\frac{\phi}{B}\,\mathrm{d}z\right| \leq \|\phi(M_B)\|_\infty \leq \sup_{|z|=1}\left|\frac{\phi}{B}\right|.$$

It can be shown that the spectrum of the model operator M_B defined above is given by the zeros of the corresponding Blaschke product B. In our case this means that $\mathcal{T}-\mathcal{T}_\infty$ and M_B have identical spectrum. Hence, to any $\mathcal{T}$ we can associate a (completely non-unitary [11]) contraction M_B having spectrum $\sigma = \sigma(\mathcal{T}-\mathcal{T}_\infty)$.

Let us proceed by studying convergence estimates for the contractive semigroup of the form of Inequality (1). To start with, we prove that if $\|\mathcal{T}\|_\infty \leq 1$ then $\|\mathcal{T}-\mathcal{T}_\infty\|_\infty \leq 1$, i.e. the semigroup $\{(\mathcal{T}-\mathcal{T}_\infty)^n\}_{n\geq 0}$ is contractive, too.

Proposition IV.5. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a contractive semigroup on a Hilbert space and let $\mathcal{T}_\infty$ be as in Equation* (8)*. Then*
(i) the semigroup $\{(\mathcal{T}-\mathcal{T}_\infty)^n\}_{n\geq 0}$ is contractive, and
(ii) if $\mathcal{T}^*(e) = \lambda e$ with $|\lambda| = 1$, then $\mathcal{T}(e) = \bar{\lambda}e$.

Proof. Both follows from the fact that any contraction on a Hilbert space admits a unique decomposition into an orthogonal direct sum of a unitary and a completely non-unitary operation ([11], Theorem 3.2). In our case, $\mathcal{T}_\infty$ corresponds exactly to the unitary part of $\mathcal{T}$ and $\mathcal{T}-\mathcal{T}_\infty$ is a (completely non-unitary) contraction, hence (i). (ii) is then a consequence of the normality of the unitary part. □

The second part of Proposition IV.5 generalizes the fact that for classical as well as for quantum Markov processes, contractivity implies that the transition map is doubly stochastic. In fact, in those cases the converse implication holds as well [17].

From the first part of Proposition IV.5 and by Inequality (12) it follows that

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty = \|(\mathcal{T}-\mathcal{T}_\infty)^n\|_\infty \leq \|z^n\|_{H^\infty}.$$

Our previous considerations from Section IV A, Lemma IV.1 furthermore imply

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq \|z^n\|_{H^\infty/mH^\infty}. \tag{23}$$

We conclude from Equation (21) that in order to upper bound (23) it is sufficient to consider $\|M_B^n\|_\infty$. This is in contrast to our discussion of bounded semigroups on Banach spaces, where we had to rely on the Cauchy-Schwarz Inequality (15). In addition, we note that $\|\phi\|_{H^\infty/mH^\infty} = \|\phi(M_B)\|_\infty$ allows us to work with $\|\cdot\|_{H^\infty/mH^\infty}$ directly and we do not require an ad hoc function to upper bound (14).

In our study of bounded semigroups in Section IV C we have encountered a factor $(1/\mu)^{|m|-1}$ in (20) that grows exponentially with the dimension of the space on which the semigroup acts if all eigenvalues of the generator are distinct. The following proposition shows that, if in a bound of the type (1) K only depends on the eigenvalue structure of $\mathcal{T}$ and on n, then K must contain such a factor. We achieve this by showing that for any contractive semigroup with generator $\mathcal{T}$ there is a contractive semigroup whose generator has the same spectrum as $\mathcal{T}$ but which converges slowly if n is small.

Proposition IV.6. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a contractive semigroup acting on a D-dimensional Hilbert space and let $m = m_{\mathcal{T}-\mathcal{T}_\infty}$ denote the minimal polynomial of $\mathcal{T} - \mathcal{T}_\infty$ and B the corresponding Blaschke product. Then there is a contractive semigroup $(\mathcal{E}^n)_{n\geq 0}$ such that $\mathcal{E}$ has the* same minimal polynomial *as $\mathcal{T}$ and*

$$\|\mathcal{E}^n - \mathcal{E}_\infty^n\|_\infty = \sup_{\substack{F\in H^1\\ \|F\|_1\leq 1}} \left|\frac{1}{2\pi i}\int_{|z|=1} \frac{z^n}{B} F \,\mathrm{d}z\right|.$$

In particular, for all $n < |m| \leq D$ it holds that

$$\|\mathcal{E}^n - \mathcal{E}_\infty^n\|_\infty = 1.$$

The supremum in Proposition IV.6 is attained by a function $\tilde{F} = f^2$, where f is in the unit ball of K_B [5]. Hence, the optimization effectively goes over a finite-dimensional vector space of rational functions with fixed poles and bounded degree (see [5] for details). One can obtain simple lower bounds on the convergence speed of $(\mathcal{E}^n)_{n\geq 0}$ by choosing a certain $f \in K_B$ and evaluating the integral with the Residue theorem.

The second assertion of Proposition IV.6 states that for any spectrum we can construct a semigroup such that the distance of the evolution to its asymptotic behavior stays maximal for at least $|m|-1$ time steps. Clearly this implies that one cannot prove that $poly(\log|m|)$ time steps bring the semigroup close to its stationary behavior if only spectral data is given.

Note that if in a bound of the form $\|\mathcal{T}^n - \mathcal{T}_\infty^n\| \leq K\mu^n$, with a bounded semigroup $(\mathcal{T}^n)_{n\geq 0}$, K only depends on the spectrum of $\mathcal{T}$ then by Proposition IV.6 we have $1 \leq K\mu^{|m|-1}$. That is, in this case we obtain the lower bound $K \geq (1/\mu)^{|m|-1}$.

Proof of Proposition IV.6. The first assertion is clear by choosing "$\mathcal{E} := \mathcal{T}_\infty \oplus M_B$" such that $\mathcal{E}_\infty = \mathcal{T}_\infty$ (on the unitary subspace) and $\mathcal{E} - \mathcal{E}_\infty = 0 \oplus M_B$. For the second we consider the extremal problem Equation (22). Let ψ be any rational function with poles away from the unit circle $|z| = 1$. Corollary 5 in [5] asserts that, we have

$$\sup_{\substack{F\in H^1\\ \|F\|_1\leq 1}} \left|\frac{1}{2\pi i}\int_{|z|=1} \psi F \,\mathrm{d}z\right| = \sup_{|z|=1} |\psi(z)|$$

if and only if ψ is a constant multiple of the quotient of two finite Blaschke products B_1, B_2 having no common zeros and such that the degree of B_1 is strictly smaller than the degree of B_2 ($|B_1| < |B_2|$), i.e., $\psi = c\frac{B_1}{B_2}$ for some $c \in \mathbb{C}$. Let B denote the Blaschke product associated with m, it follows readily that

$$\|M_B^n\|_\infty = \sup_{\substack{F\in H^1\\ \|F\|_1\leq 1}} \left|\frac{1}{2\pi i}\int_{|z|=1} \frac{z^n}{B} F \,\mathrm{d}z\right| = 1$$

holds for $n < |m|$. □

To gain a better understanding of weather the derivation of Theorem IV.3 is optimal, i.e. whether or not the obtained estimate is sharp, let us prove an analog of Theorem IV.3 for semigroups of Hilbert space contractions. The derivation is based on techniques similar to those that led to Theorem IV.3, but in the case at hand we can take a more direct approach based on the theory of model operators.

Proposition IV.7. *Let $(\mathcal{T}^n)_{n\geq 0}$ be a contractive semigroup on a D-dimensional Hilbert space and let $\mathcal{T}_\infty$ be the operator introduced in* (8) *(i.e., the unitary part of $\mathcal{T}$). We write $m = m_{\mathcal{T}-\mathcal{T}_\infty}$ for the minimal polynomial and μ for the spectral radius of $\mathcal{T}-\mathcal{T}_\infty$. B denotes the Blaschke product associated with m. Then, for $n > \frac{\mu}{1-\mu}$ we have*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq \mu^{n+1} \frac{2|m|e}{n(1-(1+1/n)^2\mu^2)} \sup_{|z|=\mu(1+1/n)} \left|\frac{1}{B(z)}\right|.$$

As before, we can bound all terms in the Blaschke product individually (see Appendix A) and find (compare Corollary IV.4)

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq \mu^n \frac{2|m|e}{1-(1+1/n)^2\mu^2} \prod_{i\neq|m|} \frac{1-|\lambda_i|\mu(1+1/n)}{\mu(1+1/n)-\lambda_i}.$$

Proof of Proposition IV.7. The derivation proceeds along the lines of Theorem IV.3. We use an H^∞ functional calculus to bound $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty$ in terms of $\|z^n\|_{H^\infty/mH^\infty}$. The latter expression can be rewritten using a contour integral similar to Equation (16), integrate by parts, and finally apply Spijker's Lemma. We have already mentioned that

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq \|z^n\|_{H^\infty/mH^\infty} = \|M_B^n\|_\infty = \sup_{\substack{F\in H^1\\ \|F\|_1\leq 1}} \left|\frac{1}{2\pi i}\int_{|z|=1} \frac{z^n}{B} F \,\mathrm{d}z\right|$$

and that the supremum in this extremal problem is attained by some function $\tilde{F} = f^2$ with $f \in K_B$ [5]. Thus, $\tilde{F}/B$ is a rational function with $2|m|$ poles located at $(\xi_1, ..., \xi_{|m|}, \bar{\xi}_1^{-1}, ..., \bar{\xi}_{|m|}^{-1})$, where ξ_i are the zeros of m. In the above integral we can change the contour of integration and integrate along the circle $\gamma : \phi \mapsto \mu(1+1/n)e^{i\phi}$. Integrating by parts and and applying Spijker's Lemma [24] we obtain

$$\begin{aligned}
\left|\frac{1}{2\pi i}\int_\gamma \frac{z^n}{B}\tilde{F}\,\mathrm{d}z\right| &= \frac{1}{2\pi(n+1)}\left|\int_\gamma z^{n+1}\left(\frac{\tilde{F}}{B}\right)'\mathrm{d}z\right| \\
&\leq \frac{\mu^{n+1}(1+1/n)^{n+1}}{2\pi(n+1)}\int_\gamma \left|\left(\frac{\tilde{F}}{B}\right)'\right||\mathrm{d}z| \\
&\leq \frac{2|m|\mu^{n+1}(1+1/n)^{n+1}}{n+1} \sup_{|z|=\mu(1+1/n)}\left|\frac{\tilde{F}}{B}\right|.
\end{aligned}$$

It is known that for $F \in H^1$ and $z \in \mathbb{D}$ one can bound $|F(z)| \leq \frac{1}{1-|z|^2} \|F\|_{H^1}$ [8] and with $(1+1/n)^n \leq e$ we finally obtain

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \leq \frac{2|m|e\mu^{n+1}}{n(1-(1+1/n)^2\mu^2)} \sup_{|z|=\mu(1+1/n)} \left|\frac{1}{B}\right|.$$

□

E. Slow convergence for Markov chains

Proposition IV.6 provides an example of a slowly converging contractive semigroup with arbitrarly given spectrum. One might wonder in how far the phenomenon extends to the Markov chain setup. When $\mathcal{T}$ is the transition map of a classical or quantum Markov chain, is it possible to prove (1) where, K should only depend on the spectrum of $\mathcal{T}$ and n but such that the stationary behavior sets in after $poly(\log(D))$ time steps? The following example shows that this can not be the case.

We construct a classical stochastic $D \times D$ matrix T with real positive spectrum such that $\|T^n - T_\infty\|_{1\to 1} = 2$ for $n \leq D-2$. Let, as always, μ denote the spectral radius of $T - T_\infty$. We write $\{e_i\}_{i=1,\ldots,D}$ for the canonical column vectors, i.e., $(e_i)_j = \delta_{ij}$ and for $\lambda_i \in [0,1)$, $1 \leq i \leq D-1$, we define

$$T := \begin{pmatrix} \lambda_1 & & & & & \\ 1-\lambda_1 & \lambda_2 & & & & \\ & 1-\lambda_2 & \lambda_3 & & & \\ & & \ddots & \ddots & & \\ & & & & \lambda_{D-1} & \\ & & & & 1-\lambda_{D-1} & 1 \end{pmatrix}.$$

T is a stochastic matrix with spectrum $\sigma(T) = \{\lambda_1, \ldots, \lambda_{D-1}, 1\}$. Since $\lambda_i < 1$ for large n the image of T^n converges to an one-dimensional subspace corresponding to the eigenvalue 1. We have that $T_\infty = \lim_{n\to\infty} T^n$ and observe that $T_\infty e_1 = e_D$. It is not difficult to see that for $n \leq D-2$ the D-th entry of the vector $T^n e_1$ is always zero, $\langle e_D | T^n e_1 \rangle = 0$. It follows that $\|(T^n - T_\infty)e_1\|_1 = 2$ (where $\|\cdot\|_1$ denotes the 1-norm, Section II B) and we conclude that $\|T^n - T_\infty\|_{1\to 1} = 2$ for $n \leq D-2$. As before, if K only depends on the spectrum of $\mathcal{T}$ this implies that $K \geq (1/\mu)^{D-2}$. Note that the above reasoning does not depend on the exact values of the eigenvalues (as long as they are non-negative). This suggests that generally the spectrum $\sigma(T)$ does not contain sufficient information to prove $poly(\log D)$ fast convergence estimates. Since every classical stochastic matrix can be embedded into a quantum channel, the lower bound on K is also true for quantum channels.

V. CONVERGENCE BOUNDS FROM DETAILED BALANCE

Applications often rely on fast convergence in the sense that $poly(\log D)$ steps should suffice for the asymptotic behavior to set in. In our previous discussion we have argued

that such bounds cannot rely on spectral data alone. To obtain better convergence estimates one requires additional knowledge about the semigroup. In this section we will derive convergence estimates for a general bounded semigroup under the condition that its generator be related to a Hermitian map in a certain way – for classical and quantum Markov processes this will correspond to the well-known *detailed balance condition* (see, e.g., [1, 7, 26]). Throughout this section we require the state space $\mathcal{V}$ to be equipped with a scalar product $\langle\cdot|\cdot\rangle$, which induces norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ on $\mathcal{V}$ and $\mathcal{L}(\mathcal{V})$, respectively, and for convenience we will sometimes assume an orthonormal basis in $\mathcal{V}$ to be fixed (cf. Subsection II A).

A. General bound

We start with a generalization of the well-known detailed balance condition for classical Markov chains. This allows us to employ the corresponding property in the context of bounded semigroups.

Definition V.1 (Detailed balance for linear maps)**.** *Let a linear map $\mathcal{T} \in \mathcal{L}(\mathcal{V})$ be given. If $\mathcal{B} \in \mathcal{L}(\mathcal{V})$ is positive-definite (i.e., $\langle v|\mathcal{B}(v)\rangle > 0\ \forall v \in \mathcal{V} \setminus \{0\}$) and satisfies $\mathcal{T}\mathcal{B} = \mathcal{B}\mathcal{T}^*$, then we say that $\mathcal{T}$ satisfies the detailed balanced condition(with respect to $\mathcal{B}$).*

This definition is equivalent to saying that $\mathcal{T}$ is Hermitian with respect to *some* scalar product on the space $\mathcal{V}$, namely the scalar product $\langle\cdot|\mathcal{B}^{-1}(\cdot)\rangle$, but we choose the formulation with *given* scalar product $\langle\cdot|\cdot\rangle$ (independent of $\mathcal{T}$) and explicit use of $\mathcal{B}$. Note further that, due to strict positive-definiteness, $\mathcal{B}$ in the above definition is in particular Hermitian and invertible. In conventional formulations of the detailed balance condition the map $\mathcal{B}$ is not required to be strictly positive-definite, but we do so here as the derived bounds become trivial otherwise (see below).

The detailed balance condition for a linear map $\mathcal{T}$ gives

$$\mathcal{B}^{-1/2}\mathcal{T}\mathcal{B}^{1/2} \;=\; \mathcal{B}^{1/2}\mathcal{T}^*\mathcal{B}^{-1/2}\ ,$$

which means that $\mathcal{B}^{-1/2}\mathcal{T}\mathcal{B}^{1/2}$ is Hermitian, therefore has only real eigenvalues $\lambda_i \in \mathbb{R}$ $(i = 1, \ldots, D)$, and is unitarily diagonalizable:

$$\mathcal{U}^*\mathcal{B}^{-1/2}\mathcal{T}\mathcal{B}^{1/2}\mathcal{U} \;=\; \Lambda \;=\; \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{pmatrix}\ .$$

This equation implies that $\mathcal{T}$ is diagonalized by the similarity transformation $S := \mathcal{B}^{1/2}\mathcal{U}$ (i.e. $S^{-1}\mathcal{T}S = \Lambda$). Note that $\mathcal{T}$ has spectrum $\{\lambda_i\}_i$, too.

If $\mathcal{T}$ is now power-bounded, i.e., the generator of a bounded semigroup, the definition in Equation (8) implies that $\mathcal{T}_\infty$ is diagonalized by S as well,

$$\mathcal{U}^*\mathcal{B}^{-1/2}\mathcal{T}_\infty\mathcal{B}^{1/2}\mathcal{U} \;=\; \Lambda_\infty,$$

where Λ_∞ is obtained from Λ by deleting all entries of magnitude smaller than 1. $\Lambda - \Lambda_\infty$ is thus diagonal with operator norm $\mu < 1$, where μ is the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. We thus arrive at the following convergence estimate:

$$\begin{aligned}\|(\mathcal{T} - \mathcal{T}_\infty)^n\|_\infty &= \left\|\mathcal{B}^{1/2}\mathcal{U}(\Lambda - \Lambda_\infty)^n\mathcal{U}^*\mathcal{B}^{-1/2}\right\|_\infty \\ &\leq \left\|\mathcal{B}^{1/2}\right\|_\infty \|\mathcal{U}\|_\infty \|(\Lambda - \Lambda_\infty)^n\|_\infty \|\mathcal{U}^*\|_\infty \left\|\mathcal{B}^{-1/2}\right\|_\infty \\ &= \mu^n \left\|\mathcal{B}^{1/2}\right\|_\infty \left\|\mathcal{B}^{-1/2}\right\|_\infty \ .\end{aligned}$$

(The latter two factors may be recognized as the *condition number* of $\mathcal{B}^{1/2}$.) We formulate this as a theorem:

Theorem V.1. *Let $\mathcal{V}$ be a (real or complex) vector space with scalar product, and $\mathcal{T} \in \mathcal{L}(\mathcal{V})$ be the generator of a bounded semigroup $(\mathcal{T}^n)_{n\geq 0}$, which satisfies detailed balanced w.r.t. a positive-definite $\mathcal{B} \in \mathcal{L}(\mathcal{V})$. Denote by μ the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. Then, for any $n \in \mathbb{N}$,*

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\infty \ \leq \ \mu^n \left\|\mathcal{B}^{1/2}\right\|_\infty \left\|\mathcal{B}^{-1/2}\right\|_\infty \ ,$$

where $\|\cdot\|_\infty$ denotes the operator norm on $\mathcal{L}(\mathcal{V})$.

We now discuss detailed balance more specifically for classical and quantum Markov chains. First observe that, if $e \in \mathcal{V}$ is a fixed point of $\mathcal{T}^*$, i.e. $\mathcal{T}^*(e) = e$, then $\pi := \mathcal{B}(e)$ satisfies

$$\mathcal{T}(\pi) = \mathcal{T}\mathcal{B}(e) = \mathcal{B}\mathcal{T}^*(e) = \mathcal{B}(e) = \pi \ ,$$

i.e., π is fixed by the semigroup generator $\mathcal{T}$. Conversely, if π is a fixed point of $\mathcal{T}$, then $e := \mathcal{B}^{-1}(\pi)$ is left invariant by $\mathcal{T}^*$. For a classical Markov chain the generator satisfies $T^*(e) = e$ with $e = \sum_{i=1}^D e_i = (1, ..., 1)$ and $\mathcal{T}^*(\mathbb{1}) = \mathbb{1}$ holds for generators of quantum Markov chains (see Section II B). Thus, for classical and quantum Markov chains the detailed balance condition immediately yields a fixed point of the transition map.

In the theory of classical Markov chains, a stochastic matrix $T \in \mathbb{R}^{d\times d}$ is usually defined to be detailed balanced w.r.t. the probability distribution $\pi \in \mathbb{R}^d$ (i.e. $\pi_i \geq 0$ and $\sum_i \pi_i = 1$), if $T_{ji}\pi_i = T_{ij}\pi_j$ holds for all i, j (see, e.g., [1, 7]). Defining a diagonal matrix B with entries $B_{ii} := \pi_i$, the latter condition can be written as $TB = BT^*$. If furthermore the fixed-point probability distribution π has full support (i.e. $\pi_i > 0 \ \forall i$), then T is detailed balanced w.r.t. B in the sense of our Definition V.1. ($\pi = Be$ will necessarily be a fixed point of T.) Due to normalization it holds that $\min_i \pi_i \leq 1/d$. Using this and the norm equivalence (2), Theorem V.1 yields the following well-known convergence estimate [1, 7] for the special case of a classical Markov chain that satisfies detailed balance w.r.t. the distribution π:

$$\|T^n - T_\infty^n\|_{1-1} \ \leq \ \mu^n\sqrt{d}\sqrt{\frac{\max_i \pi_i}{\min_i \pi_i}} \ \leq \ \frac{\mu^n}{\min_i \pi_i} \ . \tag{24}$$

This estimate may become trivial if detailed balance is defined without the full-support condition on π as one may then have $\min_i \pi_i = 0$. On the other hand, if one has a positive lower bound on $\min_i \pi_i$, Equation (24) may become a useful convergence estimate. This technique is frequently used for detailed balanced chains that have a (unique) full-rank probability distribution as fixed point, and where one can find a "good" lower bound on $\min_i \pi_i$ [1, 2, 7]. Often the situation arises that the chain converges to a Gibbs state $\pi_i = e^{-\beta H_i}/Z$ at finite inverse temperature $\beta \in [0, \infty)$ with $Z := \sum_i e^{-\beta H_i}$. An important class of Markov chains that obey the detailed balance condition are Metropolis Hastings Markov Chains [7].

There are different generalizations of the detailed balance condition to quantum Markov chains [26], which we, however, all capture by Definition V.1. Let us specialize to the quantum detailed balance condition that most immediately generalizes the classical condition from the previous paragraph to the non-commutative case in a symmetric way and that has been employed for proving convergence of quantum Markov chains before (e.g. [27]). Namely, given a positive trace-preserving map $\mathcal{T} \in \mathfrak{T}$ acting on the set $\mathcal{M}_d$ of $d \times d$ matrices, we consider the detailed balance condition induced by the map $\mathcal{B}_\sigma(X) := \sqrt{\sigma} X \sqrt{\sigma}$, where $\sigma \in \mathcal{M}_d$ is a density matrix of full rank. Again, due to trace-preservation, it is easy to see that if $\mathcal{T}$ is detailed balanced w.r.t. $\mathcal{B}_\sigma$, then $\sigma = \mathcal{B}_\sigma(\mathbb{1})$ is a fixed point of $\mathcal{T}$. This leads to the following convergence result for the quantum case:

Corollary V.2. *Let $\mathcal{T} : \mathcal{M}_d \to \mathcal{M}_d$ be a positive trace-preserving map, and $\sigma \in \mathcal{M}_d$ be a full-rank density matrix such that*

$$\sqrt{\sigma}\mathcal{T}^*(X)\sqrt{\sigma} \;=\; \mathcal{T}(\sqrt{\sigma} X \sqrt{\sigma}) \quad \forall X \in \mathcal{M}_d \; .$$

Denote by μ the spectral radius of $\mathcal{T} - \mathcal{T}_\infty$. Then, for any $n \in \mathbb{N}$,

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_{1-1} \;\leq\; \mu^n \sqrt{d} \sqrt{\frac{\lambda_{max}(\sigma)}{\lambda_{min}(\sigma)}} \;\leq\; \frac{\mu^n}{\lambda_{min}(\sigma)} \; ,$$

where $\lambda_{min}(\sigma)$ and $\lambda_{max}(\sigma)$ denote the minimal and maximal eigenvalues of σ, respectively. If, in addition, $\sigma = e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$ is the Gibbs state at inverse temperature $\beta \in [0, \infty)$ of a bounded Hamiltonian $H \in \mathcal{M}_d$, then

$$\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_{1-1} \;\leq\; \mu^n \, d \, e^{2\beta \|H\|_\infty} \; .$$

Proof of Corollary V.2. The conditions on $\mathcal{T}$ imply that it is detailed balanced w.r.t. the map $\mathcal{B}_\sigma$ defined above. Computing $\left\|\mathcal{B}_\sigma^{1/2}\right\|_\infty = \sqrt{\lambda_{max}(\sigma)}$ and $\left\|\mathcal{B}_\sigma^{-1/2}\right\|_\infty = 1/\sqrt{\lambda_{min}(\sigma)}$ and considering the norm equivalence (4) and bounding $\lambda_{min}(\sigma) \leq 1/d$ and $\lambda_{max}(\sigma) \leq 1$, we get the first assertion from Theorem V.1. In case of a thermal state, the second assertion follows from

$$\lambda_{min}\left(\frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}\right) \;\geq\; \frac{e^{-\beta \|H\|_\infty}}{\mathrm{tr}(e^{\beta \|H\|_\infty \mathbb{1}_d})} \;=\; \frac{e^{-2\beta \|H\|_\infty}}{d} \; .$$

□

This Corollary provides a possible way for proving that a state preparation or algorithm is *efficient* in the sense of computational complexity [12]. More concretely, for each N, consider a system of N particles (spins), each with finite Hilbert space dimension $s < \infty$, and a Hamiltonian H_N on each system. In many physical situations the Hamiltonian will be bounded by some polynomial of the particle number, $\|H_N\|_\infty \leq c_H N^k$; this occurs for example if $H_N = \sum_i H_{N,i}$ is a sum of k-local terms that are uniformly bounded by c_H. Assume further that the thermal state $\sigma_N = e^{-\beta H_N}/\mathrm{tr}(e^{-\beta H_N})$ at inverse temperature $\beta \in [0, \infty)$ is a fixed point of the positive trace-preserving map $\mathcal{T}_N$, and that $\mathcal{T}_N$ satisfies detailed balanced w.r.t. $\mathcal{B}_{\sigma_N}$. This assumption may be fulfilled, e.g., by Gibbs dynamics in a Markov Chain Monte Carlo Algorithm [27]. Lastly, assume that the *spectral gap* of $\mathcal{T}_N$ is asymptotically lower bounded by an inverse polynomial c_μ/N^α of N (where $c_\mu > 0$), i.e., the eigenvalue 1 corresponding to σ_N is the only eigenvalue of $\mathcal{T}_N$ with modulus 1 whereas $|\lambda_i| \leq 1 - c_\mu/N^\alpha$ for all other eigenvalues. Among these assumptions, when they apply, the latter one is usually the hardest one to prove in a given situation.

Under these presuppositions, the evolution operator $\mathcal{T}_N$ prepares the final state σ_N efficiently in the system size N. More precisely, for any initial state ρ_N of the system, the time-evolved state $\mathcal{T}_N^n(\rho_N)$ after n steps will be ε-close in trace-norm to the thermal state σ_N (i.e. $\|\mathcal{T}_N^n(\rho_N) - \sigma_N\|_1 \leq \varepsilon$) if

$$n \;\geq\; \frac{N^\alpha}{c_\mu}\left(2\beta c_H N^k + N \log s + \log\frac{1}{\varepsilon}\right) \;. \tag{25}$$

This means that the runtime to ε-convergence scales at most polynomially in the particle number N and polylogarithmically in the desired accuracy $\varepsilon \in (0, 1]$, which proves efficient state preparation.

For a proof of the runtime bound Inequality (25), note that the dimension of the N-partite system is $d_N = s^N$ and that, due to the spectral gap condition, $(\mathcal{T}_N)_\infty(\rho_N) = \sigma_N$ for any state ρ_N, which implies $\|\mathcal{T}_N^n(\rho_N) - \sigma_N\|_1 \leq \|\mathcal{T}_N^n - (\mathcal{T}_N)_\infty^n\|_{1-1}$. Finally, we use

$$\mu \;\leq\; 1 - \frac{c_\mu}{N^\alpha} \;\leq\; e^{-c_\mu/N^\alpha}$$

in the Gibbs state bound from Corollary V.2 and requiring the latter to be at most ε shows that the condition in Inequality (25) is sufficient for ε-convergence.

If one wants to bound the diamond norm $\|\mathcal{T}^n - \mathcal{T}_\infty^n\|_\diamond$ between the actual and the asymptotic evolution in Corollary V.2 instead of the trace-norm, then by Inequality (3) one incurs another factor d (or $1/\lambda_{min}(\sigma)$) in the upper bounds. This however does not affect the efficiency statement just obtained, as the asymptotic dynamics $(\mathcal{T}_N)_\infty$ is still reached, up to ε, in polynomial time.

B. An ℓ^2 bound

In this Subsection, again based on the detailed balance condition, we derive a sharper convergence bound than in Subsection V A, taking into account all eigenvalues and eigenvectors of the transition map $\mathcal{T}$. The special case of this bound for classical Markov processes has been used to prove so-called cutoff dynamics [1, 2, 7]. After describing the

approach for general bounded semigroups obeying detailed balance, we will specialize to quantum Markov chains.

Recall from above that, if $\mathcal{T} \in \mathcal{L}(\mathcal{V})$ is detailed balanced w.r.t. $\mathcal{B}$, its eigenvalues λ_i are real. Furthermore, as $\mathcal{R} := \mathcal{B}^{-1/2}\mathcal{T}\mathcal{B}^{1/2}$ is a Hermitian operator, it has a complete orthonormal eigenbasis $\{x_i\}_i$, i.e. $\mathcal{R}(x_i) = \lambda_i x_i$. From this we can define an eigensystem of the adjoint $\mathcal{T}^*$, which will play a prominent role in the bound:

$$y_i \;:=\; \mathcal{B}^{-1/2}(x_i)\,, \quad \text{which implies } \mathcal{T}^*(y_i) \;=\; \lambda_i y_i\,.$$

$\{y_i\}_i$ could alternatively be chosen as any eigensystem of $\mathcal{T}^*$ that is orthonormal w.r.t. the weighted scalar product $\langle\cdot|\mathcal{B}(\cdot)\rangle$.

The spectral decomposition $\mathcal{R}(v) = \sum_i \lambda_i \langle x_i|v\rangle x_i$ now gives:

$$\begin{aligned}
\mathcal{B}^{-1/2}\mathcal{T}^n(v) =\; & \mathcal{R}^n\mathcal{B}^{-1/2}(v) \\
=\; & \sum_{i=1}^{D} \lambda_i^n \langle x_i|\mathcal{B}^{-1/2}(v)\rangle x_i \\
=\; & \sum_{i=1}^{D} \lambda_i^n \langle \mathcal{B}^{-1/2}(x_i)|v\rangle x_i \;=\; \sum_{i=1}^{D} \lambda_i^n \langle y_i|v\rangle x_i\;.
\end{aligned}$$

Recognizing that $\mathcal{B}^{1/2}(x_i)$ is the right-eigenvector of $\mathcal{T}$ corresponding to y_i, the terms with $|\lambda_i| = 1$ in the last expression (which we assume to be $i = r+1, \ldots, n$) correspond to the asymptotic evolution $\mathcal{T}_\infty^n$. We can thus write

$$\mathcal{B}^{-1/2}(\mathcal{T}^n - \mathcal{T}_\infty^n)(v) \;=\; \sum_{i=1}^{r} \lambda_i^n \langle y_i|v\rangle x_i\;,$$

which, together with the fact that $\{x_i\}$ is an orthonormal system, gives by squaring:

$$\|(\mathcal{T}^n - \mathcal{T}_\infty^n)(v)\|_{2,\mathcal{B}^{-1}}^2 := \; \langle (\mathcal{T}^n - \mathcal{T}_\infty^n)(v)|\mathcal{B}^{-1}\left(\mathcal{T}^n - \mathcal{T}_\infty^n\right)(v)\rangle = \; \sum_{i=1}^{r} \lambda_i^{2n} |\langle y_i|v\rangle|^2\;. \quad (26)$$

This equality relates the eigensystem corresponding to the eigenvalues with modulus smaller than 1 to the convergence in a suitably modified Hilbert norm. By itself this relation does not seem very useful, although one can derive Theorem V.1 from it by rescaling the modified scalar product back to the originally given one.

When specializing to the quantum case, however, we can make a connection to the induced trace-norm, and thereby strengthen Corollary V.2:

Proposition V.3. *Let $\mathcal{T} : \mathcal{M}_d \to \mathcal{M}_d$ be a positive trace-preserving map, and $\sigma \in \mathcal{M}_d$ be a full-rank density matrix (i.e. $\mathrm{tr}(\sigma) = 1$, $\sigma > 0$) such that the detailed balance condition*

$$\sqrt{\sigma}\mathcal{T}^*(X)\sqrt{\sigma} \;=\; \mathcal{T}(\sqrt{\sigma}X\sqrt{\sigma}) \quad \forall X \in \mathcal{M}_d$$

holds. Let $\{\lambda_i\}_{i=1}^r$ be the part of the spectrum of $\mathcal{T}$ in the open interval $(-1, 1)$, and Y_i be the corresponding eigenvectors of the adjoint map $\mathcal{T}^*$, orthonormal in the sense that $\mathrm{tr}(Y_i^*\sigma^{1/2}Y_j\sigma^{1/2}) = \delta_{ij}$. Then, for every $Z \in \mathcal{M}_d$ (e.g. a quantum state):

$$\|(\mathcal{T}^n - \mathcal{T}_\infty^n)(Z)\|_1^2 \;\leq\; \sum_{i=1}^{r} |\mathrm{tr}(Y_i^* Z)|^2\, \lambda_i^{2n}\;. \quad (27)$$

Proof. One can apply the preceding general steps to the map $\mathcal{B}_\sigma(X) := \sqrt{\sigma}X\sqrt{\sigma}$ and the inner product $\langle Y|X\rangle := \mathrm{tr}(Y^*X)$. Then it remains to show that, for $A := (\mathcal{T}^n - \mathcal{T}_\infty^n)(Z)$,

$$\|A\|_1^2 \;\leq\; \langle A|\mathcal{B}_\sigma^{-1}(A)\rangle \;=\; \mathrm{tr}(A^*\sigma^{-1/2}A\sigma^{-1/2})\;.$$

To see this inequality holds in fact for all $A \in \mathcal{M}_d$, use the polar decomposition and let $U \in \mathcal{M}_d$ be a unitary such that UA is positive-semidefinite. Then cyclicity of the trace and two applications of the Cauchy-Schwarz inequality give:

$$\begin{aligned}
\|A\|_1^2 =\;& |\mathrm{tr}\,[UA]|^2 \;=\; \left|\mathrm{tr}\left[(\sigma^{1/4}U\sigma^{1/4})(\sigma^{-1/4}A\sigma^{-1/4})\right]\right|^2 \\
\leq\;& \mathrm{tr}\left[\sigma^{1/4}U\sigma^{1/2}U^*\sigma^{1/4}\right]\,\mathrm{tr}\left[\sigma^{-1/4}A^*\sigma^{-1/2}A\sigma^{-1/4}\right] \\
=\;& \mathrm{tr}\left[U\sigma^{1/2}U^*\sigma^{1/2}\right]\,\mathrm{tr}\left[A^*\sigma^{-1/2}A\sigma^{-1/2}\right] \\
\leq\;& \sqrt{\mathrm{tr}\,[U\sigma U^*]\,\mathrm{tr}\left[\sigma^{1/2}UU^*\sigma^{1/2}\right]}\,\mathrm{tr}\left[A^*\sigma^{-1/2}A\sigma^{-1/2}\right] \\
=\;& \mathrm{tr}(\sigma)\,\mathrm{tr}\left[A^*\sigma^{-1/2}A\sigma^{-1/2}\right] \\
=\;& \mathrm{tr}\left[A^*\sigma^{-1/2}A\sigma^{-1/2}\right]\;.
\end{aligned}$$

□

Detailed balance of a quantum map $\mathcal{T}$ w.r.t. certain other maps ($\mathcal{B} = \Omega_\sigma^k)^{-1}$ has been defined in [26] so that the family $(\Omega_\sigma^k)^{-1}$ includes the map $\mathcal{B}_\sigma$ from above. These detailed balance conditions also result in bounds that look essentially like Equation (27), except that in this more general case the Y_i should be orthonormal in the sense that $\mathrm{tr}(Y_i^*\mathcal{B}(Y_j)) = \delta_{ij}$. For a proof, note that Equation (26) holds generally, and the proof of Lemma 5 in [26] shows $\|(\mathcal{T}^n - \mathcal{T}_\infty^n)(Z)\|_1^2 \leq \|(\mathcal{T}^n - \mathcal{T}_\infty^n)(v)\|_{2,\mathcal{B}^{-1}}^2$ (the right-hand-side of the last inequality is a χ^2-divergence as considered in [26] only if $\mathcal{T}$ has merely one eigenvalue of modulus 1, however).

For classical detailed balanced Markov chains the analog of the convergence bound Inequality (27), which looks very similar in this setting [7], is often used for demonstrating the upper bound in cutoff results (cf. [1, 2] for an over overview and references). In this setting, most commonly the evolution $\mathcal{T}$ leads to a unique fixed point σ (often the maximally mixed state), so that the asymptotic evolution would simply be the "projection onto the fixed point", i.e. $\mathcal{T}_\infty^n(X) = \sigma\,\mathrm{tr}(X)$ for $n \geq 1$. Of course, for Proposition V.3 to be useful one also needs knowledge about the normalized eigenvectors Y_i.

Acknowledgments

We acknowledge financial support by the Elite Network of Bavaria (ENB) project QCCC, the CHIST-ERA/BMBF project CQC, the Marie-Curie project QUINTYL and

the Alfried Krupp von Bohlen und Halbach-Stiftung

---

[1] P. Diaconis. *Group representations in probability and statistics.* IMS Lecture Notes, Monograph Series, 1988.

[2] P. Diaconis. The cutoff phenomenon in finite markov chains. *Proc. Natl. Acad. Sci.*, 93:1659–1664, 1996.

[3] S. Diehl, A. Micheli, A. Kantian, B. Kraus, H. Büchler, and P. Zoller. Quantum states and phases in driven open quantum systems with cold atoms. *Nature Physics*, 4:878–883, 2008.

[4] D. Evans. Spectral properties of positive maps on c*-algebras. *J. of London Math. Soc.*, 17(2):345–355, 1978.

[5] S. R. Garcia and W. T. Ross. A nonlinear extremal problem on the Hardy space. *Computational methods and function theory*, 2009.

[6] B. Kraus, H. P. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller. Preparation of entangled states by quantum Markov processes. *Phys. Rev. A*, 78:042307, Oct 2008.

[7] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times.* Am. Math. Soc., 2009.

[8] A. D. Macintyre and W. W. Rogosinski. Extremum problems in the theory of analytic functions. *Acta Math.*, 82:275–325, 1950.

[9] A. Mitrophanv. Stability and exponential convergence of continuous-time markov chains. *J. Appl. Prob.*, 40:970–979, 2003.

[10] A. Mitrophanv. Sensitivity and convergence of uniformly ergodic markov chains. *J. Appl. Prob.*, 42:1003–1014, 2005.

[11] S. B. Nagy, C. Foias, H. Bercovici, and L. Kerchy. *Harmonic analysis of operators on Hilbert spaces.* Springer, 2010.

[12] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[13] N. Nikolski. *Treatise on the Shift Operator.* Springer: Grundlehren der mathematischen Wissenschaft, 1986.

[14] N. Nikolski. Condition numbers of large matrices and analytic capacities. *St. Petersburg Math. J.*, 17:641–682, 2006.

[15] N. Nikolski. *Operators, Functions and Systems: An Easy Reading.* AMS: Mathematical Surveys and Monographs: 93, 2009.

[16] V. I. Paulsen. *Completely bounded maps and operator algebras.* Cambridge University Press, 2002.

[17] D. Perez-Garcia, M. Wolf, D. Petz, and M. Ruskai. Contractivity of positive and trace preserving maps under lp norms. *J. Math. Phys.*, 47, 2006. ArXiv: 0601063v1.

[18] D. Reeb, M. J. Kastoryano, and M. M. Wolf. Hilbert's projective metric in quantum information theory. *Journal of Math. Phys.*, 52, 2011. ArXiv: 1102.5170.

[19] J. Rosenthal. Convergence rates of Markov chains. *SIAM*, 37(1995):387–405, 1995.

[20] D. Sarason. Generalized interpolation in H^∞ . *Trans. Amer. Math. Soc.*, 127:289–299, 1967.

[21] W. M. Schmidt. *Diophantine approximation.* Springer: Lect. Not. in Math., 1980.

[22] E. Seneta. *Non-negative matrices and Markov chains.* Springer Ser. in Stat., 2006.

[23] R. R. Smith. Completely bounded maps between C*-algebras. *J. London Math. Soc.*, 27:157–166, 1983.

[24] M. Spijker. On a conjecture by leveque and trefethen related to the kreiss matrix theorem. *BIT*, 31:551–555, 1991.

[25] O. Szehr and M. Wolf. Perturbation bounds for quantum Markov processes and their fixed

points . 2011. arXiv: 1210.1171.
[26] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete. The χ^2-divergence and mixing times of quantum Markov processes. *Journal of Math. Phys.*, 51:12, 2010.
[27] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete. Quantum Metropolis sampling. *Nature*, 471:87-90, 2011.
[28] F. Versraete, M. Wolf, and I. Cirac. Quantum computation and quantum-state engineering driven by dissipation. *Nat. Phys.*, 5 No.9:633–636, 2009.

Appendix A: An upper bound on a single Blaschke factor

For completeness we prove the following short lemma.

Lemma A.1. *Let* $|\lambda| < c \leq 1$ *then*

$$\sup_{|z|=c} \left| \frac{1-\bar{\lambda}z}{z-\lambda} \right| = \frac{1-|\lambda|c}{c-|\lambda|}.$$

Proof. We rewrite the absolute value on the left hand side using the fact that $|a|^2 = a\bar{a}$ for all $a \in \mathbb{C}$. This gives

$$\left| \frac{1-\bar{\lambda}z}{z-\lambda} \right|^2 = \frac{(1-|\lambda|c)^2 + 2|\lambda|c - 2\Re(\lambda\bar{z})}{(c-|\lambda|)^2 + 2|\lambda|c - 2\Re(\lambda\bar{z})}.$$

Note now that for $0 < \beta < \alpha$ and $0 \leq x$ we have

$$\frac{\alpha + x}{\beta + x} \leq \frac{\alpha}{\beta}.$$

Hence,

$$\frac{(1-|\lambda|c)^2 + 2|\lambda|c - 2\Re(\lambda\bar{z})}{(c-|\lambda|)^2 + 2|\lambda|c - 2\Re(\lambda\bar{z})} \leq \frac{(1-|\lambda|c)^2}{(c-|\lambda|)^2}.$$

Finally, we note that the supremum is attained for $z = \frac{c}{|\lambda|}\lambda$. □

Perturbation bounds for quantum Markov processes and their fixed points

O. Szehr and M. Wolf May 22, 2014

We investigate the stability of Markov processes with respect to perturbations of their transition maps. While this appears to be a well studied subject for classical Markov chains [1, 3, 2], it is essentially untouched for their quantum counterparts. Guided by the stability theory for classical Markov chains we prove inequalities of the form

$$\|\rho - \tilde{\rho}\| \leq \kappa \|\mathcal{T} - \tilde{\mathcal{T}}\|.$$

The above relates the distance between two stationary states ρ and $\tilde{\rho}$ that arise from two quantum channels $\mathcal{T}$ and $\tilde{\mathcal{T}}$ as fixed points $\rho = \mathcal{T}(\rho)$, $\tilde{\rho} = \tilde{\mathcal{T}}(\tilde{\rho})$. Commonly, $\mathcal{T}$ is considered the evolution of interest, while $\tilde{\mathcal{T}}$ is a small perturbation thereof. The *condition number* κ measure the sensitivity of the evolution to perturbation.

Quantum Markov processes model the behaviour of open quantum systems. It is important to study their sensitivity for at least the following two reasons. First, can there be a phase-transition in the steady states of a dissipative system due to perturbation of the generator of the evolution? Second, in the context of dissipative quantum computation the asymptotic state carries the outcome of the computation. From an engineering point of view it is inevitable to ask about the level of error in the steady states given that $\mathcal{T}$ cannot be realized ideally.

1 Sensitivity analysis

We will follow two approaches to sensitivity analysis. In the first we bound the condition number for asymptotic states using novel resolvent estimates, which provides a statement about the robustness of the asymptotic evolution [3].

Theorem 1. *Let $\mathcal{T}$ be a trace-preserving, positive linear map on $\mathcal{M}_d(\mathbb{C})$ and $\Lambda := \mathrm{spec}[\mathcal{T}]\backslash\{1\}$ the set of its non-unit eigenvalues. Then*

$$\frac{1}{\min_{\lambda\in\Lambda}|1-\lambda|} \leq \kappa \leq \frac{2(5\pi/3 + 2\sqrt{2})d^3}{\min_{\lambda\in\Lambda}|1-\lambda|}.$$

The second approach yields perturbation bounds at finite times, but we must assume in addition that the Markov processes has a unique stationary state [2].

Theorem 2. *Let $\rho_n := \mathcal{T}^n(\rho_0)$ and $\sigma_n := \tilde{\mathcal{T}}^n(\sigma_0)$ be the evolution of two density matrices with respect to two positive and trace-preserving linear maps $\mathcal{T}, \tilde{\mathcal{T}}$. If $\mathcal{T}$ has a unique stationary state and $\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1} \leq K\mu^n$ for $K \geq 0$, $\mu < 1$, then for $n > \hat{n} := \left\lceil \frac{\log(1/K)}{\log(\mu)} \right\rceil$ we have*

$$\|\rho_n - \sigma_n\|_1 \leq K\mu^n \|\rho_0 - \sigma_0\|_1 + \left(\hat{n} + K\frac{\mu^{\hat{n}} - \mu^n}{1-\mu}\right) \left\|\mathcal{T} - \tilde{\mathcal{T}}\right\|_{1\to 1}.$$

In order to apply the theorem we must provide a priori estimates on $\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1}$, which is achieved in Article I). For any μ such that $|\lambda_i| < \mu < 1\ \forall i$, where λ_i denote the eigenvalues of $\mathcal{T} - \mathcal{T}^\infty$ it holds that

$$\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1} \leq \frac{4e\sqrt{|m|}}{(1-\mu)^{3/2}} \sup_{|z|=\mu} \left| \prod_{i\in m} \frac{1-\bar{\lambda}_i z}{z-\lambda_i} \right| \mu^{n+1}$$

Here, m denotes the minimal polynomial of $\mathcal{T} - \mathcal{T}^\infty$ and $|m|$ is the number of linear factors in m. The product is taken over all i such that the corresponding factor $(z-\lambda_i)$ occurs in the prime factorization of m. Plugging this estimate into Theorem 2 yields significant improvement on known stability estimates for classical Markov chains and a strong stability result in the quantum context. The core conceptual insight is that the (pseudo-hyperbolic) distance of the subdominant eigenvalues of $\mathcal{T}$ to the spectral radius of $\mathcal{T} - \mathcal{T}^\infty$ determines the sensitivity of the chain to perturbation. This is in contrast to previous work [2, Thm. 4.1], where corresponding estimates involved (inverse) distances $|\lambda_i - \lambda_j|^{-1}$. The latter, however, *diverge* when degeneracy occurs in the spectrum making the estimate trivial.

2 Legal statement

References

[1] E. Cho and C. Meyer. Comparison of perturbations bounds for the stationary distributions of a Markov chain. *Lin. Alg. and Appl.*, 335:1:137–150, 2001.

[2] A. Mitrophanov. Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Prob.*, 42:1003–1014, 2005.

[3] E. Seneta. Sensitivity of finite Markov chains under perturbation. *Stat. and Prob. let.*, 17.2:163–168, 1993.

JOURNAL OF MATHEMATICAL PHYSICS **54**, 032203 (2013)

Perturbation bounds for quantum Markov processes and their fixed points

Oleg Szehr and Michael M. Wolf
Department of Mathematics, Technische Universität München, 85748 Garching, Germany

(Received 2 November 2012; accepted 22 February 2013; published online 19 March 2013)

We investigate the stability of quantum Markov processes with respect to perturbations of their transition maps. In the first part, we introduce a condition number that measures the sensitivity of fixed points of a quantum channel to perturbations. We establish upper and lower bounds on this condition number in terms of subdominant eigenvalues of the transition map. In the second part, we consider quantum Markov processes that converge to a unique stationary state and we analyze the stability of the evolution at finite times. In this way we obtain a linear relation between the mixing time of a quantum Markov process and the sensitivity of its fixed point with respect to perturbations of the transition map. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4795112]

I. INTRODUCTION

Quantum Markov processes naturally occur in various directions of quantum physics such as quantum statistical physics, quantum optics, or quantum information theory. Whenever the time evolution of some quantum system does not depend on its history, it can be appropriately described as a quantum Markov process. Here we have in particular in mind evolutions of open quantum systems which eventually converge to a set of stationary states.

Such evolutions either arise naturally from relaxation or equilibration, or they may be engineered for instance for the purpose of dissipative quantum computation,[18] dissipative quantum state preparation[3,5,18] or quantum Metropolis sampling.[17] In those cases, the quantum Markov chain is designed so that it drives any initial state towards a sought target—preferably as rapid as possible.

The present work is devoted to the question how sensitive stationary states are to perturbations of the transition map of the corresponding Markov chain. While this appears to be a well studied subject for classical Markov chains,[2,6,7,13,14] it is, to the best of our knowledge, essentially untouched territory for their quantum counterparts. Guided by the classical theory we will follow two alternative approaches, both of which result in an inequality of the form

$$\|\rho_1 - \rho_2\| \leq \kappa \|\mathcal{T}_1 - \mathcal{T}_2\|. \tag{1}$$

Equation (1) relates the distance between two stationary states ρ_1 and ρ_2 to the distance between two quantum channels $\mathcal{T}_1$ and $\mathcal{T}_2$ from which those states arise as fixed points $\rho_i = \mathcal{T}_i(\rho_i)$. A little thought reveals that such an inequality cannot hold in general if κ is a constant merely depending on the chosen norm and possibly the dimension of the underlying space: let the $\mathcal{T}_i$'s for instance be random dissipative perturbations of a unitary evolution. Then, irrespective of the size of the perturbation, there will generically not be any relation between the corresponding stationary states. So their distance cannot be bounded in terms of the perturbation of the transition map, unless κ depends on additional properties of at least one of the channels, e.g., $\kappa = \kappa(\mathcal{T}_1)$. A property which suggests itself in this context is the rate of convergence: intuitively, if the Markov chain generated by $\mathcal{T}_1$ converges rapidly towards ρ_1, one expects that the fixed point is rather robust with respect to perturbations of the transition map $\mathcal{T}_1$. Conversely, if the mixing time is very long, i.e., if there are other states which are almost stationary already and converge to ρ_1 only on a very large time scale, one expects a small perturbation to be sufficient in order to change the stationary state significantly.

We will follow two approaches which make this intuition rigorous. In Sec. III we will directly derive an inequality of the form in Eq. (1) where κ is expressed in terms of a particular condition number which we will relate to spectral properties of the transition map in Sec. III B. Alternatively, in Sec. IV we will derive perturbation bounds for finite times for discrete as well as for continuous time quantum Markov processes with unique stationary state. Those bounds will be expressed in terms of an assumed exponential convergence bound. Hence, they are applicable whenever such a convergence bound can be obtained via one of the various existing tools such as logarithmic Sobolev inequalities,[4] χ^2-divergence,[16] Hilbert's projective metric,[10] or spectral theory.[15] If by any of those tools a time scale is identified on which convergence is guaranteed, then the results in Sec. IV essentially provide a linear bound on the sensitivity coefficient in terms of that mixing time bound.

II. PRELIMINARIES

A. Quantum states and quantum evolutions

We begin with fixing the notation and terminology. We will throughout consider finite-dimensional Hilbert spaces isomorphic to $\mathbb{C}^d$ for some $d \in \mathbb{N}$. The notion of a *state* refers to a density matrix, i.e., a positive semidefinite matrix $\rho \in \mathcal{M}_d(\mathbb{C})$, $\rho \geq 0$ with unit trace $\mathrm{tr}[\rho] = 1$. Here, $\mathcal{M}_d(\mathbb{C})$ denotes the space of complex valued $d \times d$ matrices. The objects of interest in this work are linear maps on $\mathcal{M}_d(\mathbb{C})$ for which we reserve the letters $\mathcal{T}, \mathcal{E}$, and $\mathcal{L}$. For each such map $\mathcal{T}$ the *dual map* $\mathcal{T}^* : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ is defined by imposing that $\forall A, B \in \mathcal{M}_d(\mathbb{C}) : \mathrm{tr}[\mathcal{T}^*(A)B] = \mathrm{tr}[A\mathcal{T}(B)]$. $\mathcal{T}$ is called *Hermiticity preserving* iff $\forall A \in \mathcal{M}_d(\mathbb{C}) : \mathcal{T}(A)^\dagger = \mathcal{T}(A^\dagger)$, *positive* iff $A \geq 0 \Rightarrow \mathcal{T}(A) \geq 0$ and *trace-preserving* iff $\forall A \in \mathcal{M}_d(\mathbb{C}) : \mathrm{tr}[\mathcal{T}(A)] = \mathrm{tr}[A]$. The latter is equivalent to the fact that the dual map preserves the identity matrix $\mathbb{1} = \mathcal{T}^*(\mathbb{1})$. The identity map on $\mathcal{M}_d(\mathbb{C})$ will be denoted by id.

Our primary interest lies in *quantum channels*, i.e., *completely positive* and trace-preserving linear maps, which describe the time evolution of quantum systems for a single time step. We will, however, state all our results for maps which are positive but not necessary completely positive since the proofs do not require the stronger assumption of complete positivity.

Let now $\mathcal{T}$ be any linear, trace-preserving and positive map on $\mathcal{M}_d(\mathbb{C})$. The spectrum $\mathrm{spec}[\mathcal{T}] := \{\lambda \in \mathbb{C} | \exists X : \mathcal{T}(X) = \lambda X\}$ then contains 1, is closed with respect to complex conjugation and is contained in the unit disc. That is, $\lambda \in \mathrm{spec}[\mathcal{T}] \Rightarrow \bar{\lambda} \in \mathrm{spec}[\mathcal{T}]$ and the spectral radius $\varrho(\mathcal{T}) := \max\{|\lambda| \big| \lambda \in \mathrm{spec}[\mathcal{T}]\}$ satisfies $\varrho(\mathcal{T}) = 1 \in \mathrm{spec}[\mathcal{T}]$.

A state which satisfies $\rho = \mathcal{T}(\rho)$ will be called a *stationary state*. The set of stationary states is always non-empty and in fact spans the space of all fixed points of $\mathcal{T}$. The projection onto this space will be denoted by $\mathcal{T}^\infty$ and it can be expressed as a Cesàro mean via

$$\mathcal{T}^\infty = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathcal{T}^k,$$

where $\mathcal{T}^k = \mathcal{T} \circ \ldots \circ \mathcal{T}$ stands for the k-fold composition of $\mathcal{T}$. Clearly, if 1 is the only eigenvalue of $\mathcal{T}$ of modulus one, then this simplifies to $\mathcal{T}^\infty = \lim_{n\to\infty} \mathcal{T}^n$. Note that the spectral properties of $\mathcal{T}$ and $\mathcal{T}^\infty$ guarantee that the map

$$\mathcal{Z}(\mathcal{T}) := \left(\mathrm{id} - (\mathcal{T} - \mathcal{T}^\infty)\right)^{-1}, \tag{2}$$

always exists. For more details on spectral properties of (completely) positive maps we refer to Ref. 19.

When applied to an initial state, the sequence $\{\mathcal{T}^n\}_{n\in\mathbb{N}}$ can be regarded as a finite and homogeneous *quantum Markov chain* with $\mathcal{T}$ as its transition map. The classical case described by a stochastic matrix $S \in \mathcal{M}_d(\mathbb{R}_+)$ can be embedded into this framework by fixing an orthonormal basis $\{|i\rangle\}_{i=1}^d$ and setting $\mathcal{T}(\cdot) = \sum_{i,j=1}^d S_{ij} \langle i| \cdot |i\rangle \, |j\rangle\langle j|$.

B. Norms and contraction coefficients

For any $X \in \mathcal{M}_d(\mathbb{C})$ we denote by $\|X\|_1 := \mathrm{tr}\left[\sqrt{X^\dagger X}\right]$ the *Schatten 1-norm* or *trace norm* of X. When applied to quantum states, the induced metric $(\rho_1, \rho_2) \mapsto \|\rho_1 - \rho_2\|_1$ quantifies how well the two states can be distinguished in an optimally chosen statistical experiment.

For any linear map $\mathcal{L} : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ the induced 1-to-1-norm is defined as

$$\|\mathcal{L}\|_{1\to 1} := \sup_{X\neq 0} \frac{\|\mathcal{L}(X)\|_1}{\|X\|_1}, \ X \in \mathcal{M}_d(\mathbb{C}).$$

By Gelfand's formula we can express the spectral radius of $\mathcal{L}$ in terms of this norm as[1]

$$\varrho(\mathcal{L}) = \lim_{n\to\infty} \left\|\mathcal{L}^n\right\|_{1\to 1}^{1/n}. \tag{3}$$

If $\mathcal{T}$ is trace-preserving and positive, then $\|\mathcal{T}\|_{1\to 1} = 1$.

We will frequently use the so called *coefficient of ergodicity* or *trace norm contraction coefficient* which is defined as

$$\tau(\mathcal{L}) := \sup_{\substack{\sigma^\dagger=\sigma\neq 0 \\ \mathrm{tr}(\sigma)=0}} \frac{\|\mathcal{L}(\sigma)\|_1}{\|\sigma\|_1}.$$

This quantity can equivalently be obtained via an optimization over orthogonal pure states,[11]

$$\tau(\mathcal{L}) = \frac{1}{2} \sup_{\varphi\perp\psi} \|\mathcal{L}(|\varphi\rangle\langle\varphi|) - \mathcal{L}(|\psi\rangle\langle\psi|)\|_1. \tag{4}$$

Here the supremum is taken over all pairs of orthogonal unit vectors. For linear maps which are Hermiticity preserving and trace-preserving it follows readily from the definition of τ that

$$\tau(\mathcal{L}_1 \circ \mathcal{L}_2) \leq \tau(\mathcal{L}_1)\tau(\mathcal{L}_2), \text{ and so } \tau(\mathcal{L}^n) \leq \tau(\mathcal{L})^n \tag{5}$$

for all $n \in \mathbb{N}$. Finally, note that $0 \leq \tau(\mathcal{T}) \leq 1$ if $\mathcal{T}$ is positive and trace-preserving.

III. STABILITY OF FIXED POINTS

A. The main inequality

One of the first sensitivity analyses for fixed points of classical Markov chains was provided by Schweitzer[12] in terms of the so called *fundamental matrix* of a classical Markov chain. Here, we generalize his approach to the quantum setting. The immediate analogue of Schweitzer's fundamental matrix is the map $\mathcal{Z}(\mathcal{T}) : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ defined in Eq. (2). This leads to the main inequality:

Theorem 1. *Let $\mathcal{T}_1, \mathcal{T}_2 : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$ be trace-preserving, positive linear maps. For every stationary state ρ_2 of $\mathcal{T}_2$ the stationary state $\rho_1 := \mathcal{T}_1^\infty(\rho_2)$ of $\mathcal{T}_1$ satisfies*

$$\|\rho_1 - \rho_2\|_1 \leq \kappa \, \|\mathcal{T}_1 - \mathcal{T}_2\|_{1\to 1} \quad with \quad \kappa = \tau\big(\mathcal{Z}(\mathcal{T}_1)\big). \tag{6}$$

Proof. For all such pairs ρ_1, ρ_2 it holds that

$$\mathcal{Z}(\mathcal{T}_1)^{-1}(\rho_1 - \rho_2) = \big(\mathrm{id} - (\mathcal{T}_1 - \mathcal{T}_1^\infty)\big)(\rho_1 - \rho_2) \tag{7}$$

$$= \mathcal{T}_1(\rho_2) - \rho_2 = (\mathcal{T}_1 - \mathcal{T}_2)(\rho_2), \tag{8}$$

which leads to the identity

$$(\rho_1 - \rho_2) = \mathcal{Z}(\mathcal{T}_1) \circ (\mathcal{T}_1 - \mathcal{T}_2)(\rho_2). \tag{9}$$

Taking the Schatten 1-norm on both sides and abbreviating $\sigma := (\mathcal{T}_1 - \mathcal{T}_2)(\rho_2)$ we can write

$$\|\rho_1 - \rho_2\|_1 = \frac{\|\mathcal{Z}(\mathcal{T}_1)(\sigma)\|_1}{\|\sigma\|_1} \, \frac{\|(\mathcal{T}_1 - \mathcal{T}_2)(\rho_2)\|_1}{\|\rho_2\|_1}, \tag{10}$$

from which we obtain the claimed inequality by taking the supremum over all $\rho_2 \in \mathcal{M}_d(\mathbb{C})$ and over all traceless Hermitian σ. □

Evidently, the identity in Eq. (9) can be used to derive a plethora of different norm bounds (cf. Ref. 2 for an overview on different approaches for classical Markov chains). Here we focus on the trace norm since this seems to be the most natural choice in the quantum context. In addition, the trace norm dominates all other unitarily invariant norms on $\mathcal{M}_d(\mathbb{C})$[1] and makes the obtained bounds in this sense the strongest possible ones.

In the following proposition we bound the condition number of Theorem 1 in terms of a better studied object[10,11,16] with an operational meaning, namely the trace-norm contraction coefficient of the quantum channel:

Proposition 2. Let $\mathcal{T}$ be a trace-preserving, positive linear map on $\mathcal{M}_d(\mathbb{C})$ with a unique stationary state. Then

$$\tau\big(\mathcal{Z}(\mathcal{T})\big) \leq \big(1-\tau(\mathcal{T})\big)^{-1}. \tag{11}$$

Proof. We express $\mathcal{Z}(\mathcal{T})$ via its von Neumann series expansion $\mathcal{Z} = \sum_{k=0}^{\infty}(\mathcal{T}-\mathcal{T}^{\infty})^k$ and get that

$$\begin{aligned}\tau(\mathcal{Z}) &= \sup_{\substack{\sigma^\dagger=\sigma \\ \mathrm{tr}(\sigma)=0}} \frac{\big\|\sum_{k=0}^{\infty}(\mathcal{T}-\mathcal{T}^{\infty})^k(\sigma)\big\|_1}{\|\sigma\|_1} \\ &\leq \sum_{k=0}^{\infty}\sup_{\substack{\sigma^\dagger=\sigma \\ \mathrm{tr}(\sigma)=0}} \frac{\big\|(\mathcal{T}-\mathcal{T}^{\infty})^k(\sigma)\big\|_1}{\|\sigma\|_1} = \sum_{k=0}^{\infty}\tau(\mathcal{T}^k) \qquad (12)\\ &\leq \sum_{k=0}^{\infty}[\tau(\mathcal{T})]^k = \frac{1}{1-\tau(\mathcal{T})}, \quad \text{if} \quad \tau(\mathcal{T})<1.\end{aligned}$$

To obtain Eq. (12) we used that $(\mathcal{T}-\mathcal{T}^{\infty})^k = \mathcal{T}^k - \mathcal{T}^{\infty}$ if $k>0$ and that $\mathrm{tr}(\sigma)=0$ implies $\mathcal{T}^{\infty}(\sigma)=0$ since uniqueness of the fixed point means that $\mathcal{T}^{\infty}$ acts as $X \mapsto \mathrm{tr}[X]\rho$. □

Note that $\tau(\mathcal{T})$ has an operational meaning. Since $\tau(\mathcal{T}) = \frac{1}{2}\sup_{\varphi\perp\psi}\|\mathcal{T}(|\varphi\rangle\langle\varphi|)-\mathcal{T}(|\psi\rangle\langle\psi|)\|_1$ by Eq. (4), it is directly related to the maximum probability with which two orthogonal inputs can be distinguished at the output of $\mathcal{T}$.

B. Spectral bounds on $\tau(\mathcal{Z})$

In this subsection we prove that the sensitivity of the set of stationary states of a quantum Markov chain to perturbations is related to the closeness of the subdominant eigenvalues to 1. More precisely, we show that if there exists a subdominant eigenvalue of $\mathcal{T}$ close to 1, then the chain is ill conditioned in the sense that $\tau(\mathcal{Z})$ is large. On the other hand if all eigenvalues are well separated from 1, the process is well conditioned. The following theorem quantifies this observation. We note that the relevant spectral quantity is not equal to the *spectral gap* $\min\{1-|\lambda|\,\big|\,\lambda\in\mathrm{spec}[\mathcal{T}]\setminus\{1\}\}$ which also appears frequently in convergence analyses.

Theorem 3. *Let $\mathcal{T}$ be a trace-preserving, positive linear map on $\mathcal{M}_d(\mathbb{C})$ and $\Lambda := \mathrm{spec}[\mathcal{T}]\setminus\{1\}$ the set of its non-unit eigenvalues. Then*

$$\frac{1}{\min_{\lambda\in\Lambda}|1-\lambda|} \leq \tau\big(\mathcal{Z}(\mathcal{T})\big) \leq \frac{2(5\pi/3+2\sqrt{2})d^3}{\min_{\lambda\in\Lambda}|1-\lambda|}. \tag{13}$$

Proof. We begin with proving the left hand inequality—guided by the techniques developed for classical Markov chains in Ref. 13. We abbreviate $\mathcal{Z} := \mathcal{Z}(\mathcal{T})$ and note that $\mathcal{Z}$ is trace-preserving,

since $\mathcal{Z}^{-1}$ is trace-preserving and therefore $\mathrm{tr}[\mathcal{Z}(X)] = \mathrm{tr}[\mathcal{Z}^{-1} \circ \mathcal{Z}(X)] = \mathrm{tr}[X]$. Consequently, $\mathcal{Z}^*(\mathbb{1}) = \mathbb{1}$. We write $\mathcal{P}$ for the projection onto the invariant subspace of $\mathcal{M}_d(\mathbb{C})$ corresponding to the eigenvalue 1 of $\mathcal{Z}$. Note that this implies that

$$(\mathcal{Z} - \mathcal{P})^k = \mathcal{Z}^k \circ (\mathrm{id} - \mathcal{P}).$$

Using the fact that any matrix σ can be expressed as a sum of a Hermitian matrix $\sigma_+ := (\sigma + \sigma^\dagger)/2$ and a skew-Hermitian matrix $i\sigma_- := (\sigma - \sigma^\dagger)/2$, i.e., $\sigma = \sigma_+ + i\sigma_-$, we can bound

$$\begin{aligned}
\left\|(\mathcal{Z} - \mathcal{P})^k\right\|_{1\to 1} &= \left\|\mathcal{Z}^k \circ (\mathrm{id} - \mathcal{P})\right\|_{1\to 1} \\
&= \sup_{\sigma=\sigma_+ + i\sigma_-} \frac{\left\|\mathcal{Z}^k \circ (\mathrm{id} - \mathcal{P})(\sigma_+ + i\sigma_-)\right\|_1}{\|\sigma_+ + i\sigma_-\|_1} \\
&\leq \sup_{\sigma_+,\sigma_-} \frac{\left\|\mathcal{Z}^k \circ (\mathrm{id} - \mathcal{P})(\sigma_+)\right\|_1}{\|\sigma_+\|_1} + \frac{\left\|\mathcal{Z}^k \circ (\mathrm{id} - \mathcal{P})(\sigma_-)\right\|_1}{\|\sigma_-\|_1} \qquad (14) \\
&\leq 2 \sup_{\substack{\sigma^\dagger=\sigma \\ \mathrm{tr}(\sigma)=0}} \frac{\left\|\mathcal{Z}^k(\sigma)\right\|_1}{\|\sigma\|_1} \|\mathrm{id} - \mathcal{P}\|_{1\to 1} \qquad (15) \\
&= 2\tau(\mathcal{Z}^k) \|\mathrm{id} - \mathcal{P}\|_{1\to 1} \\
&\leq 2\tau(\mathcal{Z})^k \|\mathrm{id} - \mathcal{P}\|_{1\to 1}. \qquad (16)
\end{aligned}$$

To obtain Eq. (14) we apply the triangle inequality in the numerator and note that again by the triangle inequality $||\sigma_i||_1 \leq ||\sigma||_1$, $i \in \{+, -\}$ holds to bound the denominator. Inequality (15) exhibits the fact that both σ_+ and σ_- are Hermitian and that $(\mathrm{id} - \mathcal{P})(\sigma)$ is traceless for all σ. To obtain Eq. (16) we used Eq. (5). Taking the kth root and the limit $k \to \infty$ on both sides of the above derivation, we conclude with Eq. (3) that

$$\varrho(\mathcal{Z} - \mathcal{P}) \leq \tau(\mathcal{Z}). \qquad (17)$$

That is, $\tau(\mathcal{Z})$ provides an upper bound on the modulus of all non-unit eigenvalues of $\mathcal{Z}$. Finally, note that the spectrum of $\mathcal{Z}$ is given by $\mathrm{spec}[\mathcal{Z}] = \{1\} \cup \{\frac{1}{1-\lambda}\}_{\lambda\in\Lambda}$ from which the lower bound in the theorem follows.

For the upper bound we use known results from non-classical spectral theory. The core observation is that the map $\Delta := \mathcal{T} - \mathcal{T}^\infty$ is *power bounded* since

$$\begin{aligned}
\left\|\Delta^n\right\|_{1\to 1} &= \left\|\mathcal{T}^n - \mathcal{T}^\infty\right\|_{1\to 1} \\
&\leq \left\|\mathcal{T}^n\right\|_{1\to 1} + \left\|\mathcal{T}^\infty\right\|_{1\to 1} = 2.
\end{aligned}$$

In Ref. 20 it has been shown that the resolvent of a general power bounded operator Δ, which acts on a complex D-dimensional Banach space and whose spectrum is contained in the open unit disc, satisfies

$$\left\|(\mu\,\mathrm{id} - \Delta)^{-1}\right\| \leq \frac{C\left(\frac{5\pi}{3} + 2\sqrt{2}\right) D^{3/2}}{\min_{\lambda\in\mathrm{spec}[\Delta]} |\mu - \lambda|}, \qquad (18)$$

for all $|\mu| \geq 1$ and $C := \sup_n \|\Delta^n\|$, where $\|\cdot\|$ denotes the usual operator norm induced by the norm of the Banach space. The core observation in Ref. 20 is that one can bound the norm $\|(\mu\,\mathrm{id} - \Delta)^{-1}\|$ by employing a Wiener algebra functional calculus and bounding $\|\frac{1}{\mu - z}\|_{W/mW} := \inf\{\|\frac{1}{\mu - z} + mg\|_W \mid g \in W\}$, where $\|\cdot\|_W$ denotes the Wiener norm and $m \neq 0$ is the minimal degree polynomial annihilating Δ, i.e., $m(\Delta) = 0$. For more details concerning the techniques employed see Refs. 8, 15, and 20, and references therein.

Suppose for now that the only eigenvalue of $\mathcal{T}$ of magnitude one is 1. Then, all eigenvalues of Δ are contained in the open unit disc. Setting $D = d^2$, $\mu = 1$, observing that $\mathrm{spec}[\Delta] = \Lambda$ and bounding $C \leq 2$, Eq. (18) specializes to the upper bound claimed in the theorem.

To incorporate the case when $\mathcal{T}$ has eigenvalues of magnitude one other than 1, i.e., when the spectrum of Δ is merely contained in the closed unit disc, we employ an argument based on continuity. We consider a map $\mathcal{T}_\epsilon$ whose spectrum differs from the one of $\mathcal{T}$ in that the peripheral eigenvalues other than 1 of $\mathcal{T}$ are shifted "by ϵ" radially towards the center of the unit disc. More precisely, we define $\mathcal{T}_\epsilon := \mathcal{T} - \epsilon(\mathcal{T}_\phi - \mathcal{T}^\infty)$, where $\mathcal{T}_\phi$ denotes the part of the spectral decomposition of $\mathcal{T}$ which belongs to all eigenvalues of magnitude one, i.e., if $\mathcal{T} = \sum_k \lambda_k \mathcal{P}_k$ then $\mathcal{T}_\phi = \sum_{k:|\lambda_k|=1} \lambda_k \mathcal{P}_k$. Exploiting the relations between $\mathcal{T}$, $\mathcal{T}^\infty$, and $\mathcal{T}_\phi$ we can show that

$$\begin{aligned}\Delta_\epsilon^n &:= \left(\mathcal{T}_\epsilon - \mathcal{T}^\infty\right)^n \\ &= \mathcal{T}^n - (1-\epsilon)^n \mathcal{T}^\infty + [(1-\epsilon)^n - 1]\mathcal{T}_\phi^n.\end{aligned}$$

Since the involved maps are all positive and trace-preserving, and thus have unit norm, this implies that Δ_ϵ is power bounded with $\left\|\Delta_\epsilon^n\right\|_{1\to 1} \leq 2$ as before. Thus, for $\mu = 1$ and any (small enough) $\epsilon > 0$ the above assertion (18) holds for Δ_ϵ. Then by continuity the statement stays true even for $\epsilon = 0$.

IV. FINITE TIME PERTURBATION BOUNDS

So far, we have analysed the stability of the fixed point of a quantum channel and in this sense the robustness of the asymptotic time evolution of the corresponding quantum Markov chain. In this section we will extend the analysis to finite times, first for discrete and then for continuous time evolutions. A second point in which the following approach differs from the previous one is that it uses the assumption of an exponential convergence bound as an additional ingredient.

A. Evolution in discrete time

Theorem 4. *For $n \in \mathbb{N}_0$ let $\rho_n := \mathcal{T}^n(\rho_0)$ and $\sigma_n := \mathcal{E}^n(\sigma_0)$ be the evolution of two density matrices with respect to two positive and trace-preserving linear maps $\mathcal{T}, \mathcal{E} : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$. If $\mathcal{T}$ has a unique stationary state and $\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1} \leq K \cdot \mu^n$ for $K \geq 0$, $\mu < 1$ and all $n \in \mathbb{N}_0$, then we can bound the distance between the evolved states by*

$$\|\rho_n - \sigma_n\|_1 \leq$$
$$\begin{cases} \|\rho_0 - \sigma_0\|_1 + n\,\|\mathcal{E} - \mathcal{T}\|_{1\to 1} & \text{for } n \leq \hat{n} \\ K\mu^n \|\rho_0 - \sigma_0\|_1 + \left(\hat{n} + K\frac{\mu^{\hat{n}} - \mu^n}{1-\mu}\right)\|\mathcal{E} - \mathcal{T}\|_{1\to 1} & \text{for } n > \hat{n}, \end{cases}$$

where $\hat{n} := \left\lceil \frac{\log(1/K)}{\log(\mu)} \right\rceil$.

Remark. Before we prove this statement, let us mention known pairs (K, μ) to which the theorem might be applied. For definitions and further details we refer to the references:

(i) For $K = 1$ one can choose $\mu = \tanh(\Delta/4)$, where Δ is the *projective diameter* of the map $\mathcal{T}$, measured in terms of Hilbert's projective metric.[10]

(ii) For $K = \sup_\rho \left[\chi_k^2(\rho, \sigma)\right]^{1/2}$ a particular χ^2-*divergence* and σ the stationary state of $\mathcal{T}$, we can choose μ to be the second largest singular value of $\Omega := [\Omega_\sigma^k]^{1/2} \circ \mathcal{T} \circ [\Omega_\sigma^k]^{-1/2}$ where Ω_σ^k is a map on which the chosen χ^2-divergence is based on Ref. 16.
If λ_{min} is the smallest eigenvalue of σ, a particular choice results in $K = (\lambda_{min}^{-1} - 1)^{1/2}$ and $\Omega(X) = \sigma^{-1/4}\mathcal{T}\left(\sigma^{1/4} X \sigma^{1/4}\right)\sigma^{-1/4}$.

(iii) If λ_{min} is the smallest eigenvalue of the stationary state of $\mathcal{T}$, we can choose $K = \sqrt{-2\log\lambda_{min}}$ and μ determined by a logarithmic Sobolev inequality.[4] Strictly speaking, those bounds apply to the continuous time case, which we discuss in Theorem 6 below.

(iv) If there is a similarity transformation such that $S \circ \mathcal{T} \circ S^{-1}$ is a normal operator on $\mathcal{M}_d(\mathbb{C})$, we can choose $\mu := \max\{|\lambda| \big| \lambda \in \mathrm{spec}[\mathcal{T}] \setminus \{1\}\}$ and $K = \sqrt{2d}\kappa_{\mathcal{T}}$, where $\kappa_{\mathcal{T}} := \|S \otimes S^{-1}\|_{2\to 2}$. The latter can be upper bounded by $\kappa_{\mathcal{T}} \leq \lambda_{min}^{-1/2}$ if $\mathcal{T}$ satisfies *detailed balance* with respect to its stationary state.

(v) Finally, we note that the assumption that $\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1} \leq K \cdot \mu^n$ for $K \geq 0$, $\mu < 1$ of Theorem 4 implies that all non-unit eigenvalues of $\mathcal{T}$ are contained in the open unit disc. In this situation more elaborate bounds, which only depend on the spectrum of $\mathcal{T}$ can be given. In Ref. 15 a Wiener algebra functional calculus is employed to obtain spectral convergence bounds for classical and quantum Markov chains. The techniques of Ref. 15 are new even to the theory of classical Markov chains and do not rely on additional assumptions such as *detailed balance*. The derivation of Corollary IV.4 of Ref. 15 yields that for any μ such that $|\lambda_i| < \mu < 1 \, \forall i$, where $(\lambda_i)_{i=1,\ldots,d^2}$ denote the eigenvalues of $\mathcal{T} - \mathcal{T}^\infty$ it holds that

$$\begin{aligned} \left\|\mathcal{T}^n - \mathcal{T}^\infty\right\|_{1\to 1} &\leq \frac{4e\sqrt{|m|}}{(1-\mu)^{3/2}} \sup_{|z|=\mu} \left| \prod_{i\in m} \frac{1-\bar{\lambda}_i z}{z-\lambda_i} \right| \mu^{n+1} \\ &\leq \frac{4e\sqrt{|m|}}{(1-\mu)^{3/2}} \prod_{i\in m} \frac{1-\mu|\lambda_i|}{\mu-|\lambda_i|} \, \mu^{n+1}. \end{aligned} \tag{19}$$

Here, m denotes the minimal polynomial of $\mathcal{T} - \mathcal{T}^\infty$ and $|m|$ is the number of linear factors in m. The product in Eq. (19) is taken over all i such that the corresponding factor $(z - \lambda_i)$ occurs in the prime factorization of m.

Theorem 4 together with Eq. (19) provides a purely spectral bound on the sensitivity of a Markov chain under perturbations. Even compared to the results for classical Markov chains in Ref. 7 (on which our derivation of Theorem 4 builds), bounds based on (19) yield a significant improvement (compare, Theorem 4.1 of Ref. 7). Our bound proves that the distance of the subdominant eigenvalues of $\mathcal{T}$ to the spectral radius of $\mathcal{T} - \mathcal{T}^\infty$ determines the sensitivity of the chain to perturbation, while their mutual distances, i.e., the quantities $|\lambda_i - \lambda_j|$ for general i, j are not relevant (compare, Theorem 4.1 of Ref. 7). We refer to Ref. 15 for a discussion of Eq. (19) and related results.

It is also possible to use Corollary IV.4 of Ref. 15 directly to derive stability estimates. Note, however, that if in $\|\mathcal{T}^n - \mathcal{T}^\infty\|_{1\to 1} \leq K \cdot \mu^n$ we allow that μ equals the spectral radius of $\mathcal{T} - \mathcal{T}^\infty$ then the prefactor will depend on n. More precisely, Theorem III.2 of Ref. 15 yields that in this case $K = K(n) = Cn^{d_\mu - 1}$, where C does not depend on n and d_μ denotes the size of the largest Jordan block of any eigenvalue of magnitude μ. It is not difficult to extend the derivation of Theorem 4 to the situation, where μ is the spectral radius of $\mathcal{T} - \mathcal{T}^\infty$ and $K(n) = Cn^{d_\mu - 1}$.

Proof of Theorem 4. The proof is guided by techniques used in Ref. 7 for classical Markov processes. First we note that in general for linear maps $\mathcal{T}, \mathcal{E}$,

$$\mathcal{E}^n = \mathcal{T}^n + \sum_{i=0}^{n-1} \mathcal{T}^{n-i-1} \circ (\mathcal{E} - \mathcal{T}) \circ \mathcal{E}^i \quad n \geq 1$$

holds, which can easily be shown by induction. Applying the above to the state σ_0 and subtracting ρ_n from both sides gives

$$\sigma_n - \rho_n = \mathcal{T}^n(\sigma_0 - \rho_0) + \sum_{i=0}^{n-1} \mathcal{T}^{n-i-1} \circ (\mathcal{E} - \mathcal{T})(\sigma_i)$$

from which we conclude that

$$\|\sigma_n - \rho_n\|_1 \leq \tag{20}$$

$$\left\|\mathcal{T}^n(\sigma_0 - \rho_0)\right\|_1 + \sum_{i=0}^{n-1} \left\|\mathcal{T}^{n-i-1} \circ (\mathcal{E} - \mathcal{T})(\sigma_i)\right\|_1 .$$

We now find upper bounds for the norm terms appearing on the right-hand side of Eq. (20). The fact that $\mathcal{E}(\sigma_i) - \mathcal{T}(\sigma_i)$ is Hermitian and traceless implies that

$$\left\|\mathcal{T}^{n-i-1} \circ (\mathcal{E} - \mathcal{T})(\sigma^i)\right\|_1 \leq \tau(\mathcal{T}^{n-i-1}) \left\|\mathcal{E} - \mathcal{T}\right\|_{1\to 1},$$
$$\text{and} \quad \left\|\mathcal{T}^n(\sigma_0 - \rho_0)\right\|_1 \leq \tau(\mathcal{T}^n) \left\|\rho_0 - \sigma_0\right\|_1.$$

Thus, from Eq. (20) we conclude that

$$\left\|\rho_n - \sigma_n\right\|_1 \leq \tau(\mathcal{T}^n) \left\|\rho_0 - \sigma_0\right\|_1 + \left\|\mathcal{E} - \mathcal{T}\right\|_{1\to 1} \sum_{i=0}^{n-1} \tau(\mathcal{T}^i). \tag{21}$$

The term $\tau(\mathcal{T}^n)$ can in turn be bounded using Eq. (4) and the assumed convergence properties of $\mathcal{T}$ by

$$\tau(\mathcal{T}^n) = \sup_{\substack{\sigma^\dagger = \sigma \\ \mathrm{tr}(\sigma)=0}} \frac{\left\|\mathcal{T}^n(\sigma)\right\|_1}{\left\|\sigma\right\|_1} = \frac{1}{2} \sup_{|\phi\rangle, |\psi\rangle} \left\|\mathcal{T}^n(\phi) - \mathcal{T}^n(\psi)\right\|_1$$
$$\leq \sup_{|\phi\rangle} \left\|\mathcal{T}^n(\phi) - \mathcal{T}^\infty(\phi)\right\|_1 \leq K \cdot \mu^n.$$

Note that the first inequality requires uniqueness of the stationary state, i.e., that $\mathcal{T}^\infty(\phi) = \mathcal{T}^\infty(\psi)$.

Alternatively, we can use that $\mathcal{T}^n$ is trace-preserving and positive, so that in total

$$\tau(\mathcal{T}^n) \leq \begin{cases} 1 & \text{for } n < \hat{n} \\ K \cdot \mu^n & \text{for } n \geq \hat{n}. \end{cases}$$

We now find a suitable upper bound on $\sum_i \tau(\mathcal{T}^i)$ by always choosing the better of the two bounds for $\tau(\mathcal{T}^i)$. In this way we obtain

$$\sum_{i=0}^{n-1} \tau(\mathcal{T}^i) \leq \hat{n} + \sum_{i=\hat{n}}^{n-1} \tau(\mathcal{T}^i) \leq \hat{n} + K \cdot \mu^{\hat{n}} \sum_{i=0}^{n-\hat{n}-1} \mu^i$$
$$= \hat{n} + K \cdot \mu^{\hat{n}} \frac{1 - \mu^{n-\hat{n}}}{1 - \mu}. \tag{22}$$

Plugging this expression into Eq. (21) and again choosing the better bound for $\tau(\mathcal{T}^n)$ concludes the proof of the theorem. □

If we take the limit $n \to \infty$ in Theorem 4 and use that $K \cdot \mu^{\hat{n}} \leq 1$ is by definition of $\hat{n}$ basically an equality, we obtain a perturbation bound for the asymptotic states:

Corollary 5. Under the conditions of Theorem 4

$$\limsup_{n\to\infty} \left\|\rho_n - \sigma_n\right\|_1 \leq \left(\hat{n} + \frac{1}{1-\mu}\right) \left\|\mathcal{E} - \mathcal{T}\right\|_{1\to 1}. \tag{23}$$

B. Evolution in continuous time

The following is the quantum counterpart of the results on classical Markov chains in Ref. 6:

Theorem 6. *Let $\mathcal{T}^t = e^{t\mathfrak{L}_\mathcal{T}}$ and $\mathcal{E}^t = e^{t\mathfrak{L}_\mathcal{E}}$ with $t \in \mathbb{R}_+$ be two one-parameter semi groups of positive and trace-preserving linear maps on $\mathcal{M}_d(\mathbb{C})$. Write $\rho(t) := \mathcal{T}^t(\rho_0)$ and $\sigma(t) := \mathcal{E}^t(\sigma_0)$ for the evolution of two density matrices and assume that $\mathcal{T}^t$ has a unique stationary state and that*

$\forall t > 0 : \left\|\mathcal{T}^t - \mathcal{T}^\infty\right\|_{1\to 1} \leq K e^{-\nu\, t}$ *for some* $K, \nu > 0$. *Then*

$$\|\rho(t) - \sigma(t)\|_1 \leq \begin{cases} \|\rho_0 - \sigma_0\|_1 + t\, \|\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T}\|_{1\to 1}\,, & \text{for } t < \hat{t} \\ K e^{-\nu t}\, \|\rho_0 - \sigma_0\|_1 + \frac{\log(K)+1-Ke^{-\nu t}}{\nu}\, \|\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T}\|_{1\to 1} & \\ \qquad \text{for } t \geq \hat{t} \end{cases},$$

where $\hat{t} := \frac{\log(K)}{\nu}$.

Proof. The proof goes along the lines of the proof of Theorem 4. The difference between two dynamical semi-groups can be expressed using their generators as[9]

$$\mathcal{E}_t = \mathcal{T}_t + \int_0^t \mathcal{T}_{t-s} \circ (\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T}) \circ \mathcal{E}_s \, \mathrm{d}s.$$

Following the derivation of Theorem 4 and using that $\forall X : \mathrm{tr}[(\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T})(X)] = 0$, we obtain the continuous time analogue of Eq. (21),

$$\|\rho(t) - \sigma(t)\|_1 \leq \tau(\mathcal{T}_t)\, \|\rho_0 - \sigma_0\|_1 + \|\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T}\|_{1\to 1} \int_0^t \tau(\mathcal{T}_u)\, \mathrm{d}u.$$

Again, it is possible to state upper bounds for $\tau(\mathcal{T}_t)$ for small and large t, respectively. We have that

$$\tau(\mathcal{T}_t) \leq \begin{cases} 1 & \text{for } t \leq \hat{t} \\ K \cdot e^{-\nu t} & \text{for } t > \hat{t}, \end{cases}$$

where $\hat{t} := \frac{\log(K)}{\nu}$. The proof is then concluded following exactly the same lines as in the proof of Theorem 4. □

Again we can consider the limit $t \to \infty$ and thereby obtain a perturbation bound for the asymptotic evolution in terms of the distance between the generators and as a function of the convergence rate ν:

Corollary 7 Under the conditions of Theorem 6

$$\limsup_{t\to\infty} \|\rho(t) - \sigma(t)\|_1 \leq \frac{\log(K)+1}{\nu}\, \|\mathfrak{L}_\mathcal{E} - \mathfrak{L}_\mathcal{T}\|_{1\to 1}\,.$$

V. OUTLOOK

We have established general perturbation bounds for fixed points of quantum Markov chains. The results focus on the trace norm, but it is clear from their derivation, that analogous bounds can be obtained for essentially any norm. For practical purposes and large systems, the derived bounds may be weaker than desired—owing to the fact that we do not impose and exploit any additional structure of transition map and perturbation. Investigating bounds in more structured frameworks, where for instance Liouvillians as well as perturbations are geometrically local, seems to be a worthwhile direction for future studies.

We have also seen that perturbation bounds are linked to convergence bounds so that stronger perturbation bounds can be obtained from better convergence bounds. A detailed analysis of the latter, leading to bounds of the form in Eq. (19), will be given.[15]

Clearly, one may also exploit the relation in the other direction and use the derived perturbation bounds in order to obtain lower bounds on mixing times for quantum Markov processes.

ACKNOWLEDGMENTS

We acknowledge financial support from the European project QUEVADIS, the CHIST-ERA/BMBF project CQC, and the QCCC programme of the Elite Network of Bavaria.

[1] R. Bhatia, *Matrix Analysis* (Springer, 1996).
[2] E. Cho and C. Meyer, "Comparison of perturbations bounds for the stationary distributions of a Markov chain," Linear Algebr. Appl. **335**(1), 137–150 (2001).
[3] S. Diehl, A. Micheli, A. Kantian, B. Kraus, H. Büchler, and P. Zoller, "Quantum states and phases in driven open quantum systems with cold atoms," Nat. Phys. **4**, 878–883 (2008).
[4] M. Kastoryano and K. Temme, "Quantum logarithmic Sobolev inequalities and rapid mixing," preprint arXiv:1207.3261 (2012).
[5] B. Kraus, H. P. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller, "Preparation of entangled states by quantum Markov processes," Phys. Rev. A **78**, 042307 (2008).
[6] A. Mitrophanov, "Stability and exponential convergence of continuous-time Markov chains," J. Appl. Probab. **40**, 970–979 (2003).
[7] A. Mitrophanov, "Sensitivity and convergence of uniformly ergodic Markov chains," J. Appl. Probab. **42**, 1003–1014 (2005).
[8] N. K. Nikolski, "Condition numbers of large matrices and analytic capacities," St. Petersbg. Math. J. **17**, 641–682 (2006).
[9] R. Phillips, "Perturbation theory for semi-groups of linear operators," Trans. Am. Math. Soc. **74**, 199–221 (1953).
[10] D. Reeb, M. J. Kastoryano, and M. M. Wolf, "Hilbert's projective metric in quantum information theory," J. Math. Phys. **52**(8), 082201 (2011).
[11] M. Ruskai, "Beyond strong subadditivity? Improved bounds on the contraction of generalized relative entropy," Rev. Math. Phys. **6**, 1147–1161 (1994).
[12] P. Schweitzer, "Perturbation theory and finite Markov chains," J. Appl. Probab. **5**, 401–413 (1968).
[13] E. Seneta, "Ergodicity coefficients and spectrum localization," Linear Algebr. Appl. **60**, 187–197 (1984).
[14] E. Seneta, "Perturbation of the stationary distribution measured by ergodicity coefficients," Adv. Appl. Probab. **20**, 228–230 (1988).
[15] O. Szehr, D. Reeb, and M. Wolf, "Spectral convergence bounds for classical and quantum Markov processes," preprint arXiv:1301.4827 (2013).
[16] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete, "The chi[sup 2]-divergence and mixing times of quantum Markov processes," J. Math. Phys. **51**(12), 122201 (2010).
[17] K. Temme, T. Osborne, K. Vollbrecht, D. Poulin, and F. Verstraete, "Quantum metropolis sampling," Nature (London) **471**, 87–90 (2011).
[18] F. Versraete, M. Wolf, and I. Cirac, "Quantum computation and quantum-state engineering driven by dissipation," Nat. Phys. **5**(9), 633–636 (2009).
[19] M. Wolf and D. Perez-Garcia, "The inverse eigenvalue problem for quantum channels," preprint arXiv:1005.4545 (2010).
[20] R. Zarouf, "Une amélioration d'un résultat de E.B. Davies et B. Simon," Comptes Rendus Mathematique **347**(15), 939–942 (2009); e-print arXiv:0903.2743v2.

Eigenvalue estimates for the resolvent of a non-normal matrix

O. Szehr December 17, 2013

We investigate the relation between the spectrum of a non-normal matrix and the norm of its resolvent. This relation plays an important role in various situations in pure and applied mathematics. For example in operator theory, when constructing a functional calculus [2]; in the theory of random orthogonal polynomials when localizing their zeroes [1] and in computational linear algebra when one is concerned about numeric stability of solutions of linear equations.

1 Resolvent estimates

We provide the strongest known resolvent estimates for two classes of matrices. Theorem 1 treats matrices with operator norm $\|A\|_\infty \leq 1$ and is optimal. Theorem 2 only requires that for *any* norm $\sup_{k\geq 0}\|A^k\| \leq C < \infty$. Our results improve on previous contributions by E.B. Davies and B. Simon [1] as well as N. Nikolski [2].

Theorem 1. *Let A be an $n \times n$ matrix with spectrum $\sigma(A)$ and $\|A\|_\infty \leq 1$. If $m = \prod_{i=1}^{|m|}(z-\lambda_i)$ is the minimal polynomial of A and $\zeta \in \mathbb{C} - \sigma(A)$ then it holds that*

$$\left\|(\zeta - A)^{-1}\right\|_\infty \leq \left\|(\zeta - M_B)^{-1}\right\|_\infty$$

and

$$\left((\zeta - M_B)^{-1}\right)_{ij} = \begin{cases} 0 & if\ i<j \\ \frac{1}{\zeta-\lambda_i} & if\ i=j \\ \frac{(1-|\lambda_i|^2)^{1/2}}{\zeta-\lambda_i}\frac{(1-|\lambda_j|^2)^{1/2}}{\zeta-\lambda_j}\prod_{\mu=j+1}^{i-1}\left(\frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu}\right) & if\ i>j \end{cases}.$$

The assertion of the theorem is surprising. In order to estimate the resolvent of a given matrix A it is sufficient to consider a certain matrix M_B, whose resolvent has largest operator norm among all (normalized) matrices. The optimality of the estimate is trivial. We provide the entries of $\left((\zeta - M_B)^{-1}\right)_{ij}$, which makes it possible to simply bound $\left\|(\zeta - M_B)^{-1}\right\|_\infty$. The second observation treats power-bounded matrices. We state the theorem for $\zeta \in \partial\mathbb{D} - \sigma(A)$ as this most clearly demonstrates its strength. (The statement for $\zeta \in \partial\mathbb{D} - \sigma(A)$ can be found is the contributed article.)

Theorem 2. *Let A be an $n \times n$ matrix with minimal polynomial m of degree $|m|$ and let $\|\cdot\|$ be* any norm *with $\sup_{k\geq 0}\|A^k\| = C < \infty$. For any $\zeta \in \partial\mathbb{D} - \sigma(A)$ it holds that*

$$\left\|(\zeta - A)^{-1}\right\| \leq \frac{\sqrt{16e-4}\,|m|C}{\min_{\lambda_i\in\sigma(A)}|\zeta-\lambda_i|}.$$

The core conceptual insight is that the resolvent of a power-bounded operator grows linearly with the number of factors in the minimal polynomial. Previous results could only establish estimates involving $|m|^{3/2}$ [3]. As compared to (optimal) bounds for A with $\|A\|_\infty \leq 1$ (resulting from Theorem 1), Theorem 2 is weaker only by a constant prefactor of about 5.

2 Legal statement

The content of the article is the exclusive work of the author.

References

[1] B. Davies and B. Simon. Eigenvalue estimates for non-normal matrices and the zeros of random orthogonal polynomials on the unit circle. *Jour. Approx. Theo.*, 141:2:189–213, 2006.

[2] N.K. Nikolski. Condition numbers of large matrices and analytic capacities. *St. Petersburg Math. J.*, 17:641–682, 2006.

[3] R. Zarouf. Une amélération d'un résultat de E.B. Davies et B. Simon. *Acad. Scien. Paris*, I 347:785–790, 2009. arXiv: 0903.2743v2.

Eigenvalue estimates for the resolvent of a non-normal matrix

Oleg Szehr[1,*]

[1]*Department of Mathematics, Technische Universität München, 85748 Garching, Germany*

(Dated: March 27, 2014)

We investigate the relation between the spectrum of a non-normal matrix and the norm of its resolvent. We provide spectral estimates for the resolvent of matrices whose largest singular value is bounded by 1 (so-called Hilbert space contractions) and for power-bounded matrices. In the first case our estimate is optimal and we present explicit matrices that achieve equality in the bound. This result recovers and generalizes previous estimates obtained by E.B. Davies and B. Simon in the study of orthogonal polynomials on the unit circle. In case of power-bounded matrices we achieve the strongest estimate so far. Our result unifies previous approaches, where the resolvent was estimated in certain restricted regions of the complex plane. To achieve our estimates we relate the problem of bounding the norm of a function of a matrix to a Nevanlinna-Pick interpolation problem in a corresponding function space. In case of Hilbert space contractions this problem is connected to the theory of compressed shift operators to which we contribute by providing explicit matrix representations for such operators. Finally, we apply our results to study the sensitivity of the stationary states of a classical or quantum Markov chain with respect to perturbations of the transition matrix.

Keywords: Resolvent, Non-normal matrix, Markov chain
2010 Mathematics Subject Classification: Primary: 15A60; Secondary: 65F35, 65J05

Contents

*Electronic address: oleg.szehr@posteo.de

I. INTRODUCTION

The contribution of this article is to provide new estimates on the norm of the resolvent of a matrix A and to prove their optimality under certain conditions. We derive bounds of the form

$$\left\|(\zeta - A)^{-1}\right\| \leq \Phi(\zeta, n, \sigma(A)), \tag{1}$$

where Φ is a function of $\zeta \in \mathbb{C}$, the dimension n and the spectrum $\sigma(A)$ of A. In the first part of the article, (cf. Section III) we assume that the largest singular value (the spectral norm) of A is bounded by 1 i.e. $\|A\|_\infty \leq 1$. Note that this can always be achieved by a suitable normalization. Under this assumption we obtain optimal bounds for $\zeta \in \mathbb{C} - \sigma(A)$ and present explicit matrices that establish equality in (1). Thus we identify the relation between the localization of the spectrum of A and the norm of its resolvent. In the second part (cf. Section IV) we study (1) under the assumption that each power of A can be bounded with respect to *any* given norm by the same constant, $\sup_{k\geq 0} \left\|A^k\right\| \leq C$. In this case we derive the strongest estimates so far.

The problem of finding good functions Φ in (1) was studied extensively in the literature [5, 6, 17, 29]. We contribute to the topic by strengthening the results of the cited publications. Our approach is based on the theory of certain (Hilbert/ Banach) function spaces. We associate to a given class of matrices Γ a certain Banach algebra $\mathcal{A}$ of functions and instead of working with matrices directly we estimate the norm of a representative function in the function algebra. A key role is played by inequalities of the type

$$\|f(A)\| \leq C \|f\|_{\mathcal{A}}, \tag{2}$$

which relate for a given $A \in \Gamma$ the norm $\|f(A)\|$ to the norm of f in $\mathcal{A}$. At first glance this appears to be of little use since the right hand side no longer depends on A. However, it is possible to exploit spectral properties of A to significantly strengthen the inequality in (2). Let m_A be the minimal polynomial of A. For any $f, g \in \mathcal{A}$ we have then that $\|(f + m_A g)(A)\|_{\mathcal{A}} = \|f(A)\|_{\mathcal{A}}$ and an application of (2) reveals that for all $g \in \mathcal{A}$ we have $\|f(A)\| \leq C \|f + m_A g\|_{\mathcal{A}}$. This reduces the problem of bounding $\|f(A)\|$ to the problem of finding the least norm function $f + m_A g$ in $\mathcal{A}$, which is equivalent to a Nevanlinna-Pick interpolation problem in $\mathcal{A}$ [17]. If $\|A\|_\infty \leq 1$ the resulting interpolation problem can be solved using an operator theoretic approach pioneered by D. Sarason [1, 22]. This approach is intrinsically connected to the theory of compressed shift operators on Hardy space. We contribute to this theory by providing a framework that allows us to compute explicit matrix representations for functions of model operators. In case that A is power-bounded we choose a rational approximation function in $\mathcal{A}$ and bound its norm to achieve our result.

Bounds on the norm of a resolvent are required in various situations in pure and applied mathematics. For example in operator theory, when constructing a functional calculus

$$f \mapsto f(X) = \frac{1}{2\pi i} \int_\gamma f(\zeta)(\zeta - X)^{-1} \mathrm{d}\zeta$$

one needs to bound the norm of the resolvent in terms of distance of ζ to the spectrum of X [17]. In the theory of orthogonal polynomials resolvent estimates are used to study the

location of zeros of a random orthogonal polynomials on the unit circle [6]. In computational linear algebra the following are classical problems that can be approached through appropriate estimates for $\|(\zeta - A)^{-1}\|$.

1. To analyze the stability of solutions x of the matrix equation $Ax - \zeta x = b$ under perturbations in b and A, see [5].

2. To study whether an approximate eigenvalue ζ of A (in the sense that $\|Ax - \zeta x\| \leq \varepsilon\|x\|$ for some vector x) is close to an actual eigenvalue of A, see [5, 19, 28].

3. To estimate the distance of the spectrum of a matrix B to the spectrum of a matrix A in terms of $B - A$, see [2, 19, 21].

Our resolvent bounds are stronger than the ones used for example in [21] to obtain estimates on the spectral variation of non-normal matrices. In Section V we apply our resolvent estimate for power-bounded matrices to study the sensitivity of stationary states of a classical or quantum Markov chain under perturbations of the transition matrix. We recover known stability results for classical Markov chains and prove new estimates in the quantum case. A similar approach, based on the power-boundedness of the transition matrix, was previously applied in [26] to investigate spectral convergence properties of classical and quantum Markov chains.

II. PRELIMINARIES

We will take a function space based approach to the problem of bounding the norm of the resolvent of a certain matrix. This section lays down the required definitions and basic results.

A. Notation

We denote by $\mathcal{M}_n$ the set of $n \times n$ matrices with complex entries. For $A \in \mathcal{M}_n$ we denote by $\sigma(A)$ its spectrum and by m its minimal polynomial. We write $|m|$ for the degree of m. To the minimal polynomial m we associate the Blaschke product

$$B(z) := \prod_i \frac{z - \lambda_i}{1 - \bar{\lambda}_i z}.$$

The product is taken over all i such that (respecting multiplicities) the corresponding linear factor $z - \lambda_i$ occurs in the minimal polynomial m. Thus, the numerator of B as defined here is exactly the associated minimal polynomial.

We denote by $\|A\|$ *any* particular norm of A while the ∞-norm is defined by

$$\|A\|_\infty = \sup_{\|v\|_2=1} \|Av\|_2 \,,$$

where $\|v\|_2^2 = \sum_i |v_i|^2$ is the usual Euclidean norm. That means $\|A\|_\infty$ simply denotes the largest singular value of A. We will slightly abuse nomenclature and call matrices with

$$\|A\|_\infty \leq 1$$

Hilbert space contractions, although of course the underlying space always has finite dimension. Similarly, the class of $A \in \mathcal{M}_n$ with

$$\sup_{k\geq 0} \|A^k\| \leq C < \infty$$

will be called Banach space power-bounded operators with respect to $\|\cdot\|$ and constant C. (Note that here the norm is general.)

To achieve our estimates we rely on the theory of certain function spaces. Let $\mathbb{D} = \{z \in \mathbb{C} \,|\, |z| < 1\}$ denote the open unit disk in the complex plane and $\bar{\mathbb{D}}$ its closure. The space of analytic functions on $\mathbb{D}$ is denoted by $Hol(\mathbb{D})$. The Hardy spaces considered here are

$$H_2 := \big\{f \in Hol(\mathbb{D}) |\, \|f\|_{H_2}^2 := \sup_{0\leq r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\phi})|^2 \mathrm{d}\phi < \infty\big\},$$

and

$$H_\infty := \big\{f \in Hol(\mathbb{D}) |\, \|f\|_{H_\infty} := \sup_{z\in\mathbb{D}} |f(z)| < \infty\big\}.$$

The H_2-norm can be written in terms of the Taylor coefficients of the analytic function f. We write $f(z) = \sum_{k\geq 0} \hat{f}(k) z^k$ and use Plancherel's identity to conclude that

$$\sup_{0\leq r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\phi})|^2 \mathrm{d}\phi = \sum_{k\geq 0} |\hat{f}(k)|^2.$$

Thus, $f \in Hol(\mathbb{D})$ is in H_2 if and only if $\sum_{k\geq 0} |\hat{f}(k)|^2 < \infty$. The Wiener algebra is defined as the subset of $Hol(\mathbb{D})$ of absolutely convergent Taylor series,

$$W := \{f = \sum_{k\geq 0} \hat{f}(k) z^k |\, \|f\|_W := \sum_{k\geq 0} |\hat{f}(k)| < \infty\}.$$

B. Model spaces and operators

Let $A \in \mathcal{M}_n$ with $\sigma(A) \subset \mathbb{D}$ and let B be the Blaschke product associated to the minimal polynomial of A. We define the $|m|$-dimensional *model space*

$$K_B := H_2 \ominus BH_2 := H_2 \cap (BH_2)^\perp,$$

where we employ the usual scalar product from the Hilbert space $L^2(\partial\mathbb{D})$,

$$\langle f|g\rangle := \int_{\partial\mathbb{D}} f(z)\overline{g(z)}\, \frac{|\mathrm{d}z|}{2\pi}.$$

If the zeros $\{\lambda_i\}_{i=1,\ldots,|m|}$ of B are distinct (that is A can be diagonalized) it is not difficult to verify that K_B is spanned by the *Cauchy kernels*

$$K_B = span\left\{\frac{1}{1-\bar{\lambda}_i z}\right\}_{i=1,\ldots,|m|}.$$

Thus K_B is a space of rational functions f of the form

$$f(z) = \frac{p(z)}{\prod_i (1-\bar{\lambda}_i z)},$$

where $p(z)$ is a polynomial of degree at most $|m|-1$. If the zeros of B are not distinct the above remains valid but the Cauchy kernels have to be replaced by

$$\frac{z^{k-1}}{(1-\bar{\lambda}_i z)^k}, \quad 1 \leq k \leq k_i,$$

where k_i denotes the multiplicity of λ_i. In our consecutive proofs, however, we omit this case and assume that A is diagonalizable. This does not result in any difficulties since upper bounds obtained in the special case extend by continuity to bounds for non-diagonalizable matrices. The assumption that A can be diagonalized is not principal; virtually all computations in the manuscript can be carried out in the more general case. We avoid non-diagonalizable A and rely on continuity only for notational convenience.

One natural orthonormal basis for K_B is the Malmquist-Walsh basis $\{e_k\}_{k=1,\ldots,|m|}$ with ([16], page 117)

$$e_k(z) := \frac{(1-|\lambda_k|^2)^{1/2}}{1-\bar{\lambda}_k z} \prod_{i=1}^{k-1} \frac{z-\lambda_i}{1-\bar{\lambda}_i z},$$

where, as it will remain throughout the manuscript, the empty product is defined to be 1 i.e.

$$e_1(z) = \frac{(1-|\lambda_1|^2)^{1/2}}{1-\bar{\lambda}_1 z}.$$

The *model operator* M_B acts on K_B as

$$\begin{aligned} M_B :\ & K_B \to K_B \\ & f \mapsto M_B(f) := P_B(zf), \end{aligned}$$

where P_B denotes the orthogonal projection on K_B. In other words, M_B is the compression of the multiplication operation by z to the model space K_B (see [16] for a detailed discussion of model operators and spaces). As multiplication by z has operator norm 1 it is clear that M_B is a Hilbert space contraction. Moreover, it is not hard to show that the eigenvalues of M_B are exactly the zeros of the corresponding Blaschke product (see [18], page 228 and Proposition III.5 in the article at hand).

C. Spectral bounds on the norm of a function of a matrix

This subsection contains a brief outline of methods to obtain spectral bounds on a function of a matrix. For a more detailed account see [16–18] and the references therein. Suppose that f is holomorphic on a domain containing all eigenvalues of A and let γ be a smooth curve in this domain that encloses the eigenvalues. The matrix $f(A)$ is defined by the Dunford-Taylor integral [9]

$$f(A) = \frac{1}{2\pi i} \int_\gamma f(\zeta)(\zeta - A)^{-1} \mathrm{d}\zeta.$$

It is easily seen that if $f(z) = \sum_{k=0}^n a_k z^k$ is a polynomial then $f(A) = \sum_{k=0}^n a_k A^k$ and that the correspondence $f \mapsto f(T)$ is an algebra homomorphism from the algebra of holomorphic functions (on the given domain) to $\mathcal{M}_n$ i.e $(f+g)(T) = f(T) + g(T)$ and $(fg)(T) = f(T)g(T)$ (see [9], Chapter I.6). A unital Banach algebra $\mathcal{A}$ with elements in $Hol(\mathbb{D})$ will be called a function algebra if

i) $\mathcal{A}$ contains all polynomials and $\lim_{n\to\infty} \|z^n\|_{\mathcal{A}}^{1/n} = 1$ and

ii) ($f \in \mathcal{A}$, $\lambda \in \mathbb{D}$, $f(\lambda) = 0$) implies that $\frac{f}{z-\lambda} \in \mathcal{A}$.

Following the conventions of [17] we say that a set of matrices Γ obeys an $\mathcal{A}$ functional calculus with constant C if

$$\|f(A)\| \leq C \|f\|_{\mathcal{A}},$$

holds for any $A \in \Gamma$ and $f \in \mathcal{A}$. Here $\|f\|_{\mathcal{A}}$ denotes the norm of f in $\mathcal{A}$. Clearly, this is only possible if all eigenvalues of A are contained in $\bar{\mathbb{D}}$. For us, two instances of such inequalities will be important. In the first example we consider Hilbert space contractions, while the second one treats power-bounded Banach space operators.

i) The family of Hilbert space contractions $\Gamma = \{A \in \mathcal{M}_n | \, \|A\|_\infty \leq 1\}$ is related to an H_∞ functional calculus, since by von Neumann's inequality [5, 14] we have for any f in the disk algebra $H_\infty \cap \mathcal{C}(\bar{\mathbb{D}})$ (the set of bounded holomorphic functions on $\mathbb{D}$ that admit a continuous extension to the boundary) and $A \in \Gamma$ with $\sigma(A) \subset \mathbb{D}$

$$\|f(A)\|_\infty \leq \|f\|_{H_\infty}.$$

ii) Consider a family $\Gamma = \{A \in \mathcal{M}_n | \, \|A^k\| \leq C \ \forall k \in \mathbb{N}\}$ of Banach space operators that are power bounded by some constant $C < \infty$. This family admits a Wiener algebra functional calculus since for any $f \in W$ and $A \in \Gamma$

$$\|f(A)\| = \left\| \sum_{k\geq 0} \hat{f}(k) A^k \right\| \leq \sum_{k\geq 0} |\hat{f}(k)| \left\| A^k \right\| \leq C \sum_{k\geq 0} |\hat{f}(k)| = C \|f\|_W$$

holds.

At first glance, the inequalities of *i*) and *ii*) seem to be of little use when it comes to finding spectral bounds on $\|f(A)\|$ since the obtained upper bounds do not depend on

A anymore. To obtain better estimates one can rely on the following insight. Instead of considering the function f directly, we add multiples of m (or any other annihilating polynomial) to this function and consider $h = f + mg$, $g \in \mathcal{A}$ instead of f. It is immediate that $\|f(X)\| = \|h(X)\|$. The following simple but crucial lemma summarizes this point:

Lemma II.1 ([17] Lemma 3.1)**.** *Let $m \neq 0$ be a polynomial and let Γ be a set of matrices that obey an $\mathcal{A}$ functional calculus with constant C and that satisfy $m(A) = 0\ \forall A \in \Gamma$. Then*

$$\|f(A)\| \leq C \|f\|_{\mathcal{A}/m\mathcal{A}}, \ \forall A \in \Gamma,$$

where $\|f\|_{\mathcal{A}/m\mathcal{A}} = \inf \{\|h\|_{\mathcal{A}} |\ h = f + mg,\ g \in \mathcal{A}\}$.

Proof. For any $g \in A$ we have that $\|f(A)\| = \|(f + mg)(A)\| \leq C \|f + mg\|_{\mathcal{A}}$. □

If $\sigma(A) \subset \mathbb{D}$ (and A can be diagonalized) it follows directly from the definition of the function algebra (see also [17], Section 3.1 (iii) or [30], Section 1.2 P4) that

$$\|f\|_{\mathcal{A}/m\mathcal{A}} = \inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g(\lambda_i) = f(\lambda_i)\ \forall \lambda_i \in \sigma(A)\},$$

which is a Nevanlinna-Pick type interpolation problem in $\mathcal{A}$. If the eigenvalue λ_i carries a multiplicity $k_i > 1$ in m the above remains valid but at λ_i the first $k_i - 1$ derivatives of f and g must coincide. Since for $\sigma(A) \subset \mathbb{D}$ the Blaschke product is holomorphic on a set containing $\bar{\mathbb{D}}$ we can define $\|f\|_{\mathcal{A}/B\mathcal{A}}$ as in Lemma II.1 and note ([30], Lemma 3.1) that as before

$$\|f\|_{\mathcal{A}/B\mathcal{A}} = \inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g(\lambda_i) = f(\lambda_i)\ \forall \lambda_i \in \sigma(A)\}.$$

In the special case $\mathcal{A} = H_\infty$ it is possible to evaluate $\|f\|_{H_\infty/BH_\infty}$ using Sarason's approach to the Nevanlinna-Pick problem [1, 22] or the Commutant lifting theorem of B. Sz.-Nagy and C. Foiaş [1, 8, 13].

Lemma II.2 ([17] Theorem 3.12, [18] Theorem 3.1.11)**.** *For any $f \in H_\infty$ it holds that*

$$\|f\|_{H_\infty/BH_\infty} = \|f(M_B)\|_\infty .$$

III. HILBERT SPACE CONTRACTIONS

Spectral bounds on the resolvent of a Hilbert space contraction were derived in [6]. The authors provide an upper bound in terms of a certain Toeplitz matrix, compute the norm of this matrix and present a sequence of matrices that approaches their upper bound. The following theorem summarizes the basic three assertions from the discussion of Hilbert spaces contractions in [6].

Theorem III.1 ([5, 6])**.**

1. *Let A be an $n \times n$ matrix with $\|A\|_\infty \leq 1$ and $1 \notin \sigma(A)$. Then*

$$\left\|(\mathbb{1} - A)^{-1}\right\|_\infty \leq \frac{\|M_n\|_\infty}{\min_{\lambda_i \in \sigma(A)} |1 - \lambda_i|},$$

with the $n \times n$ matrix

$$M_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 2 & \dots & 2 & 1 \end{pmatrix}.$$

2. *It holds that $\|M_n\|_\infty = \cot\left(\frac{\pi}{4n}\right)$.*

3. *For any $a \in (0,1)$ there are $n \times n$ matrices $A_n(a)$ with $\|A_n(a)\|_\infty \leq 1$ and $\sigma(A) = \{a\}$ such that*

$$\lim_{a \to 1} (1-a)(\mathbb{1} - A_n(a))^{-1} = M_n.$$

In this paper we present generalizations of all three assertions of the theorem. We recover the statements *1* and *3* using a unified approach based on the techniques developed in [17]. Here, our strategy is to directly compute and bound the entries of the model operator in Malmquist-Walsh basis. Our approach has the advantage that it yields spectral bounds for any $\zeta \in \mathbb{C} - \sigma(A)$ and that the optimality statement *3* is automatic. Concerning the second point of the theorem we present a technique going back to [7] in order to compute the norm of Toeplitz matrices of the form

$$M_n(\beta) := \begin{pmatrix} 1 & 0 & \dots & 0 \\ \beta & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \beta & \dots & \beta & 1 \end{pmatrix}, \qquad \beta \in [0,2]. \tag{3}$$

Theorem III.2. *Let A be an $n \times n$ matrix with $\|A\|_\infty \leq 1$ and minimal polynomial $m = \prod_{i=1}^{|m|}(z - \lambda_i)$ with $\sigma(A) \subset \mathbb{D}$. Then for any $\zeta \in \mathbb{C} - \sigma(A)$ it holds that*

$$\left\|(\zeta - A)^{-1}\right\|_\infty \leq \left\|(\zeta - M_B)^{-1}\right\|_\infty$$

and

$$\left((\zeta - M_B)^{-1}\right)_{ij} = \begin{cases} 0 & if\ i < j \\ \frac{1}{\zeta - \lambda_i} & if\ i = j \\ \frac{(1-|\lambda_i|^2)^{1/2}}{\zeta - \lambda_i} \frac{(1-|\lambda_j|^2)^{1/2}}{\zeta - \lambda_j} \prod_{\mu=j+1}^{i-1} \left(\frac{1 - \bar{\lambda}_\mu \zeta}{\zeta - \lambda_\mu}\right) & if\ i > j \end{cases}$$

with respect to the Malmquist-Walsh basis. (The empty product is defined to be 1.)

To compare our new result Theorem III.2 to Theorem III.1 we note that for any $n \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, the condition $|a_{ij}| \leq b_{ij}\ \forall i, j$ implies that $\|A\|_\infty \leq \|B\|_\infty$. Suppose for instance that $|\zeta| \leq 1$. Then we can estimate the off-diagonal components of $(\zeta - M_B)^{-1}$ by

$$\left| \frac{(1-|\lambda_i|^2)^{1/2}}{1-\bar{\lambda}_i\zeta} \frac{(1-|\lambda_j|^2)^{1/2}}{1-\bar{\lambda}_j\zeta} \prod_{\mu=j}^{i} \left(\frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu} \right) \right| \leq \max_i \frac{1-|\lambda_i|^2}{|1-\bar{\lambda}_i\zeta|^2} \prod_{\mu=1}^{|m|} \left| \frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu} \right|$$

$$\leq \max_i \frac{1}{|1-\bar{\lambda}_i\zeta|} \max_i \frac{1-|\lambda_i|^2}{|1-\bar{\lambda}_i\zeta|} \prod_{\mu=1}^{|m|} \left| \frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu} \right| \leq \max_i \frac{2}{|1-\bar{\lambda}_i\zeta|} \prod_{\mu=1}^{|m|} \left| \frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu} \right|,$$

which yields the component-wise estimate

$$\left| \left((\zeta - M_B)^{-1} \right)_{ij} \right| \leq \frac{1}{\min_{\lambda_k \in \sigma(A)} |1-\bar{\lambda}_k\zeta|} \prod_{\mu=1}^{|m|} \left| \frac{1-\bar{\lambda}_\mu\zeta}{\zeta-\lambda_\mu} \right| \cdot \begin{cases} 0 & if\ i < j \\ 1 & if\ i = j \\ 2 & if\ i > j \end{cases}.$$

Corollary III.3. *Under the assumptions of Theorem III.2 suppose that $|\zeta| \leq 1$. It follows*

$$\left\| (\zeta - A)^{-1} \right\|_\infty \leq \frac{\left\| M_{|m|} \right\|_\infty}{\min_{\lambda_k \in \sigma(A)} |1-\bar{\lambda}_k\zeta|} \frac{1}{|B(\zeta)|},$$

where $B(\zeta) = \prod_{i=1}^{|m|} \frac{\zeta-\lambda_i}{1-\bar{\lambda}_i\zeta}$ is the Blaschke product associated with m.

We can pass to the general case $\sigma(A) \subset \overline{\mathbb{D}}$ by continuous extension. Setting $\zeta = 1$ Corollary III.3 is the first assertion of Theorem III.1 with the bonus that on the right hand side the norm of an $|m| \times |m|$ matrix occurs (compare [6] Section 6 B). However, if $\max_i \frac{1-|\lambda_i|^2}{|1-\lambda_i|} = \beta$ is given we have (with the same computation as above) for $\zeta = 1$

$$\left| \left((\mathbb{1} - M_B)^{-1} \right)_{ij} \right| \leq \frac{1}{\min_{\lambda_k \in \sigma(A)} |1-\lambda_k|} \cdot \begin{cases} 0 & if\ i < j \\ 1 & if\ i = j \\ \beta & if\ i > j \end{cases}$$

and we can improve the bound in Theorem III.1 if we can compute $\|M_n(\beta)\|_\infty$ (see (3)). The following theorem generalizes the discussion of Toeplitz matrices in [6]. It establishes an indirect possibility to compute $\|M_n(\beta)\|_\infty$.

Proposition III.4. *Let $M_n(\beta)$ with $\beta \in (0, 2]$ be the $n \times n$ Toeplitz matrix introduced in (3). Then the equation*

$$\beta \cot(n\theta) + (2-\beta)\cot(\theta/2) = 0, \qquad \theta \in \mathbb{R} \tag{4}$$

has a unique solution $\theta^* \in [\frac{2n-1}{2n}\pi, \pi)$ and

$$\|M_n(\beta)\|_\infty = \frac{1}{2}\sqrt{(\beta-2)^2 + \frac{\beta^2}{\cot^2(\theta^*/2)}}.$$

In particular it holds that $\|M_n(0)\|_\infty = 1$ and $\|M_n(1)\|_\infty = \frac{1}{2\sin(\frac{\pi}{4n+2})}$ and $\|M_n(2)\|_\infty = \cot(\frac{\pi}{4n})$.

It is possible to expand $\cot(n\theta)$ in Equation (4) in terms of $\cot(\theta/2)$, which yields a polynomial equation in $\cot(\theta/2)$. Since $\|M_n(\beta)\|_\infty$ only depends on $\cot(\theta/2)$ (and β) computing $\|M_n(\beta)\|_\infty$ is equivalent to finding the unique zero of the resulting polynomial in the interval $(0, \cot\left(\frac{2n-1}{4n}\pi\right)]$ as a function of β.

Finally, statement *3* of Theorem III.1 can be recovered from Theorem III.2 with the choice of a minimal polynomial $m = (z-a)^n$, $a \in (0,1)$ and setting $A_n(a) = M_B$. In this case we have for $1 \leq i, j \leq n$ that

$$\left|\left((\mathbb{1} - M_B)^{-1}\right)_{ij}\right| = \frac{1}{1-a} \cdot \begin{cases} 0 & if\ i < j \\ 1 & if\ i = j \\ 1+a & if\ i > j. \end{cases}$$

Letting $a \to 1$ proves item *3* of Theorem III.1. In the following Subsection III A we compute the entries of M_B with respect to the Malmquist-Walsh basis. This yields a simple form for matrices that achieve equality in Theorem III.2 i.e. for A with largest $\left\|(\zeta - A)^{-1}\right\|_\infty$ for a given spectrum.

Proposition III.5. *The components of the model operator M_B with respect to Malmquist-Walsh basis are given by*

$$(M_B)_{ij} = \begin{cases} 0 & if\ i < j \\ \lambda_i & if\ i = j \\ (1-|\lambda_i|^2)^{1/2}(1-|\lambda_j|^2)^{1/2}\prod_{\mu=j+1}^{i-1}\left(-\bar{\lambda}_\mu\right) & if\ i > j. \end{cases}$$

Hence, an explicit form of the matrices $A_n(a)$ in Theorem III.1 is

$$A_n(a) := \begin{pmatrix} a & 0 & \dots & \dots & 0 \\ 1-a^2 & a & \ddots & & \vdots \\ -a(1-a^2) & 1-a^2 & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ (-1)^n a^{(n-2)}(1-a^2) & \dots & -a(1-a^2) & 1-a^2 & a \end{pmatrix}.$$

Finally, we note that Theorem III.2 is stronger than Theorem III.1 in that it holds for *general* ζ and yields an *optimal* bound for *general* spectra.

The rest of this section is organized in two subsections. The first, Subsection III A, contains a proof of Theorem III.2 and Proposition III.5 while in Subsection III B we prove Proposition III.4.

A. A model operator approach to resolvent bounds

As mentioned before our approach is to bound a function of a matrix in terms of the norm of a representative function. A key role is played by Lemma II.1, which however requires that $f \in \mathcal{A}$. In order to derive upper bounds for rational functions such as the

resolvent we need to extend Lemma II.1. The following is based on the techniques of [17], Lemma 3.2 for the discussion of inverses. Here, we present an extension, which is adapted to our purposes.

Lemma III.6. *Let A be an $n \times n$ matrix with $\sigma(A) \subset \mathbb{D}$ and let ψ be a rational function with poles $(\xi_i)_{i=1,...,k}$ such that $\bigcup_i \{\xi_i\} \cap \sigma(A) = \emptyset$.*

1. *If A obeys an $\mathcal{A}$-functional calculus with constant C then $\|\psi(A)\| \leq C \inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g(\lambda_i) = \psi(\lambda_i)\ i = 1, ..., n\}$.*
2. *If $\|A\|_\infty \leq 1$ holds then $\|\psi(A)\|_\infty \leq \|\psi(M_B)\|_\infty$.*

Proof. We extend Lemma II.1 to the situation, when ψ is rational. Define $\varphi := \psi \cdot \prod_{j=1}^{k} \left(\frac{m(\xi_j)-m}{m(\xi_j)}\right)^{k_j}$, where k_j denotes the multiplicity of the pole at ξ_j and note that φ is polynomial and that $\psi(A) = \varphi(A)$. It follows using Lemma II.1 that

$$\begin{aligned}\|\psi(A)\| = \|\varphi(A)\| &\leq C \|\varphi\|_{\mathcal{A}/m\mathcal{A}} = C \inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g(\lambda_i) = \varphi(\lambda_i)\ i = 1, ..., n\} \\ &= C \inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g(\lambda_i) = \psi(\lambda_i)\ i = 1, ..., n\},\end{aligned}$$

which proves the first assertion. For the second one we consider the same φ as above and note that

$$\|\psi(A)\|_\infty = \|\varphi(A)\|_\infty \leq \|\varphi\|_{H_\infty/BH_\infty} = \|\varphi(M_B)\|_\infty ,$$

where we applied Lemma II.2 in the last step. But as $m(M_B) = 0$ it follows that $\varphi(M_B) = \psi(M_B)$. □

Let us remark that Lemma III.6 remains valid if the eigenvalue λ_i carries degeneracy k_i in m. The point here is to replace the inf on the right hand side of *1* with $\inf\{\|g\|_{\mathcal{A}} \mid g \in \mathcal{A},\ g^{(k)}(\lambda_i) = \psi^{(k)}(\lambda_i)\ , 0 \leq k < k_i\}$, where the superscript k denotes the k-th derivative.

Lemma III.7. *Let $\{\lambda_i\}_{i=1,...,n} \subset \mathbb{D}$ and let $\zeta \in \mathbb{C} - \{\lambda_i\}_{i=1,...,n}$ and $j < i$ then*

$$\sum_{\mu=j}^{i} \frac{1}{\zeta - \lambda_\mu} \frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1 - \bar{\lambda}_\nu \lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu - \lambda_\nu)} = \frac{1}{(1 - \bar{\lambda}_i\zeta)(1 - \bar{\lambda}_j\zeta)} \prod_{\mu=j}^{i} \left(\frac{1 - \bar{\lambda}_\mu\zeta}{\zeta - \lambda_\mu}\right).$$

Proof of Lemma III.7. We present two proofs for this lemma. The first one arises naturally in the context of H_2 spaces (see the proof of Theorem III.2), while the second one is more direct and simple. We define $t(z) := \frac{z}{\zeta - z} \frac{1}{(1-\bar{\lambda}_i z)(1-\bar{\lambda}_j z)}$ and the (truncated) Blaschke product $B_{ji}(z) := \prod_{\mu=j}^{i} \frac{z-\lambda_\mu}{1-\bar{\lambda}_\mu z}$ and compute the $L_2(\partial\mathbb{D})$ scalar product. Suppose for now that $|\zeta| > 1$ then

$$\begin{aligned}\langle t | B_{ji} \rangle &= \int_0^{2\pi} t(z)\overline{B_{ji}(z)}\Big|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi} = \int_0^{2\pi} t(z) \prod_\mu \frac{1 - \bar{\lambda}_\mu z}{z - \lambda_\mu}\Big|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi} \\ &= \frac{1}{2\pi i} \int_{\partial\mathbb{D}} \frac{1}{(\zeta - z)(1 - \bar{\lambda}_i z)(1 - \bar{\lambda}_j z)} \prod_\mu \frac{1 - \bar{\lambda}_\mu z}{z - \lambda_\mu} \,\mathrm{d}z = \sum_{\mu=j}^{i} \frac{1}{\zeta - \lambda_\mu} \frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1 - \bar{\lambda}_\nu \lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu - \lambda_\nu)},\end{aligned}$$

where in the last step we applied the Residue theorem and made use of the assumption $|\zeta| > 1$. On the other hand

$$\begin{aligned}\langle B_{ji}|t\rangle &= \int_0^{2\pi} B_{ji}(z)\overline{t(z)}\Big|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi} = \frac{1}{2\pi i}\int_{\partial\mathbb{D}} \prod_\mu \frac{z-\lambda_\mu}{1-\bar\lambda_\mu z}\frac{1}{\bar\zeta z - 1}\frac{z}{(z-\lambda_i)(z-\lambda_j)}\mathrm{d}z \\ &= \frac{1}{(1-\lambda_i\bar\zeta)(1-\lambda_j\bar\zeta)}\prod_\mu \frac{1-\lambda_\mu\bar\zeta}{\bar\zeta-\bar\lambda_\mu}.\end{aligned}$$

Clearly, $\langle t|B_{ji}\rangle = \overline{\langle B_{ji}|t\rangle}$ from which the lemma follows for $|\zeta| > 1$. In case that $|\zeta| < 1$ we compute similarly

$$\begin{aligned}\langle t|B_{ji}\rangle &= \frac{1}{2\pi i}\int_{\partial\mathbb{D}} \frac{1}{(\zeta-z)(1-\bar\lambda_i z)(1-\bar\lambda_j z)}\prod_\mu \frac{1-\bar\lambda_\mu z}{z-\lambda_\mu}\,\mathrm{d}z \\ &= \sum_{\mu=j}^{i}\frac{1}{\zeta-\lambda_\mu}\frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1-\bar\lambda_\nu\lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu-\lambda_\nu)} - \frac{1}{(1-\bar\lambda_i\zeta)(1-\bar\lambda_j\zeta)}\prod_\mu\frac{1-\bar\lambda_\mu\zeta}{\zeta-\lambda_\mu}\end{aligned}$$

and

$$\langle B_{ji}|t\rangle = 0.$$

The case $|\zeta| = 1$ follows by continuity. For the second proof we multiply both sides of the lemma with $\prod_{\mu=j}^{i}(\zeta-\lambda_\mu)$ to obtain a polynomial equation in ζ

$$\sum_{\mu=j}^{i}\prod_{\nu:\nu\neq\mu}(\zeta-\lambda_\nu)\frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1-\bar\lambda_\nu\lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu-\lambda_\nu)} = \prod_{\mu:\mu\neq i,\mu\neq j}(1-\bar\lambda_\mu\zeta).$$

The polynomial on the left hand side has degree at most $i-j$ and the degree of the polynomial on the right hand side is $i-j-1$. Two polynomials of a certain degree n are the same if and only if they coincide at $n+1$ nodes. We choose the $i-j+1$ values $\{\lambda_\alpha\}_{j\leq\alpha\leq i}$ and verify that for this choice equality indeed holds:

$$\begin{aligned}\sum_{\mu=j}^{i}\prod_{\nu:\nu\neq\mu}(\zeta-\lambda_\nu)\frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1-\bar\lambda_\nu\lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu-\lambda_\nu)}\Bigg|_{\zeta=\lambda_\alpha} &= \prod_{\nu:\nu\neq\alpha}(\lambda_\alpha-\lambda_\nu)\frac{\prod_{\nu:\nu\neq i,\nu\neq j}(1-\bar\lambda_\nu\lambda_\alpha)}{\prod_{\nu:\nu\neq\alpha}(\lambda_\alpha-\lambda_\nu)} \\ &= \prod_{\nu:\nu\neq i,\nu\neq j}(1-\bar\lambda_\nu\lambda_\alpha).\end{aligned}$$

□

We are now ready to present a proof of Theorem III.2.

Proof of Theorem III.2. The first assertion follows directly from Lemma III.6. To compute the matrix entries of $(\zeta - M_B)^{-1}$ with respect to Malmquist-Walsh basis we recall that

$$(\zeta - M_B)^{-1} = \varphi(M_B),$$

where $\varphi(z) := \frac{1}{\zeta - z} \frac{m(\zeta) - m(z)}{m(\zeta)}$ is a polynomial. We have that

$$((\zeta - M_B)^{-1})_{ij} = \langle \varphi(M_B) e_j | e_i \rangle = \langle P_B(\varphi\, e_j) | e_i \rangle = \langle \varphi\, e_j | e_i \rangle = \int_0^{2\pi} \varphi(z) e_j(z) \overline{e_i(z)} \Big|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi}$$
$$= \frac{((1-|\lambda_i|^2)(1-|\lambda_j|^2))^{1/2}}{2\pi i} \int_{\partial\mathbb{D}} \varphi(z) \frac{1}{(1-\bar{\lambda}_i z)(1-\bar{\lambda}_j z)} \prod_{\mu=1}^{j-1} \frac{z-\lambda_\mu}{1-\bar{\lambda}_\mu z} \prod_{\nu=1}^{i} \frac{1-\bar{\lambda}_\nu z}{z-\lambda_\nu}\, \mathrm{d}z. \quad (5)$$

In case that $j > i$ the integrand is holomorphic on $\mathbb{D}$. Hence, the integral in (5) is zero. If $j = i$ we have that

$$\frac{(1-|\lambda_i|^2)}{2\pi i} \int_{\partial\mathbb{D}} \frac{1}{\zeta - z} \frac{m(\zeta) - m(z)}{m(\zeta)} \frac{1}{(1-\bar{\lambda}_i z)(z-\lambda_i)}\, \mathrm{d}z = \frac{1}{\zeta - \lambda_i}.$$

Finally if $j < i$ then (5) becomes

$$\frac{((1-|\lambda_i|^2)(1-|\lambda_j|^2))^{1/2}}{2\pi i} \int_{\partial\mathbb{D}} \frac{1}{\zeta - z} \frac{m(\zeta) - m(z)}{m(\zeta)} \frac{1}{(1-\bar{\lambda}_i z)(1-\bar{\lambda}_j z)} \prod_{\nu=j}^{i} \frac{1-\bar{\lambda}_\nu z}{z-\lambda_\nu}\, \mathrm{d}z$$
$$= ((1-|\lambda_i|^2)(1-|\lambda_j|^2))^{1/2} \sum_{\mu=j}^{i} \frac{1}{\zeta - \lambda_\mu} \frac{\prod_{\nu:\nu\neq i, \nu\neq j}(1-\bar{\lambda}_\nu \lambda_\mu)}{\prod_{\nu:\nu\neq\mu}(\lambda_\mu - \lambda_\nu)}.$$

An application of Lemma III.7 concludes the proof of Theorem III.2. □

Proposition III.5 is verified via a direct calculation.

Proof of Proposition III.5. We proceed as in the derivation of Theorem III.2 and conclude

$$(M_B)_{ij} = ((1-|\lambda_i|^2)(1-|\lambda_j|^2))^{1/2} \int_0^{2\pi} \frac{z^2}{(1-\bar{\lambda}_i z)(1-\bar{\lambda}_j z)} \prod_{\mu=1}^{j-1} \frac{z-\lambda_\mu}{1-\bar{\lambda}_\mu z} \prod_{\nu=1}^{i} \frac{1-\bar{\lambda}_\nu z}{z-\lambda_\nu} \Bigg|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi}.$$

If $j > i$ the Residue theorem reveals that the integral is zero. Similarly, if $i = j$ the integral is given by λ_i. Finally if $i > j$ we compute

$$\int_0^{2\pi} \frac{z^2}{(1-\bar{\lambda}_i z)(1-\bar{\lambda}_j z)} \prod_{\mu=j}^{i} \frac{1-\bar{\lambda}_\mu z}{z-\lambda_\mu} \Bigg|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi} = \overline{\int_0^{2\pi} \frac{1}{(z-\lambda_i)(z-\lambda_j)} \prod_{\mu=j}^{i} \frac{z-\lambda_\mu}{1-\bar{\lambda}_\mu z} \Bigg|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi}}$$
$$= \overline{\frac{1}{2\pi i} \int_{\partial\mathbb{D}} \frac{1}{z(z-\lambda_i)(z-\lambda_j)} \prod_{\mu=j}^{i} \frac{z-\lambda_\mu}{1-\bar{\lambda}_\mu z} \mathrm{d}z} = \overline{\prod_{\mu=j+1}^{i-1} (-\lambda_\mu)},$$

where the last step again uses the Residue theorem. □

B. Computing the norm of certain Toeplitz matrices

In this subsection we prove Proposition III.4 with a direct computation of $\|M_n(\beta)\|_\infty$. Our approach is guided by the techniques developed in [7]. The quantities $\|M_n(1)\|_\infty$ and $\|M_n(2)\|_\infty$ are computed in [6] and [5] (Lemma 9.6.5) following a different approach.

Proof of Proposition III.4. Instead of working with

$$M_n(\beta) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \beta & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \beta & \dots & \beta & 1 \end{pmatrix}$$

directly, we consider the matrix

$$\tilde{M}_n(\beta) := \begin{pmatrix} \beta & \dots & \beta & 1 \\ \vdots & \iddots & 1 & 0 \\ \beta & \iddots & \iddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

and note that

$$\|M_n(\beta)\|_\infty = \left\|\tilde{M}_n(\beta)\right\|_\infty.$$

As $\tilde{M}_n(\beta)$ is Hermitian all its eigenvalues are real and its ∞-norm is simply the largest in magnitude eigenvalue. The eigenvalues of $\tilde{M}_n(\beta)^2$ are the eigenvalues of $\tilde{M}_n(\beta)$ squared. Hence, we are looking for the largest λ^2 such that

$$0 = \det\left(\tilde{M}_n(\beta)^2 - \lambda^2 \mathbb{1}\right) = \det\left(\tilde{M}_n(\beta) - \lambda\mathbb{1}\right)\left(\tilde{M}_n(\beta) + \lambda\mathbb{1}\right).$$

Direct computation reveals that

$$\left(\tilde{M}_n(\beta) - \lambda\mathbb{1}\right)\left(\tilde{M}_n(\beta) + \lambda\mathbb{1}\right) = \begin{pmatrix} \beta-\lambda & \beta & \beta & \dots & \beta & 1 \\ \beta & \beta-\lambda & & & 1 & 0 \\ \beta & & & & & 0 \\ \vdots & & & & & \vdots \\ \beta & 1 & & & -\lambda & 0 \\ 1 & 0 & 0 & \dots & 0 & -\lambda \end{pmatrix} \cdot \begin{pmatrix} \beta+\lambda & \beta & \beta & \dots & \beta & 1 \\ \beta & \beta+\lambda & & & 1 & 0 \\ \beta & & & & & 0 \\ \vdots & & & & & \vdots \\ \beta & 1 & & & \lambda & 0 \\ 1 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

$$= \begin{pmatrix} (n-1)\beta^2 - \lambda^2 + 1 & (n-2)\beta^2 + \beta & (n-3)\beta^2 + \beta & \dots & \beta^2 + \beta & \beta \\ (n-2)\beta^2 + \beta & (n-2)\beta^2 - \lambda^2 + 1 & (n-3)\beta^2 + \beta & \dots & \beta^2 + \beta & \beta \\ (n-3)\beta^2 + \beta & (n-3)\beta^2 + \beta & & & & \beta \\ \vdots & \vdots & & & & \vdots \\ \beta^2 + \beta & \beta^2 + \beta & & & \beta^2 - \lambda^2 + 1 & \beta \\ \beta & \beta & \beta & \dots & \beta & -\lambda^2 + 1 \end{pmatrix}.$$

We rearrange the resulting determinant by subtracting successively the second column from the first, the third from the second, the n-th from the $n-1$-th and leave the n-th unchanged. This yields

$$\det\left(\tilde{M}_n(\beta)^2-\lambda^2\mathbb{1}\right)=$$
$$\det\begin{pmatrix}\beta^2-\beta-\lambda^2+1 & \beta^2 & \beta^2 & \dots & \beta^2 & \beta\\ \beta+\lambda^2-1 & \beta^2-\beta-\lambda^2+1 & \beta^2 & \dots & \beta^2 & \beta\\ 0 & \beta+\lambda^2-1 & & & & \beta\\ \vdots & \vdots & & & & \vdots\\ 0 & 0 & & & \beta^2-\beta-\lambda^2+1 & \beta\\ 0 & 0 & 0 & \dots & \beta+\lambda^2-1 & -\lambda^2+1\end{pmatrix}.$$

Similarly, we subtract the second row from the first, the third from the second, the n-th from the $n-1$-th and leave the n-th unchanged. We conclude

$$\det\left(\tilde{M}_n(\beta)^2-\lambda^2\mathbb{1}\right)=$$
$$\det\begin{pmatrix}\beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1 & 0 & \dots & 0 & 0\\ \beta+\lambda^2-1 & \beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1 & 0 & \dots & 0\\ 0 & \beta+\lambda^2-1 & & & & 0\\ \vdots & 0 & & & & \vdots\\ 0 & \vdots & & & \beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1\\ 0 & 0 & 0 & \dots & \beta+\lambda^2-1 & -\lambda^2+1\end{pmatrix}=$$
$$\det\begin{pmatrix}\beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1 & 0 & \dots & 0 & 0\\ \beta+\lambda^2-1 & \beta^2-2\beta-2\lambda^2+2 & & & \dots & 0\\ 0 & \beta+\lambda^2-1 & & & & 0\\ \vdots & 0 & & & & \vdots\\ 0 & \vdots & & & \beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1\\ 0 & 0 & 0 & \dots & \beta+\lambda^2-1 & \beta^2-2\beta-2\lambda^2+2\end{pmatrix}$$
$$+\det\begin{pmatrix}\beta^2-2\beta-2\lambda^2+2 & \beta+\lambda^2-1 & 0 & \dots & 0 & 0\\ \beta+\lambda^2-1 & \beta^2-2\beta-2\lambda^2+2 & & & \dots & 0\\ 0 & \beta+\lambda^2-1 & & & & 0\\ \vdots & 0 & & & & \vdots\\ 0 & \vdots & & & \beta^2-2\beta-2\lambda^2+2 & 0\\ 0 & 0 & 0 & \dots & \beta+\lambda^2-1 & \lambda^2-(\beta-1)^2\end{pmatrix},\qquad(6)$$

where the last equality is a consequence of the linearity of det in the last column. The following is a classical formula for the determinant of an $n\times n$ tri-diagonal Toeplitz ma-

trix [7, 20]

$$\det \begin{pmatrix} x & 1 & 0 & \dots & 0 \\ 1 & x & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & x & 1 \\ 0 & \dots & 0 & 1 & x \end{pmatrix} = \frac{\sin(n+1)\theta}{\sin\theta}, \qquad x = 2\cos\theta. \tag{7}$$

To apply this result we exclude the trivial case $\beta = 0$ and note that we can always assume that $\lambda^2 \geq 1$ such that $\beta + \lambda^2 - 1 > 0$ and $\frac{\beta^2}{\beta+\lambda^2-1} \in (0, \beta]$. Hence, we can divide all columns of both determinants of (6) by $\beta + \lambda^2 - 1$. We then expand the second determinant along its last column and apply (7) to both terms resulting from (6). We find

$$\det\left(\tilde{M}_n(\beta)^2 - \lambda^2 \mathbb{1}\right) = (\beta + \lambda^2 - 1)^n \left(\frac{\sin(n+1)\theta}{\sin\theta} + \frac{\lambda^2 - (\beta-1)^2}{\lambda^2 + (\beta-1)} \frac{\sin n\theta}{\sin\theta} \right) \tag{8}$$

with

$$2\cos\theta = \frac{\beta^2 - 2\beta - 2\lambda^2 + 2}{\lambda^2 + \beta - 1} = \frac{\beta^2}{\lambda^2 + \beta - 1} - 2.$$

Solving the latter for λ^2 gives

$$\lambda^2 = \frac{1}{4}\left((\beta-2)^2 + \beta^2 \tan^2(\theta/2)\right),$$

where $\beta \neq 0$ implies that θ is such that the tangent is well defined. This enables us to eliminate λ^2 from (8) as

$$\frac{\lambda^2 - (\beta-1)^2}{\lambda^2 + (\beta-1)} = \frac{1}{\beta}(-\beta + 2 - 2\beta\cos\theta + 2\cos\theta).$$

It follows that (8) is zero if and only if

$$\begin{aligned} 0 &= \beta \frac{\sin(n+1)\theta}{\sin\theta} + (-\beta + 2 - 2\beta\cos\theta + 2\cos\theta) \frac{\sin n\theta}{\sin\theta} \\ &= \beta \cos n\theta + (2-\beta)(1+\cos\theta) \frac{\sin n\theta}{\sin\theta}, \end{aligned}$$

which in turn is equivalent to

$$\cot n\theta = \frac{\beta - 2}{\beta} \cot(\theta/2). \tag{9}$$

In total, we are looking for the solution θ^* of (9) such that λ^2 is maximal i.e. $\cot^2(\theta^*/2)$ is minimal. Since for any $\theta \in [\frac{2n-1}{2n}\pi, \pi)$ we have $\frac{\beta-2}{\beta}\cot(\theta/2) \leq 0$ with $\cot(\pi/2) = 0$ and $\lim_{\theta\uparrow\pi} \cot n\theta \to -\infty$ and $\cot\frac{2n-1}{2}\pi = 0$, it follows that there is a unique solution

$\theta^* \in [\frac{2n-1}{2n}\pi, \pi)$ of Equation (9). Moreover, by the same fact, $\cot(\pi/2) = 0$, this solution maximizes λ^2 as desired.

Sometimes it is possible to obtain a solution for Equation (9) in closed form. Suppose $\beta = 2$, then $\cot n\theta^* = 0$ and $\theta^* = \frac{2n-1}{2n}\pi$. It follows

$$\|M_n(2)\|_\infty^2 = \tan^2\left(\frac{2n-1}{4n}\pi\right) = \cot^2(\pi/4n)$$

as in [6]. If $\beta = 1$ we have

$$\lambda^2 = \frac{1}{4\cos^2(\theta/2)}$$

and

$$\sin(2n+1)\theta/2 = 0$$

such that $\theta^* = \frac{2n\pi}{2n+1}$. It follows

$$\|M_n(1)\|_\infty^2 = \frac{1}{4\cos^2(\frac{n\pi}{2n+1})} = \frac{1}{4\sin^2(\frac{\pi}{4n+2})}$$

as in [7]. The trivial fact $\|M_n(0)\|_\infty = 1$ can be recovered by continuous extension as $\beta \to 0$. □

IV. POWER-BOUNDED OPERATORS

It is natural to ask if power-boundedness of A is sufficient to obtain estimates on $\|(\zeta - A)^{-1}\|$ qualitatively similar to the results of Theorem III.1, III.2 and Corollary III.3. In this section we prove that this is indeed the case and present a new bound on the norm of the resolvent of a power-bounded operator.

Theorem IV.1. *Let A be an $n \times n$ matrix with minimal polynomial m of degree $|m|$ and let $\|\cdot\|$ be an arbitrary matrix norm with $\sup_{k\geq 0}\|A^k\| = C < \infty$. For any $\zeta \in \overline{\mathbb{D}} - \sigma(A)$ it holds that*

$$\|(\zeta - A)^{-1}\| \leq \frac{2|m|C}{\min_{\lambda_i\in\sigma(A)}|1-\bar{\zeta}\lambda_i|^{1/2}(2|m| - 2|m||\zeta|^2 + |\zeta|^2\min_{\lambda_i\in\sigma(A)}|1-\bar{\zeta}\lambda_i|)^{1/2}}\left(\frac{4e}{|B(\zeta)|^2} - 1\right)^{1/2},$$

where $B(\zeta) = \prod_{i=1}^{|m|}\frac{\zeta-\lambda_i}{1-\bar{\lambda}_i\zeta}$ is the Blaschke product associated with m. For $|\zeta| > 1$, we have the obvious estimate $\|(\zeta - A)^{-1}\| \leq \frac{C}{|\zeta|-1}$.

Theorem IV.1 is the analogue of Corollary III.3 for power-bounded operators. Spectral bounds on the norm of the resolvent of a power-bounded operator are well studied in the literature. Theorem 6.4 of [6] treats the same problem in the special case that A is

power-bounded with respect to operator norm $\|\cdot\|_\infty$. In [17, Theorem 3.24] the behavior of $\left\|(\zeta - A)^{-1}\right\|$ is studied for $|\zeta| < 1$ and in [29] an upper bound is derived for $|\zeta| \geq 1$. Theorem IV.1 unifies the mentioned results and yields a quantitatively better bound in each case. To compare suppose that $|\zeta| < 1$ and note that in this case

$$1 - |\zeta|^2 + \frac{1}{2|m|}|\zeta|^2 \min_{\lambda_i \in \sigma(A)} |1 - \bar{\zeta}\lambda_i| \geq (1 - |\zeta|)^2$$

and of course $\min_{\lambda_i \in \sigma(A)} |1 - \bar{\zeta}\lambda_i| \geq 1 - |\zeta|$. Hence, it follows

$$\left\|(\zeta - A)^{-1}\right\| \leq \frac{\sqrt{8e|m|}C}{(1 - |\zeta|)^{3/2}} \frac{1}{|B(\zeta)|},$$

which is qualitatively the same as Theorem 3.24 in [17] but has a better numerical prefactor. If we choose $|\zeta| = 1$ it follows $|B(\zeta)| = 1$ and therefore

$$\left\|(\zeta - A)^{-1}\right\| \leq \frac{\sqrt{16e - 4}\,|m|C}{\min_{\lambda_i \in \sigma(A)} |\zeta - \lambda_i|}. \tag{10}$$

This bound improves on the result in [29] (which in turn is stronger than [6, Theorem 6.4]) as the new bound only grows linearly with $|m|$ as opposed to $|m|^{3/2}$ in [29]. That for power-bounded $A \in \mathcal{M}_n$ the correct asymptotic growth order for an upper bound is $\mathcal{O}(n)$ was already suspected in [6] and [31]. The bound obtained almost reaches the optimal estimate of Theorem III.1 for Hilbert-space contractions. In the latter case we have that $\cot(\frac{\pi}{4n})/n \leq \frac{4}{\pi}$, while the prefactor of (10) is $\sqrt{16e - 4} \approx 6.28$. However, as is clear from the derivation, Inequality (10) is not optimal. We will use Inequality (10) to study the sensitivity of a classical or quantum Markov chain to perturbations in Section V.

To prove Theorem IV.1 we take a similar approach as to Theorem III.2. We note that power-bounded operators admit a Wiener algebra functional calculus. Thus an application of Lemma III.6 reveals that

$$\left\|(\zeta - A)^{-1}\right\| \leq C \inf\{\|g\|_W \mid g \in W,\ g(\lambda_i) = \frac{1}{\zeta - \lambda_i}\}. \tag{11}$$

The strategy of our proof will be to consider one specific representative function g in (11) and to bound its norm. To achieve this we employ the following method. Instead of considering g directly we choose a "smoothing parameter" r and pass to a "stretched" interpolation function.
Given any function $f \in H_2$ and $r \in (0, 1)$, we write $f_r(z) := f(rz) = \sum_{k \geq 0} \hat{f}(k) r^k z^k$ and observe that by the Cauchy-Schwarz inequality and the Plancherel identity

$$\|f_r\|_W \leq \sqrt{\sum_{k \geq 0} |\hat{f}(k)|^2} \sqrt{\frac{1}{1 - r^2}} = \|f\|_{H_2} \sqrt{\frac{1}{1 - r^2}}. \tag{12}$$

This idea was used to obtain bounds on the inverse and resolvent of a power-bounded operator in [17] and to study spectral convergence bounds for bounded semigroups in [27]. We use the Blaschke products $B(z) = \prod_i \frac{z - \lambda_i}{1 - \lambda_i z}$ and $\tilde{B}(z) = \prod_i \frac{z - r\lambda_i}{1 - r\lambda_i z}$, where in the latter

product the spectrum is stretched by a factor of r. (The products are taken over all prime factors of m, but to avoid cumbersome notation we do not write this explicitly.) Consider now the function g with

$$g(z) = \sum_k \left(\frac{1}{\zeta - \lambda_k} \frac{\prod_j (1 - \bar{\lambda}_j \lambda_k)}{\prod_{j \neq k} (\lambda_k - \lambda_j)} \right) \frac{B(z)}{z - \lambda_k}.$$

Note that g is analytic in the unit disc and $g(\lambda_i) = \frac{1}{\zeta - \lambda_i}$ for all $\lambda_i \in \sigma(A)$. In order to use the estimate (12) we perform the aforementioned smoothing. We define the modified function $\tilde{g}$ by

$$\tilde{g}(z) = \sum_k \left(\frac{1}{\zeta - \lambda_k} \frac{\prod_j (1 - r^2 \bar{\lambda}_j \lambda_k)}{\prod_{j \neq k} (r\lambda_k - r\lambda_j)} \right) \frac{\tilde{B}(z)}{z - r\lambda_k}$$

and observe that $\tilde{g}_r$ enjoys the same basic properties as g i.e. $\tilde{g}_r$ is analytic in $\mathbb{D}$ and $\tilde{g}_r(\lambda_i) = \frac{1}{\zeta - \lambda_i}$ for any $\lambda_i \in \sigma(A)$. Thus, by Inequality (11), we have that $\left\| (\zeta - A)^{-1} \right\| \leq C \left\| \tilde{g}_r \right\|_W$ and it follows from Inequality (12) that

$$\left\| \tilde{g}_r \right\|_W \leq \sqrt{\frac{1}{1 - r^2}} \left\| \tilde{g} \right\|_{H_2}.$$

It turns out that one can directly compute $\left\| \tilde{g} \right\|_{H_2}$. The computation relies on two combinatorial observations similar to Lemma III.7, which we shall prove before we proceed with our discussion of $\left\| \tilde{g} \right\|_{H_2}$.

Lemma IV.2. *Let $|m| \in \mathbb{N} - \{0\}$ and $\{\lambda_i\}_{i=1,\ldots,|m|} \subset \mathbb{D}$. Furthermore, let $\zeta \in \mathbb{C} - \{\lambda_i\}_{i=1,\ldots,|m|}$ and $r \in (0,1)$. It follows that*

1. $$\sum_{i=1}^{|m|} \frac{1}{\zeta - \lambda_i} \frac{\prod_{j:j \neq l} (1 - r^2 \bar{\lambda}_j \lambda_i)}{\prod_{j:j \neq i} (r\lambda_i - r\lambda_j)} = \frac{r}{1 - r^2 \bar{\lambda}_l \zeta} \prod_{i=1}^{|m|} \frac{1 - r^2 \bar{\lambda}_i \zeta}{r\zeta - r\lambda_i},$$

2. $$\sum_{i=1}^{|m|} \frac{1}{\zeta - \lambda_i} \frac{1}{1 - r^2 \bar{\zeta} \lambda_i} \frac{\prod_j (1 - r^2 \bar{\lambda}_j \lambda_i)}{\prod_{j:j \neq i} (r\lambda_i - r\lambda_j)} = \frac{r}{1 - r^2 |\zeta|^2} \left(\prod_{i=1}^{|m|} \frac{1 - r^2 \bar{\lambda}_i \zeta}{r\zeta - r\lambda_i} - \prod_{i=1}^{|m|} \frac{r\bar{\zeta} - r\bar{\lambda}_i}{1 - r^2 \lambda_i \bar{\zeta}} \right),$$

3. $$\left\| \tilde{g} \right\|_{H_2}^2 = \frac{r^2}{1 - r^2 |\zeta|^2} \left(\prod_{i=1}^{|m|} \left| \frac{1 - r^2 \bar{\lambda}_i \zeta}{r\zeta - r\lambda_i} \right|^2 - 1 \right).$$

Our proof is based on the Residue theorem. (It is also possible to prove the lemma with the second technique outlined in the proof of Lemma III.7. However, the approach via the Residue theorem is more convenient for the second assertion.)

Proof. For the first assertion set $t_1(z) := \frac{rz}{r\zeta - z}\frac{1}{1-r\bar{\lambda}_l z}$ and suppose for now that $r|\zeta| < 1$. We have that

$$\begin{aligned}\langle t_1|\tilde{B}\rangle &= \int_0^{2\pi} \frac{rz}{r\zeta - z}\frac{1}{1-r\bar{\lambda}_l z}\prod_i \frac{1-r\bar{\lambda}_i z}{z-r\lambda_i}\Big|_{z=e^{i\phi}}\frac{\mathrm{d}\phi}{2\pi}\\ &= \frac{1}{2\pi i}\int_{\partial\mathbb{D}} \frac{r}{r\zeta - z}\frac{1}{1-r\bar{\lambda}_l z}\prod_i \frac{1-r\bar{\lambda}_i z}{z-r\lambda_i}\mathrm{d}z\\ &= \sum_i \frac{1}{\zeta-\lambda_i}\frac{\prod_{j:j\neq l}(1-r^2\bar{\lambda}_j\lambda_i)}{\prod_{j:j\neq i}(r\lambda_i - r\lambda_j)} - \frac{r}{1-r^2\bar{\lambda}_l\zeta}\prod_i \frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta - r\lambda_i}\end{aligned}$$

and that

$$\begin{aligned}\langle \tilde{B}|t_1\rangle &= \int_0^{2\pi}\prod_i \frac{z-r\lambda_i}{1-r\bar{\lambda}_i z}\frac{r}{r\bar{\zeta}z-1}\frac{z}{z-r\lambda_l}\Big|_{z=e^{i\phi}}\frac{\mathrm{d}\phi}{2\pi}\\ &= \frac{1}{2\pi i}\int_{\partial\mathbb{D}}\prod_i \frac{z-r\lambda_i}{1-r\bar{\lambda}_i z}\frac{r}{r\bar{\zeta}z-1}\frac{1}{z-r\lambda_l}\,\mathrm{d}z = 0.\end{aligned}$$

Hence, for $r|\zeta| < 1$

$$\sum_i \frac{1}{\zeta-\lambda_i}\frac{\prod_{j:j\neq l}(1-r^2\bar{\lambda}_j\lambda_i)}{\prod_{j:j\neq i}(r\lambda_i - r\lambda_j)} = \frac{r}{1-r^2\bar{\lambda}_l\zeta}\prod_i \frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta - r\lambda_i}$$

as claimed. An identical computation reveals that the above remains correct if $r|\zeta| > 1$ and the case $r|\zeta| = 1$ follows by continuity. For the second assertion suppose again that $r|\zeta| < 1$ and set $t_2(z) := \frac{rz}{r\zeta - z}\frac{1}{1-r\bar{\zeta}z}$ and compute

$$\begin{aligned}\langle t_2|\tilde{B}\rangle &= \int_0^{2\pi} \frac{rz}{r\zeta - z}\frac{1}{1-r\bar{\zeta}z}\prod_i \frac{1-r\bar{\lambda}_i z}{z-r\lambda_i}\Big|_{z=e^{i\phi}}\frac{\mathrm{d}\phi}{2\pi}\\ &= \frac{1}{2\pi i}\int_{\partial\mathbb{D}} \frac{r}{r\zeta - z}\frac{1}{1-r\bar{\zeta}z}\prod_i \frac{1-r\bar{\lambda}_i z}{z-r\lambda_i}\,\mathrm{d}z\\ &= \sum_i \frac{1}{\zeta-\lambda_i}\frac{1}{1-r^2\bar{\zeta}\lambda_i}\frac{\prod_j(1-r^2\bar{\lambda}_j\lambda_i)}{\prod_{j:j\neq i}(r\lambda_i - r\lambda_j)} - \frac{r}{1-r^2|\zeta|^2}\prod_i \frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta - r\lambda_i}.\end{aligned}$$

Similarly,

$$\langle \tilde{B}|t_2\rangle = \frac{1}{2\pi i}\int_{\partial\mathbb{D}}\prod_i \frac{z-r\lambda_i}{1-r\bar{\lambda}_i z}\frac{r}{r\bar{\zeta}z-1}\frac{1}{z-r\zeta}\,\mathrm{d}z = \frac{r}{r^2|\zeta|^2-1}\prod_i \frac{r\zeta - r\lambda_i}{1-r^2\bar{\lambda}_i\zeta}.$$

It follows that

$$\sum_i \frac{1}{\zeta-\lambda_i}\frac{1}{1-r^2\bar{\zeta}\lambda_i}\frac{\prod_j(1-r^2\bar{\lambda}_j\lambda_i)}{\prod_{j:j\neq i}(r\lambda_i - r\lambda_j)} = \frac{r}{1-r^2|\zeta|^2}\left(\prod_i \frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta - r\lambda_i} - \prod_i \frac{r\bar{\zeta}-r\bar{\lambda}_i}{1-r^2\lambda_i\bar{\zeta}}\right).$$

The same computations prove the validity of this statement for $r|\zeta| > 1$. One can make sense of the formula in case that $r|\zeta| = 1$ by continuous extension. Using these observations one can compute

$$\begin{aligned}
\|\tilde{g}\|_{H_2}^2 &= \int_0^{2\pi} \tilde{g}(z)\overline{\tilde{g}(z)}\Big|_{z=e^{i\phi}} \frac{\mathrm{d}\phi}{2\pi} \\
&= \frac{1}{2\pi i}\sum_{k,l} \frac{1}{\zeta-\lambda_k}\frac{\prod_\mu(1-r^2\bar{\lambda}_\mu\lambda_k)}{\prod_{\mu\neq k}(r\lambda_k - r\lambda_\mu)}\overline{\frac{1}{\zeta-\lambda_l}\frac{\prod_\nu(1-r^2\bar{\lambda}_\nu\lambda_l)}{\prod_{\nu\neq l}(r\lambda_l - r\lambda_\nu)}}\int_{\partial\mathbb{D}}\frac{1}{z-r\lambda_k}\frac{1}{1-r\bar{\lambda}_l z}\mathrm{d}z \\
&= \sum_l \left(\overline{\frac{1}{\zeta-\lambda_l}\frac{\prod_\nu(1-r^2\bar{\lambda}_\nu\lambda_l)}{\prod_{\nu\neq l}(r\lambda_l - r\lambda_\nu)}}\sum_k\left(\frac{1}{\zeta-\lambda_k}\frac{1}{1-r^2\bar{\lambda}_l\lambda_k}\frac{\prod_\mu(1-r^2\bar{\lambda}_\mu\lambda_k)}{\prod_{\mu\neq k}(r\lambda_k - r\lambda_\mu)}\right)\right) \\
&= \prod_i \frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta - r\lambda_i}\overline{\left(\sum_l \frac{1}{\zeta-\lambda_l}\frac{r}{1-r^2\lambda_l\bar{\zeta}}\frac{\prod_\nu(1-r^2\bar{\lambda}_\nu\lambda_l)}{\prod_{\nu\neq l}(r\lambda_l - r\lambda_\nu)}\right)} \qquad (13) \\
&= \frac{r^2}{1-r^2|\zeta|^2}\prod_i\frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta-r\lambda_i}\overline{\left(\prod_i\frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta-r\lambda_i} - \prod_i\frac{r\bar{\zeta}-r\bar{\lambda}_i}{1-r^2\lambda_i\bar{\zeta}}\right)} \qquad (14) \\
&= \frac{r^2}{1-r^2|\zeta|^2}\left(\prod_i\left|\frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta-r\lambda_i}\right|^2 - 1\right),
\end{aligned}$$

where we used the first assertion of the lemma for (13) and the second assertion for (14). Note that for all $\zeta \in \mathbb{C} - \sigma(A)$ and $r \in (0,1)$ the final quantity is real and positive. □

With this preparatory work done a proof of Theorem IV.1 is simple.

Proof of Theorem IV.1. We assume that $\sigma(A) \subset \mathbb{D}$. From Equations (11), (12) and Lemma IV.2 we have that for any $\zeta \in \mathbb{C} - \sigma(A)$

$$\left\|(\zeta - A)^{-1}\right\| \leq C\sqrt{\frac{1}{1-r^2}}\ \|\tilde{g}\|_{H_2} = C\sqrt{\frac{1}{1-r^2}}\sqrt{\frac{r^2}{1-r^2|\zeta|^2}\left(\prod_{i=1}^{|m|}\left|\frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta-r\lambda_i}\right|^2 - 1\right)}. \qquad (15)$$

Clearly,

$$\prod_{i=1}^{|m|}\left|\frac{1-r^2\bar{\lambda}_i\zeta}{r\zeta-r\lambda_i}\right|^2 = \frac{1}{r^{2|m|}}\ \frac{1}{|B(\zeta)|^2}\prod_{i=1}^{|m|}\left|1+\bar{\lambda}_i\zeta\,\frac{1-r^2}{1-\bar{\lambda}_i\zeta}\right|^2.$$

To obtain an upper bound we assume that $\zeta \in \overline{\mathbb{D}} - \sigma(A)$ and choose $r \in (0,1)$ such that

$$1 - r^2 = \frac{\min_i|1-\bar{\zeta}\lambda_i|}{2|m|}.$$

It follows that

$$\prod_{i=1}^{|m|}\left|1+\bar{\lambda}_i\zeta\,\frac{1-r^2}{1-\bar{\lambda}_i\zeta}\right|^2 \leq \left(1+\frac{1}{2|m|}\right)^{2|m|} \leq e$$

and that (for $|m| \geq 2$)

$$r^{2|m|} = \left(1 - \frac{\min_i |1 - \bar{\zeta}\lambda_i|}{2|m|}\right)^{|m|} \geq (1 - 1/2)^2 = 1/4.$$

We conclude that

$$\left\|(\zeta - A)^{-1}\right\| \leq C \left(\frac{2|m|}{\min_i |1 - \bar{\zeta}\lambda_i|}\right)^{1/2} \left(\frac{1}{1 - |\zeta|^2 \left(1 - \frac{\min_i |1 - \bar{\zeta}\lambda_i|}{2|m|}\right)}\right)^{1/2} \left(\frac{4e}{|B(\zeta)|^2} - 1\right)^{1/2},$$

which is claimed in the theorem. As always the general case $\sigma(A) \subset \overline{\mathbb{D}}$ follows by continuous extension. Finally, we note that for $|\zeta| > 1$ one can choose $r = \sqrt{\frac{1}{|\zeta|}}$ in (15) and recover the obvious estimate

$$\left\|(\zeta - A)^{-1}\right\| \leq C \frac{1}{|\zeta| - 1} \left(1 - \prod_{i=1}^{|m|} \left|\frac{|\zeta| - \bar{\lambda}_i \zeta}{\sqrt{|\zeta|}\zeta - \sqrt{|\zeta|}\lambda_i}\right|^2\right)^{1/2} \leq C \frac{1}{|\zeta| - 1}.$$

□

V. STABILITY OF MARKOV CHAINS

If T is a classical stochastic matrix or a quantum channel (a trace-preserving and completely positive map, see [15]) the sequence $\{T^n\}_{n \geq 0}$ can be regarded as a finite and homogenous classical or quantum Markov chain with transition map T. In this section we apply Theorem IV.1 to study the stability of the stationary states of a Markov chain to perturbations in the transition map. A core observation is that the transition matrix of the Markov chain is power-bounded with respect to the 1-to-1 norm and constant 1, i.e. the Markov chain constitutes a bounded semigroup, see [27]. A similar approach based on this observation was taken in [26] to prove spectral convergence estimates for classical and quantum Markov chains. We begin by recalling the basic framework of sensitivity analysis of Markov chains. A detailed introduction, however, is beyond the scope of this article. We refer to [4] and the references therein for an overview of the existing perturbation bounds for classical Markov chains and to [27] for an introduction to the stability theory of quantum Markov chains.

Let T, $\tilde{T}$ denote two classical stochastic matrices or two quantum channels. The inequality

$$\|\rho - \tilde{\rho}\| \leq \kappa \left\|T - \tilde{T}\right\|$$

relates the distance between two stationary states ρ and $\tilde{\rho}$ arising from T and $\tilde{T}$, $\rho = T(\rho)$, $\tilde{\rho} = \tilde{T}(\tilde{\rho})$, to the distance between T and $\tilde{T}$. Commonly T is considered to be the transition matrix of the Markov chain of interest while $\tilde{T}$ is a small perturbation thereof. The *condition number* κ measures the relative sensitivity of the stationary states

to perturbations. If T has a unique stationary state the above inequality quantifies the stability of the asymptotic behavior of $\{T^n\}_{n\geq 0}$ with respect to perturbations in the transition matrix. Elementary linear algebra shows that if T has a unique stationary state one can choose (see [23–25]) the condition number

$$\kappa_{cl} = \sup_{\substack{\delta\in\mathbb{R}^n \\ (1,\ldots,1)\cdot\delta=0}} \frac{\|Z(\delta)\|_1}{\|\delta\|_1}, \quad Z := (\mathbb{1} - T + T^\infty)^{-1}$$

in the classical case and similarly (see [27])

$$\kappa_{qu} = \sup_{\substack{\sigma=\sigma^\dagger\in\mathcal{M}_n \\ \mathrm{tr}(\sigma)=0}} \frac{\|Z(\sigma)\|_1}{\|\sigma\|_1}, \quad Z := (\mathbb{1} - T + T^\infty)^{-1}$$

in the quantum setup. Here, T^∞ denotes the projection onto the stationary state of T and $\|\cdot\|_1$ denotes the absolute entry sum in the classical and the Schatten 1-norm in the quantum case. In either case the spectral properties of T and T^∞ guarantee that the map Z exists.

If the transition matrix has a unique stationary state and a subdominant eigenvalue of this matrix is close to 1 it is clear that the chain is ill conditioned in the sense that the stationary state is sensitive to perturbations in the transition map. It is a well-studied question (see [10–12, 25, 27]) whether the reverse conclusion also holds, i.e. whether the closeness of the sub-dominant eigenvalues of T to 1 provides complete information about the sensitivity of $\{T^n\}_{n\geq 0}$. It was established that this is indeed the case by deriving spectral lower and upper bounds for certain choices of κ. In particular, as shown in [25] it holds that

$$\frac{1}{\min_{\lambda_i\in\sigma(T-T^\infty)}|1-\lambda_i|} \leq \kappa_{cl} \leq \frac{n}{\min_{\lambda_i\in\sigma(T-T^\infty)}|1-\lambda_i|}.$$

A similar quantum bound occurs in [27].

The techniques developed in this article yield a direct approach to spectral stability estimates in both the classical and quantum case. The core observation is that if T is a stochastic matrix or a quantum channel the map $T - T^\infty$ is power bounded with (see [26] Lemma III.1)

$$\|(T-T^\infty)^n\|_{1\to 1} = \|T^n - (T^\infty)^n\|_{1\to 1} \leq \|T^n\|_{1\to 1} + \|(T^\infty)^n\|_{1\to 1} \leq 2,$$

where $\|A\|_{1\to 1} = \sup_v \frac{\|Av\|_1}{\|v\|_1}$. With an application of Inequality (10) we conclude that

$$\kappa_{cl} \leq \|Z\|_{1\to 1} \leq \frac{2\sqrt{16e-4}n}{\min_{\lambda_i\in\sigma(T-T^\infty)}|1-\lambda_i|},$$

which is qualitatively the same as the estimate in [25] but has a worse numerical prefactor ($2\sqrt{16e-4}$ instead of 1). However, the bound in [25] uses the additional properties a classical stochastic matrix has as well as the fact that the supremum in the definition of

κ_{cl} is taken over vectors with 0 column sum. Our bound proves that in this case power-boundedness alone is sufficient and the additional assumptions on T and κ_{cl} are basically superfluous. Other spectral stability estimates for classical Markov chains such as [10] are weaker than (10). In the quantum context we can use Inequality (10) to improve on the spectral stability estimates of [27].

Theorem V.1. *Let T be a trace-preserving, positive linear map on $\mathcal{M}_n$ and $\Lambda := \sigma(T) - \{1\}$ the set of its non-unit eigenvalues. Then*

$$\frac{1}{\min_{\lambda_i \in \Lambda} |1 - \lambda_i|} \leq \kappa_{qu} \leq \frac{2\sqrt{16e-4}n^2}{\min_{\lambda_i \in \Lambda} |1 - \lambda_i|}.$$

The proof of the theorem is identical as in [27] up to an application of (10) instead of the theorem in [29].

Acknowledgments

OS acknowledges financial support by the Elite Network of Bavaria (ENB) project QCCC and the CHIST-ERA/BMBF project CQC. OS is thankful to Michael M. Wolf for creating conditions that made this work possible and to Alexander Müller-Hermes for proofreading the manuscript and for pointing out the simpler proof for Lemma III.7. OS is equally thankful to E. B. Davies for valuable comments on a previous version of the manuscript.

---

[1] J. Agler and J. McCarthy. *Pick Interpolation and Hilbert function spaces.* AMS: Graduate studies in mathematics: 44, 2002.
[2] O. Bandtlow. Estimates for norms and resolvents and an application to the perturbation of spectra. *Math. Nachr.*, 267:3–11, 2004.
[3] R. Bhatia. *Matrix Analysis.* Springer, 1996.
[4] E. Cho and C. Meyer. Comparison of perturbations bounds for the stationary distributions of a Markov chain. *Lin. Alg. and Appl.*, 335:1:137–150, 2001.
[5] E. Davies. *Linear operators and their spectra.* Cambridge Studies in Advanced Mathematics, 2007.
[6] E. Davies and B. Simon. Eigenvalue estimates for non-normal matrices and the zeros of random orthogonal polynomials on the unit circle. *J. of Approx. Theory*, 141:2:189–213, 2006.
[7] E. Egerváry. Über gewisse Extremumprobleme der Funktionentheorie. *Mathematische Annalen*, 99:542–561, 1928.
[8] F. Foiaş and A.E. Frazho. The commutant lifting approach to interpolation problems. Birkhäuser, Basel, 1990.
[9] T. Kato. *Perturbation Theory for Linear Operators.* Springer, Berlin, 1995.
[10] C. Meyer. Sensitivity of the stationary distribution of a Markov chain. *SIAM J. Mat. Anal. Appl.*, 15:3:715–728, 1994.
[11] A. Mitrophanov. Stability and exponential convergence of continuous-time Markov chains. *J. Appl. Prob.*, 40:970–979, 2003.

[12] A. Mitrophanov. Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Prob.*, 42:1003–1014, 2005.
[13] S. B. Nagy, C. Foiaş. Commutants de certains opérateurs. *Act. Sci. Math.*, 29:1–17, 1968.
[14] S. B. Nagy, C. Foiaş, H. Bercovici, and L. Kerchy. *Harmonic analysis of operators on Hilbert spaces.* Springer, 2010.
[15] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.
[16] N. Nikolski. *Treatise on the Shift Operator.* Springer: Grundlehren der mathematischen Wissenschaft, 1986.
[17] N. Nikolski. Condition numbers of large matrices and analytic capacities. *St. Petersburg Math. J.*, 17:641–682, 2006.
[18] N. Nikolski. *Operators, Functions and Systems: An Easy Reading.* AMS: Mathematical Surveys and Monographs: 93, 2009.
[19] Henrici P. Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. *Numerische Mathematik*, 4:24–40, 1962.
[20] E. Pascal. *Determinanten.* Teubners Sammlung 3, 1900.
[21] D. Phillips. Improving Spectral-Variation bounds with Chebyshef Polynomials. *Lin. Alg. and Appl.*, 133:165–173, 1990.
[22] D. Sarason. Generalized interpolation in H^∞ . *Trans. Amer. Math. Soc.*, 127:289–299, 1967.
[23] P. Schweitzer. Perturbation theory and finite Markov chains. *J. Appl. Prob.*, 5:401–413, 1968.
[24] E. Seneta. Perturbation of the stationary distribution measured by ergodicity coefficients. *Adv. Appl. Prob.*, 20:228–230, 1988.
[25] E. Seneta. Sensitivity of finite Markov chains under perturbation. *Stat. and Prob. let.*, 17.2:163–168, 1993.
[26] O. Szehr, D. Reeb, and M. Wolf. Spectral convergence bounds for classical and quantum Markov processes. 2013. arXiv: 1301.4827.
[27] O. Szehr and M. Wolf. Perturbation bounds for quantum Markov processes and their fixed points . 2011. *J. Math. Phys.*, 54: 032203, 2013.
[28] H. Wielandt. Inclusion theorems for eigenvalues. *Nat. Bur. Standards Appl. Math.*, 29:75–78, 1951.
[29] R. Zarouf. Une amélioration d'un résultat de E.B. Davies et B. Simon. *Comptes Rendus Mathematique*, 347:15:939–942, 2009. arXiv: 0903.2743v2.
[30] R. Zarouf. Effective H^∞ interpolation. 2011. arXiv: 0905.0573.
[31] R. Zarouf. Iterated resolvent estimates for power bounded matrices. 2011. arXiv:1103.5019v1.

Perturbation Theory for Parent Hamiltonians of Matrix Product States

O. Szehr and M. Wolf December 17, 2013

The parent Hamiltonian model is one of the most important models in rigorous solid-state theory. Almost all Hamiltonian descriptions of quantum spin chains studied in the literature fall into this class (see for example [2, 1, 3]). This article investigates the stability of the ground state of a parent Hamiltonian of a generic Matrix product state against local perturbations. We prove that the spectral gap of a parent Hamiltonian remains stable under weak local perturbations even in the thermodynamic limit, where the entire perturbation is not bounded. This is important for the following reasons: First, the contrary prediction that arbitrarily small perturbations of a physical system lead to a phase transition of the system would be in conflict with physical observation. This would be a strong point against the reasonableness of the model. Second, as any model, the parent Hamiltonian model is an idealization. In the sense of the description of a physical system it is more natural to consider "near-by" models as descriptions of a messy experimental reality and study their sensitivity with respect to local interactions.

1 Stability under weak local perturbation

Let $|\Psi\rangle$ be a translationally invariant MPS with periodic boundary conditions on a chain Λ. We suppose that $|\Psi\rangle$ is generic (in the sense of condition (G1)). The canonical parent Hamiltonian H_Λ is frustration-free and has a positive spectral gap above the ground state energy. Due to the aforementioned reasons a key question is what happens to the ground state of H_Λ if the system is perturbed with small local interaction terms. We study perturbations of the form $\Phi_\Lambda = \sum_{x\in\Lambda} \phi_x$ that are weak in the sense that

$$\|\phi_x\|_\infty \leq \beta$$

holds for sufficiently small β. Observe that although each local term is small, in the limit of a large Λ the entire perturbation Φ_Λ needs not remain bounded. We present a heuristic statement of our theorem in this overview. A rigorous formulation is contained in the attached article.

Theorem 1. *Let $|\Psi\rangle$ be a* generic *MPS, which is translationally invariant and has periodic boundary conditions. For sufficiently weak local perturbations Φ_Λ the following conclusions hold for the perturbed parent Hamiltonian $\widetilde{H}_\Lambda = H_\Lambda + \Phi_\Lambda$.*

1. *$\widetilde{H}_\Lambda$ has a non-degenerate ground state with positive spectral gap.*
2. *There exists a thermodynamic limit of the ground state as the length of the chain grows to infinity.*

3. *There is exponential decay of correlations in the infinite volume ground state ω: If A_1, A_2 are observables that are supported on disjoint subsets Λ_1, Λ_2 of the chain then*

$$|\omega(A_1 A_2) - \omega(A_1)\omega(A_2)| \leq C\epsilon^{\mathrm{dist}(\Lambda_1,\Lambda_2)}$$

for some constant C and $\epsilon < 1$.

4. *If in a certain parameter range the local perturbations change smoothly, then the infinite volume ground state changes smoothly.*

Our theorem generalizes known stability results for particular Hamiltonians, such as classical models or the AKLT model [3] as these Hamiltonians can be interpreted as parent Hamiltonians.

2 Legal statement

The project was assigned by M. Wolf. The elaboration of the topic and the derivation of stability estimates is the work of the first author.

References

[1] I. Affleck, T. Kennedy, E. Lieb, and H. Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.*, 59(7):799–802, 1987.

[2] D. Perez-Garcia, F. Verstaete, M.M. Wolf, and J.I. Cirac. Matrix product state representations. *Quantum Inf. Comput.*, 7:401–430, 2007.

[3] D. A. Yarotsky. Ground states in relatively bounded quantum perturbations of classical lattice systems. *Comm. in Math. Phys.*, 261(3):799–819, 2004.

Perturbation Theory for Parent Hamiltonians of Matrix Product States

Oleg Szehr[1,*] and Michael M. Wolf[1,†]

[1] *Zentrum Mathematik, Technische Universität München, 85748 Garching, Germany*

(Dated: February 24, 2014)

This article investigates the stability of the ground state subspace of a canonical parent Hamiltonian of a Matrix product state against local perturbations. We prove that the spectral gap of such a Hamiltonian remains stable under weak local perturbations even in the thermodynamic limit, where the entire perturbation might not be bounded. Our discussion is based on preceding work by D.A. Yarotsky that develops a perturbation theory for relatively bounded quantum perturbations of classical Hamiltonians. We exploit a renormalization procedure, which on large scale transforms the parent Hamiltonian of a Matrix product state into a classical Hamiltonian plus some perturbation. We can thus extend D.A. Yarotsky's results to provide a perturbation theory for parent Hamiltonians of Matrix product states and recover some of the findings of the independent contributions [4, 9].

Contents

I. INTRODUCTION

The purpose of this article is to investigate the low energy sector of certain models of many-body quantum systems with local interaction. We are interested in the stability of quantum phases when small perturbations act on the system. In particular, we aim at understanding the conditions under which certain physical properties of the ground state change smoothly when an interaction is added to the model Hamiltonian. While for general models this question is intractably hard, in

*Electronic address: oleg.szehr@posteo.de
†Electronic address: wolf@ma.tum.de

this article we focus our eyesight on a restricted class, namely on parent Hamiltonians of Matrix product states.

Matrix product states (MPS) have been an extremely useful tool in the study of the ground state physics of many-body quantum systems. With their local structure MPS provide an efficient description of states arising from local interactions and constitute a natural framework for the analysis of local gapped Hamiltonians in 1D. In fact, the matrix product state representation lies at the heart of the very successful density matrix renormalization group method [19, 20]. To any MPS a local frustration-free and gapped Hamiltonian having this MPS as a unique ground state can be associated. A canonical choice of such Hamiltonians was introduced in [5] and is referred to as *parent Hamiltonian* of the MPS. On the one hand the local structure of the MPS endows the canonical parent Hamiltonian with the structure necessary for a rigorous analysis. On the other hand canonical parent Hamiltonians constitute a wide class of local Hamiltonians and include many important special cases such as the AKLT-Hamiltonian [1].

We are interested in how the parent Hamiltonian model behaves under small perturbations, as this allows one to use the idealization to predict the behaviour of actual physical systems. It seems generally expected that if a ground state of a quantum many-body system is in a non-critical regime characterized by the presence of a local spectral gap and exponential decay of correlations, then the system remains in this phase under sufficiently weak perturbations. We prove that for translationally invariant parent Hamiltonians of generic MPS this is indeed the case i.e. we show that the spectral gap of such a Hamiltonian is stable under arbitrary local perturbations even in the thermodynamic limit. This result itself is not new. It was shown in [9] that local Hamiltonians that satisfy the *Local Topological Quantum Order* (LTQO) condition and that are *locally gapped* are stable under local perturbations. It was also claimed in [9] and shown in [4] that parent Hamiltonians of MPS have LTQO. (However, in spin systems of higher dimension the presence of LTQO is hard to verify.) The fact that parent Hamiltonians are locally gapped was already known from [10]. Hence, the stability of the spectral gap against sufficiently weak perturbations follows.

The contribution at hand contains a new proof of this result. Our derivation is based on the observation that with increasing system scale a matrix product state "looks more and more classical" [16]. We exploit a renormalization group flow on parent Hamiltonians to prove that on sufficiently large scale a (generic) parent Hamiltonian can be seen as a perturbation of a classical system. Hence, any sufficiently small quantum perturbation of a parent Hamiltonian is equivalent to a relatively bounded perturbation of a classical model. We then draw on the theory for ground states in quantum perturbations of classical lattice systems by D.A. Yarotsky [23] to conclude our proof.

The results presented in this article were achieved independently of the contributions [4, 9] as a part of the doctoral thesis of the first author in the summer of 2011, before the publication of [4, 9].

II. PRELIMINARIES

As mentioned in the introduction, this article investigates how the ground state subspace of an MPS parent Hamiltonian behaves under small perturbations. This section reviews the required definitions and basic results.

A. Notation

We model quantum spin chains as connected subsets $\Lambda \subset \mathbb{Z}$, where each site $x \in \Lambda$ is equipped with a d-dimensional, complex Hilbert space $\mathcal{H}_x$. The total Hilbert space associated with a finite

subset $\Lambda \subset \mathbb{Z}$ will be denoted by $\mathcal{H}_\Lambda = \bigotimes_{x\in\Lambda} \mathcal{H}_x$. The interactions on the spin chain are given by a translationally invariant (TI) Hamiltonian with some fixed interaction range Λ_0. Such Hamiltonians can formally be written as

$$H_\Lambda = \sum_{x\in\Lambda} h_x,$$

where h_x is a positive semi-definite operator acting (non-trivially) on $\mathcal{H}_{\Lambda_0+x}$ and $\Lambda_0 + x$ is a translate of Λ_0 by x. We will assume that H_Λ has a non-degenerate ground state $|\Omega\rangle_\Lambda$ and that H_Λ has a *spectral gap* $\gamma > 0$ above the ground state energy

$$H_\Lambda\big|_{\mathcal{H}_\Lambda \ominus |\Omega\rangle_\Lambda} \geq \gamma\, \mathbb{1}.$$

Moreover, the Hamiltonians considered in this article will be *frustration free*, that is each interaction term h_x minimizes the global ground state energy: for all x we have $h_x|\Omega\rangle_\Lambda = 0$. We analyse how the spectral gap behaves if the Hamiltonian is perturbed with local interactions. Formally, we add a perturbation

$$\Phi_\Lambda = \sum_{x\in\Lambda} \phi_x,$$

where each of the terms ϕ_x acts locally on a finite subset of Λ. Often, we will find it convenient to identify the first and last site of Λ to impose *periodic boundary conditions* (PBC) on the system.

To distinguish particular Hilbert subspaces of $\mathcal{H}_\Lambda$ we will add Latin subscripts, for example $\mathcal{H}_A$ and $\mathcal{H}_B$. For any operator X acting on a finite subset of the chain we denote by $\|X\|_p$ the Schatten p-norm of X. If X acts on an infinite subsets we will only employ the $\|\cdot\|_\infty$-norm, which coincides with the usual operator norm.

As mentioned before we will consider a renormalization group flow that transforms the MPS parent Hamiltonian into a classical Hamiltonian. This flow will be modeled using a consecutive application of a linear map $\mathcal{T}$ acting on matrices X. More precisely, we define the map $\mathcal{T}$ by $\mathcal{T}(X) := \sum_i A_i X A_i^\dagger$, where the summation goes over a set of so-called Kraus operators $\{A_i\}_i$. Maps with this structure are completely positive (CP). For each such map the dual map $\mathcal{T}^*$ is defined by $\mathcal{T}^*(X) := \sum_i A_i^\dagger X A_i$. $\mathcal{T}^*$ is simply the adjoint of $\mathcal{T}$ with respect to the Hilbert-Schmidt inner product $\langle X|Y\rangle = \mathrm{tr}(X^\dagger Y)$. $\mathcal{T}$ is called unital (CPU) iff it preserves the identity operator $\mathcal{T}(\mathbb{1}) = \mathbb{1}$ and $\mathcal{T}$ is called trace-preserving (CPTP) iff $\mathcal{T}^*(\mathbb{1}) = \mathbb{1}$.

B. Matrix Product States

We consider a finite subset $\Lambda \subset \mathbb{Z}$ consisting of N sites, whose Hilbert spaces are each of dimension d. Every pure state of the spin system of Λ can be written as

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A_{i_1}^{[1]} \cdot A_{i_2}^{[2]} \cdot \ldots \cdot A_{i_N}^{[N]})|i_1 \ldots i_N\rangle$$

with site dependent $D_k \times D_{k+1}$ matrices $A_{i_k}^{[k]}$ [12, 17]. States of this structure are called Matrix product states. In the case of periodic boundary conditions and translational invariance of the MPS it is possible to show [12] that the matrices can be chosen in a site-independent way, i. e.

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{d} \mathrm{tr}(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_N})|i_1 \ldots i_N\rangle$$

with $D \times D$ matrices $\{A_i\}_{i=1,...d}$. In our consecutive discussion a special class of MPS will be of particular importance. This class is characterized by the following generic condition.

Condition (G1):
There is a finite number L_0 such that for all $L \geq L_0$ the list of matrices

$$\{A_{i_1} \cdot ... \cdot A_{i_L}\}_{i_j \in \{1...d\}}$$

spans the entire algebra of $D \times D$ matrices.

Condition (G1) is generic in the sense that d matrices chosen randomly according to some reasonable measure comply with this condition with probability one. It is not hard to see that (G1) holds iff the map

$$\Gamma_L : X \mapsto \sum_{i_1,...,i_L}^{d} \mathrm{tr}(X A_{i_1} A_{i_2} ... A_{i_L}) |i_1 ... i_L\rangle$$

is injective for $L \geq L_0$. The correspondence between sets $\{A_i\}_{i=1,...d}$ and MPS is not bijective; for example the set $\{X A_i X^{-1}\}_{i=1,...d}$ with invertible X belongs to the same state. It is shown in [12], Chapter 3 that the matrices of any MPS satisfying (G1) can be chosen to constitute a CPU map $\mathcal{T}$. More precisely, we can choose $\{A_i\}_{i=1,...d}$ such that the map $\mathcal{T}(X) = \sum_i A_i X A_i^\dagger$ satisfies $\mathcal{T}(\mathbb{1}) = \mathbb{1}$ and $\mathcal{T}^*(\Xi) = \Xi$ for some diagonal and strictly positive matrix Ξ. In addition, $\mathbb{1}$ is the only fixed point of $\mathcal{T}$. For a more detailed discussion of MPS we refer to [12].

C. Canonical Parent Hamiltonians

We consider a TI state $|\Psi\rangle = \sum_{i_1...i_N} \mathrm{tr}(A_{i_1} \cdot ... \cdot A_{i_N}) |i_1 ... i_N\rangle$ of a spin system with PBC on a chain Λ. For fixed $L \in \mathbb{N}$ we define $\mathcal{G}_L \subset (\mathbb{C}^d)^{\otimes L}$ to be the subspace spanned by the vectors $|\Psi(X)\rangle = \sum_{i_1...i_L} \mathrm{tr}(X A_{i_1} \cdot ... \cdot A_{i_L}) |i_1 ... i_L\rangle$, where X are complex $D \times D$ matrices. Note that if condition (G1) holds for the matrices A_i then for $L \geq L_0$ the space spanned by $|\Psi(X)\rangle$ has dimension D^2. We write $h_{\mathcal{G}_L}$ for the projector onto the orthogonal complement of $\mathcal{G}_L$ in $(\mathbb{C}^d)^{\otimes L}$. The *canonical parent Hamiltonian* for $|\Psi\rangle$ (and fixed L) is defined as the formal expression $H_\Lambda = \sum_i^N \tau^i(h_{\mathcal{G}_L})$ where τ denotes the translation operation by one site [5, 12]. For a parent Hamiltonian with nearest neighbour interaction (L=2) we will write $H_\Lambda = \sum_k h_{k,k+1}$ to emphasize this fact. It is clear from the definition that $H_\Lambda |\Psi\rangle = 0$ and that H_Λ is frustration free. Moreover, as a result of condition (G1) $|\Psi\rangle$ is the unique ground state of H_Λ if $L > L_0$ and $N \geq 2L_0$, [12, Theorem 10]. More generally, under (G1) H_Λ can be shown to have a spectral gap $\gamma > 0$ above the ground state energy [5, 12] even in the limit of an *infinite* chain. Let $\Lambda_1 \subset \Lambda$ and let G_{Λ_1} denote the projector onto the kernel of $H_{\Lambda_1} = \sum_{i:\{i+1,...,i+L\} \subset \Lambda_1} \tau^i(h_{\mathcal{G}_L})$. The *local gap* is defined to be the largest number γ_{Λ_1} such that

$$H_{\Lambda_1} \geq \gamma_{\Lambda_1} (\mathbb{1} - G_{\Lambda_1}) .$$

The local gap does not depend on Λ but only on the number of sites in Λ_1. The "Local-Gap condition" of [9] refers to the property of a general frustration-free Hamiltonian that the local gap decays at most polynomially in the number of lattice sites. It is one core assumption for the stability proof for frustration-free Hamiltonians (the other one being LTQO). In [10, 14] a constant lower bound on the local gap of one-dimensional, frustration-free Hamiltonians is derived. In particular, this implies that parent Hamiltonians satisfy the Local-Gap condition and we will naturally make use of this in our derivation. A more detailed discussions of parent Hamiltonians for MPS can be found in [12].

D. Stability of the spectral gap under quantum Perturbations of classical Hamiltonians

In this section we recall a fundamental result by D.A. Yarotsky [23] that asserts the stability of the spectral gap of a classical Hamiltonian under certain local perturbations. The effect of small quantum perturbations to classical Hamiltonians was discussed for example in [2, 7, 21, 22]. In [6, 23] this was extended to perturbations that need not necessarily be small but are required to consist of a small bounded part and a term that is bounded relatively to the unperturbed Hamiltonian. In the following we describe rigorously this perturbation theory.

We start with a chain $\Lambda \subset \mathbb{Z}$ with PBC and we consider a TI frustration-free Hamiltonian $H_\Lambda = \sum_{x\in\Lambda} h_x$. We will call H_Λ *classical* if in each space $\mathcal{H}_x$ there is a preferred vector $|\Omega\rangle_x$ and an orthogonal basis containing that vector such that the product basis in $\mathcal{H}_{\Lambda_0+x}$ diagonalizes h_x. Furthermore we assume that H_Λ has non-degenerate ground state $|\Omega\rangle_\Lambda = \bigotimes_{x\in\Lambda} |\Omega\rangle_x$ and strictly positive spectral gap above $|\Omega\rangle_\Lambda$. We consider perturbations $\Phi_\Lambda = \sum_{x\in\Lambda} \phi_x$ whose local terms act on finite subchains and that can be split into a purely bounded part $\phi_x^{(b)}$ and a relatively bounded part $\phi_x^{(r)}$ as

$$\phi_x = \phi_x^{(r)} + \phi_x^{(b)}. \tag{1}$$

The bounded part is characterized by

$$\|\phi_x^{(b)}\|_\infty \leq \beta. \tag{2}$$

For the relatively bounded part we suppose that for any $|\psi\rangle$ and any $I \subset \Lambda$

$$\left| \sum_{x\in I} \langle\psi|\phi_x^{(r)}|\psi\rangle \right| \leq \alpha \langle\psi|H_\Lambda|\psi\rangle. \tag{3}$$

Theorem 1 ([23, Theorem 2])**.** *Let $H_\Lambda = \sum_x h_x$ be a classical Hamiltonian on a chain Λ with PBC and non-degenerate gapped ground state $|\Omega\rangle_\Lambda$. Consider the perturbed Hamiltonian $\widetilde{H}_\Lambda = H_\Lambda + \Phi$, where $\Phi = \sum_x \phi_x$ is a perturbation that satisfies (1)-(3). For any $\kappa > 1$ there is $\delta(\kappa) > 0$ such that for any $\alpha \in (0,1)$ and $\beta = \delta(1-\alpha)^{2\kappa}$ the following conclusions hold:*

1. *$\widetilde{H}_\Lambda$ has a non-degenerate gapped ground state $|\widetilde{\Omega}\rangle_\Lambda$:*

$$\widetilde{H}_\Lambda |\widetilde{\Omega}\rangle_\Lambda = \widetilde{E}_\Lambda |\widetilde{\Omega}\rangle_\Lambda$$

and for some $\gamma > 0$ that does not depend on Λ

$$\widetilde{H}_\Lambda|_{\mathcal{H}_\Lambda \ominus |\widetilde{\Omega}\rangle_\Lambda} \geq (\widetilde{E}_\Lambda + \gamma)\, \mathbb{1}.$$

2. *There exists a thermodynamic weak*-limit of the ground states $|\widetilde{\Omega}\rangle_\Lambda$: For $\Lambda \to \mathbb{Z}$ one has that*

$$\langle A\widetilde{\Omega}_\Lambda | \widetilde{\Omega}_\Lambda\rangle \to \omega(A), \qquad A \in \bigcup_{|\Lambda|<\infty} \mathcal{B}(\mathcal{H}_\Lambda),$$

where $\mathcal{B}(\mathcal{H}_\Lambda)$ denotes the bounded operators on $\mathcal{H}_\Lambda$.

3. *There is an exponential decay of correlations in the infinite volume ground state ω: for some positive c and $\epsilon < 1$*

$$|\omega(A_1A_2) - \omega(A_1)\omega(A_2)| \leq c^{|\Lambda_1|+|\Lambda_2|} \epsilon^{\mathrm{dist}(\Lambda_1,\Lambda_2)} \|A_1\|_\infty \|A_2\|_\infty, \quad A_i \in \mathcal{B}(\mathcal{H}_{\Lambda_i}).$$

4. If within the allowed range of perturbations the term ϕ_x depends analytically on some parameters, then the ground state ω is also weakly-analytic in these parameters.*

Theorem 1 establishes that the spectral gap of a classical Hamiltonian is stable under perturbations that comply with the above assumptions. We will use this result to prove that parent Hamiltonians of MPS have a spectral gap that is stable under sufficiently weak bounded perturbations. To achieve this we will view the MPS parent Hamiltonian as a perturbation of a classical Hamiltonian, which is within a parameter range where Theorem 1 applies. The bounded part of this perturbation will decay faster under scaling of the system size than $\delta(1-\alpha)^{2\kappa}$. For sufficiently large systems this implies that under a small bounded perturbation ϕ'_x the parent Hamiltonian remains a perturbation of a classical Hamiltonian such that Theorem 1 applies. This provides us with the desired perturbation result.

III. STABILITY OF THE SPECTRAL GAP OF A CANONICAL PARENT HAMILTONIAN

In this section we state our main theorem. We consider a MPS that satisfies the generic condition (G1) and prove that the spectral gap of the corresponding parent Hamiltonian is stable under sufficiently weak perturbations. In the following corollary we extend this result and show that our discussion includes D.A. Yarotsky's perturbation theory for the AKLT model [23] as an important special case.

Theorem 2. *Let $|\Psi\rangle$ be a TI MPS on a finite ring Λ with PBC and suppose that for the matrices of $|\Psi\rangle$ condition (G1) holds. Suppose $N \geq 2L_0$ and choose $L > L_0$ and let $H_\Lambda = \sum_i \tau^i(h_{G_L})$ be the canonical parent Hamiltonian for $|\Psi\rangle$. Furthermore let $\Phi_\Lambda = \sum_k \phi_k$ be any finite range interaction with $\|\phi_k\|_\infty \leq \beta$ for a sufficiently small β depending on the range of Φ. Then all conclusions of Theorem 1 hold for the perturbed parent Hamiltonian $\widetilde{H}_\Lambda = H_\Lambda + \Phi_\Lambda$.*

Note that the above does not apply to important special cases as the AKLT model. There one considers a Hamiltonian with local nearest neighbour interaction but the matrices at each site do not span the whole algebra. The following simple corollary is to remedy this issue.

Corollary 3. *Let $H_\Lambda = \sum_i \tau^i(h_{G_L})$ be a canonical parent Hamiltonian such that Theorem 2 applies. Consider a Hamiltonian $\hat{H}_\Lambda = \sum_i h_{i,i+1}$ and suppose that there are positive constants c_1 and c_2 such that*

$$c_1\, h_{G_L} \leq \sum_{j=1}^{L-1} h_{j,j+1} \leq c_2\, h_{G_L}.$$

Then all conclusions of Theorem 1 also hold for $\hat{H}$.

The ground states of the AKLT model are MPS with $\{A_i\} = \{\sigma^z, \sqrt{2}\sigma^+, -\sqrt{2}\sigma^-\}$ [1, 12], where the σ's are the Pauli matrices. If we choose $\hat{H}$ to be he AKLT Hamiltonian Corollary 3 applies with $L = 3$ and implies the stability of the spectral gap of the AKLT model.

IV. PROOF OF STABILITY

We start this section with an outline of the proof of Theorem 2. In Section IV B we prove some lemmas from the theory of quantum channels and MPS. The following Subsection IV C contains a proof of Theorem 2 under the stronger assumption that the matrices $\{A_i\}_{i=1,\ldots,d}$ at each site of the chain span the whole algebra of $D \times D$ matrices. However, this assumption is not principal and in Section IV D we extend the previous discussion to prove stability under (G1).

A. Outline of the proof

For the readers convenience, before we proceed with the derivation of Theorem 2, we start with an exposition of core observations that will provide us with the proof.

1. We are given a MPS parent Hamiltonian H_Λ. We divide Λ into subchains Λ_k and we consider local sub-Hamiltonians $H_{\Lambda_k\cup\Lambda_{k+1}}$ of H_Λ acting on $\Lambda_k\cup\Lambda_{k+1}$. We analyze the behavior of the ground state subspace of $H_{\Lambda_k\cup\Lambda_{k+1}}$ under scaling of Λ_k. To this end we introduce density matrices $\rho_{\Lambda_k\cup\Lambda_{k+1}}$ whose image subspace is exactly the kernel of $H_{\Lambda_k\cup\Lambda_{k+1}}$.

2. Using a renormalization group flow we construct local unitaries W_{Λ_k} such that on sufficiently large scale the image of $W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}\rho_{\Lambda_k\cup\Lambda_{k+1}}W_{\Lambda_k}^\dagger\otimes W_{\Lambda_{k+1}}^\dagger$ has particularly simple structure. It turns out that in the asymptotic limit of large system size this image corresponds to the ground state subspace of a classical Hamiltonian.

3. We use convergence estimates from the theory of quantum Markov chains to show that the projectors $G_{\Lambda_k\cup\Lambda_{k+1}}$ onto the kernel of $H_{\Lambda_k\cup\Lambda_{k+1}}$ and $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ onto the kernel of the asymptotic classical Hamiltonian can be made exponentially close. We prove that $\|W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}G_{\Lambda_k\cup\Lambda_{k+1}}W_{\Lambda_k}^\dagger\otimes W_{\Lambda_{k+1}}^\dagger - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\|_\infty\leq\mathcal{O}(|\lambda_2|^{L/2})$.

4. We provide an explicit perturbation consisting of a bounded part $\sum_k\phi_k^{(b)}$ and a relatively bounded part $\sum_k\phi_k^{(r)}$ that transform the classical Hamiltonian into $\bigotimes_k W_{\Lambda_k}H_\Lambda\bigotimes_k W_{\Lambda_k}^\dagger$. Using the estimate from *3.* we show that these perturbations are in accordance with the conditions of Theorem 2. When adding a sufficiently small bounded perturbation to $\sum_k\phi_k^{(b)}$ the total perturbation remains in the range where Theorem 2 applies. Hence, the ground state subspace of H_Λ is stable.

B. Some Lemmas

We already mentioned (Section II B) that to any TI MPS we can associate a certain CPU map $\mathcal{T}$. To better keep track of the kernel of the canonical parent Hamiltonian it will be useful to introduce the operator $\rho_{EE'}=\frac{1}{D}\sum_{i_1,i_2,j_1,j_2}^d\operatorname{tr}(A_{i_1}A_{i_2}A_{j_2}^\dagger A_{j_1}^\dagger)|i_1\rangle\langle j_1|_E\otimes|i_2\rangle\langle j_2|_{E'}$, which is defined via the Kraus operators of $\mathcal{T}$. The subscripts E and E' have no physical significance but are introduced to more conveniently distinguish the systems involved. The following lemma shows that if two CPU maps $\mathcal{T}$ and $\widetilde{\mathcal{T}}$ are close, then the corresponding operators $\rho_{EE'}$ and $\widetilde{\rho}_{EE'}$ can be made close using a local unitary transformation.

Lemma 4. *Let $\mathcal{T}(X)=\sum_i^d A_iXA_i{}^\dagger$ and $\widetilde{\mathcal{T}}(X)=\sum_i^d\widetilde{A}_iX\widetilde{A}_i^\dagger$ be CPU maps. Consider the operators $\rho_{EE'}:=\frac{1}{D}\sum_{i_1,i_2,j_1,j_2}^d\operatorname{tr}(A_{i_1}A_{i_2}A_{j_2}^\dagger A_{j_1}^\dagger)|i_1\rangle\langle j_1|_E\otimes|i_2\rangle\langle j_2|_{E'}$ and $\widetilde{\rho}_{EE'}:=\frac{1}{D}\sum_{i_1,i_2,j_1,j_2}^d\operatorname{tr}(\widetilde{A}_{i_1}\widetilde{A}_{i_2}\widetilde{A}_{j_2}^\dagger\widetilde{A}_{j_1}^\dagger)|i_1\rangle\langle j_1|_E\otimes|i_2\rangle\langle j_2|_{E'}$. The following conclusions hold:*

1. *The operators $\rho_{EE'}$ and $\widetilde{\rho}_{EE'}$ are positive semidefinite and $\operatorname{tr}(\rho_{EE'})=\operatorname{tr}(\widetilde{\rho}_{EE'})=1$ (i.e. they are density operators).*

2. *There is a local unitary U_E such that*

$$\|U_E\otimes U_{E'}\,\rho_{EE'}\,U_E^\dagger\otimes U_{E'}^\dagger-\widetilde{\rho}_{EE'}\|_1\leq 4d^2\,\|\mathcal{T}-\widetilde{\mathcal{T}}\|_{CB}^{1/2},$$

where by $\|\cdot\|_{CB}$ we denote the norm of complete boundedness [11, 18].

Proof. The first assertion of the lemma follows by straightforward computations. For the second assertion we extend the CPU maps $\mathcal{T}$ and $\widetilde{\mathcal{T}}$ using Stinespring representations $V := \sum_i^d A_i^\dagger \otimes |i\rangle_E$ and $\widetilde{V} := \sum_i^d \widetilde{A}_i^\dagger \otimes |i\rangle_E$, respectively. Since

$$\mathcal{T}(\rho) = V^\dagger(\rho \otimes \mathbb{1}_E)V \quad \forall \rho,$$

V is indeed a Stinespring extension of $\mathcal{T}$ with dilation space $\mathcal{H}_E$. By assumption $\mathcal{T}$ is unital and thus V is an isometry, i. e. $V^\dagger V = \mathbb{1}$. Moreover, it is not hard to verify that the operator $\rho_{EE'}$ can be rewritten as

$$\rho_{EE'} = \left(\frac{1}{D}\,\mathrm{tr}_{\mathbb{C}^D}\left((V \otimes \mathbb{1}_E)VV^\dagger(V^\dagger \otimes \mathbb{1}_E)\right)\right)^T,$$

where $(\cdot)^T$ denotes transposition with respect to the computational basis. The corresponding statements hold for the operators $\widetilde{\mathcal{T}}$, $\widetilde{V}$ and $\widetilde{\rho}_{EE'}$. To shorten the notation we introduce the isometry $W := (\mathbb{1} \otimes (U_E)^T)V$, where U_E denotes a unitary acting on the E subsystem. Using the above expression for $\rho_{EE'}$ and the monotonicity of the Schatten 1-norm under the partial trace, we conclude that

$$\begin{aligned}
&\|U_E^\dagger \otimes U_{E'}^\dagger\,(\rho_{EE'})\,U_E \otimes U_{E'} - \widetilde{\rho}_{EE'}\|_1 \\
&= \|(U_E \otimes U_{E'})^T\,(\rho_{EE'})^T\,(U_E^\dagger \otimes U_{E'}^\dagger)^T - (\widetilde{\rho}_{EE'})^T\|_1 \\
&\leq \frac{1}{D}\left\|(W \otimes \mathbb{1}_E)WW^\dagger(W^\dagger \otimes \mathbb{1}_E) - (\widetilde{V} \otimes \mathbb{1}_E)\widetilde{V}\widetilde{V}^\dagger(\widetilde{V}^\dagger \otimes \mathbb{1}_E)\right\|_1 \\
&\leq d^2\left\|(W \otimes \mathbb{1}_E)WW^\dagger(W^\dagger \otimes \mathbb{1}_E) - (\widetilde{V} \otimes \mathbb{1}_E)\widetilde{V}\widetilde{V}^\dagger(\widetilde{V}^\dagger \otimes \mathbb{1}_E)\right\|_\infty \\
&\leq 4d^2\,\left\|W - \widetilde{V}\right\|_\infty
\end{aligned}$$

It follows from the continuity of the Stinespring extension (see [8, Theorem 1]) that the unitary $(U_E)^T$ acting on the dilation space can be chosen such that

$$\|W - \widetilde{V}\|_\infty^2 = \|(\mathbb{1} \otimes (U_E)^T)V - \widetilde{V}\|_\infty^2 \leq \|\mathcal{T} - \widetilde{\mathcal{T}}\|_{CB}.$$

□

As mentioned before the operators $\rho_{EE'}$ will help us to keep track of the behaviour of the kernels of local parent Hamiltonians under scaling. The images of $\rho_{EE'}$ will correspond to the kernels of the Hamiltonians. We write $P_{EE'}$ and $\tilde{P}_{EE'}$ for the projectors onto the images of $\rho_{EE'}$ and $\tilde{\rho}_{EE'}$. In the following we shall obtain conditions under which the distance of these projectors is small, i.e. the kernels of the parent Hamiltonians are almost the same.

Lemma 5. *Let ρ and $\tilde{\rho}$ be two Hermitian operators and let ρ^{-1} and $\tilde{\rho}^{-1}$ be their (Moore-Penrose-) pseudo inverses. Let $P = \rho\rho^{-1}$ and $\tilde{P} = \tilde{\rho}\tilde{\rho}^{-1}$ denote the projectors onto the images of ρ and $\tilde{\rho}$. Then for any Schatten p-norm $\|\cdot\|_p$ we have that*

$$\|P - \tilde{P}\|_p \leq \|\rho - \tilde{\rho}\|_p \left(\|\rho^{-1}\|_\infty + \|\rho^{-2}\|_\infty + \|\tilde{\rho}^{-2}\|_\infty + \|\rho^{-1}\|_\infty\|\tilde{\rho}^{-1}\|_\infty\right).$$

Proof. We rewrite the projectors P and $\tilde{P}$ using ρ^{-1} and $\tilde{\rho}^{-1}$ to conclude that

$$\begin{aligned}
\|P - \tilde{P}\|_p &= \|\rho\rho^{-1} - \tilde{\rho}\tilde{\rho}^{-1} - \tilde{\rho}\rho^{-1} + \tilde{\rho}\rho^{-1}\|_p \\
&\leq \|\rho^{-1}\|_\infty\|\rho - \tilde{\rho}\|_p + \|\tilde{\rho}\|_\infty\|\rho^{-1} - \tilde{\rho}^{-1}\|_p.
\end{aligned}$$

The distance $\|\rho^{-1} - \tilde{\rho}^{-1}\|_p$ can be bounded using the fact that

$$\rho^{-1} - \tilde{\rho}^{-1} = \rho^{-2}(\rho - \tilde{\rho})(\mathbb{1} - \tilde{P}) + (\mathbb{1} - P)(\rho - \tilde{\rho})\tilde{\rho}^{-2} - \rho^{-1}(\rho - \tilde{\rho})\tilde{\rho}^{-1}.$$

Applying the triangle inequality and the Hölder Inequality yields

$$\|\rho^{-1} - \tilde{\rho}^{-1}\|_p \leq \|\rho - \tilde{\rho}\|_p \left(\|\rho^{-2}\|_\infty + \|\tilde{\rho}^{-2}\|_\infty + \|\rho^{-1}\|_\infty \|\tilde{\rho}^{-1}\|_\infty\right)$$

which implies that

$$\|P - \tilde{P}\|_p \leq \|\rho - \tilde{\rho}\|_p \left(\|\rho^{-1}\|_\infty + \|\rho^{-2}\|_\infty + \|\tilde{\rho}^{-2}\|_\infty + \|\rho^{-1}\|_\infty \|\tilde{\rho}^{-1}\|_\infty\right).$$

□

In our main derivation we will encounter the situation, where $\tilde{\rho}$ is fixed whereas ρ depends on an integer, $\rho = \rho(L)$, and approaches $\tilde{\rho}$ as L goes to infinity. All operators $\rho(L)$ as well as the asymptotic operator $\tilde{\rho}$ will be density operators of the same rank. We write $\mu = \mu(L)$ for the smallest non-zero eigenvalue of $\rho(L)$ and accordingly $\tilde{\mu}$ for smallest non-zero eigenvalue of $\tilde{\rho}$. By Lemma 5 the convergence behaviour of the projectors $P = P(L)$ towards $\tilde{P}$ is governed by the distance $\|\rho - \tilde{\rho}\|_p$ and the largest eigenvalues $1/\mu$ and $1/\tilde{\mu}$ of ρ^{-1} and $\tilde{\rho}^{-1}$. The upper bound for the distance between the projectors P and $\tilde{P}$ obtained from Lemma 5 depends explicitly on $1/\mu$. However, when $\|\rho - \tilde{\rho}\|_\infty$ is small enough it follows from the continuity of eigenvalues that one can replace the dependence on $1/\mu$ by $1/\tilde{\mu}$.

Lemma 6. *Let ρ and $\tilde{\rho}$ be two density matrices of the same rank and let $\tilde{\mu}$ be the smallest positive eigenvalue of $\tilde{\rho}$. If $\|\rho - \tilde{\rho}\|_\infty < \tilde{\mu}$ then*

$$\|P - \tilde{P}\|_\infty \leq \frac{4\|\rho - \tilde{\rho}\|_\infty}{(\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty)^2}.$$

Proof. An application of Weyl's Perturbation Theorem [3] under exploitation of the fact that ρ and $\tilde{\rho}$ have the same rank shows that $|\mu - \tilde{\mu}| \leq \|\rho - \tilde{\rho}\|_\infty$. This yields an upper bound on the operator norm of ρ^{-1}:

$$\|\rho^{-1}\|_\infty = \frac{1}{\mu} \leq \frac{1}{\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty}.$$

We use Lemma 5 to conclude that

$$\begin{aligned}\|P - \tilde{P}\|_\infty &\leq \|\rho - \tilde{\rho}\|_\infty \left(\frac{1}{\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty} + \frac{1}{(\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty)^2} + \frac{1}{\tilde{\mu}^2} + \frac{1}{\tilde{\mu}(\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty)}\right) \\ &\leq \frac{4\|\rho - \tilde{\rho}\|_\infty}{(\tilde{\mu} - \|\rho - \tilde{\rho}\|_\infty)^2}.\end{aligned}$$

□

The proof of Theorem 2 relies on a renormalization group technique as introduced in [16]. We define local Hamiltonians acting on subchains of Λ. We then group the sites upon which these Hamiltonians act to blocks. The core observation is that the number of matrices required for the representation of the MPS will not increase from a certain point on. On the other hand with each grouping the blocked Hamiltonians "look more and more classical". The following lemma is taken from [16] and describes this blocking procedure more precisely. The consecutive application of this result to larger and larger subchains of Λ will be referred to as the *renormalization group flow*.

Lemma 7. *Let $\{A_i\}_{i=1,...,d}$ be a set of $D \times D$ matrices and consider the set $\{A_{i_1} \cdot ... \cdot A_{i_L}\}_{i_j=1,...,d}$ of all matrix products formed by matrices from $\{A_i\}_{i=1,...,d}$. There is a $d^L \times d^L$ unitary matrix U and matrices $A_m^{(L)}$ with*

$$\sum_{i_1,...,i_L}^{d} U_{m(i_1...i_L)} A_{i_1} \cdot ... \cdot A_{i_L} = A_m^{(L)} \tag{4}$$

such that $A_m^{(L)} = 0$ for all $m > \min\{D^2, d^L\}$. Moreover, it holds that $\mathcal{T}^L = \mathcal{T}^{(L)}$, where $\mathcal{T}^{(L)}$ denotes the CP map with Kraus operators $A_m^{(L)}$.

Proof. We write $(A_{i_1} \cdot ... \cdot A_{i_L})_{\alpha,\beta}$ with $\alpha, \beta \in \{1, ..., D\}$ for the entry of the matrix $A_{i_1} \cdot ... \cdot A_{i_L}$ in row α and column β. Let $\tilde{A}$ be the $d^L \times D^2$ matrix which has the entry $(A_{i_1} \cdot ... \cdot A_{i_L})_{\alpha,\beta}$ in its $(i_1...i_L)$-th row and (α, β)-th column. We perform a singular value decomposition of $\tilde{A}$ writing

$$\tilde{A}_{(i_1...i_L),(\alpha\beta)} = \sum_{l=1}^{\min(D^2, d^L)} (U^\dagger)_{(i_1...i_L),l}\, \rho_l\, V_{l,(\alpha\beta)}.$$

For the m-th row of $U\tilde{A}$, $(U\tilde{A})^{(m)}$, then

$$(U\tilde{A})^{(m)} = \begin{cases} \rho_m V^{(m)} & ;\ m \leq \min\{d^L, D^2\} \\ 0 & ;\ m > \min\{d^L, D^2\} \end{cases}$$

holds. The rows of the matrix $U\tilde{A}$ now correspond to the matrices $A_i^{(L)}$ and thus the first assertion of the lemma follows.
For the second assertion simply observe that for any X the quantity

$$\mathcal{T}^L(X) = \sum_{i_i,...,i_L}^{d} A_{i_1} \cdot ... \cdot A_{i_L} X A_{i_L}^\dagger \cdot ... \cdot A_{i_1}^\dagger$$

is invariant under unitary summations i.e.

$$\mathcal{T}^L(X) = \sum_m A_m^{(L)} X (A_m^{(L)})^\dagger = \mathcal{T}^{(L)}(X).$$

□

In the following lemma we analyse the asymptotic behaviour of the renormalization group flow and show that at large scale a generic TI MPS "looks classical". To achieve this, we consider large powers of the CPU map associated to the MPS and prove that the corresponding Kraus operators have a certain structure. It is well known that condition (G1) implies that the peripheral spectrum of $\mathcal{T}$ is trivial i.e. 1 is the only eigenvalue of $\mathcal{T}$ whose magnitude is one [5, 12, 13].

Lemma 8. *Let $\mathcal{T}(X) = \sum_i A_i X A_i^\dagger$ be a CPU map such that 1 is the unique eigenvalue of magnitude one and suppose that $\Xi = \mathrm{diag}(\xi_1, ..., \xi_n)$ with $\xi_i > 0$ is the corresponding fixed point of $\mathcal{T}^*$. Then the following conclusions hold:*

1. *The limit $\mathcal{T}^\infty := \lim_{n\to\infty} \mathcal{T}^n$ exists and we can write $\mathcal{T}^\infty(X) = \sum_{i=1}^{D^2} A_i^{(\infty)} X (A_i^{(\infty)})^\dagger$ with matrices $A_{(pq)}^{(\infty)} = \sqrt{\xi_q}|p\rangle\langle q|$ and $p, q \in \{1, ..., D\}$.*

2. The projector $P^{(\infty)}_{EE'}$ onto the image of

$$\rho^{(\infty)}_{EE'} := \frac{1}{D}\sum_{i_1,i_2,j_1,j_2}^{D^2} \mathrm{tr}\left(A^{(\infty)}_{i_1}A^{(\infty)}_{i_2}\left(A^{(\infty)}_{j_2}\right)^\dagger\left(A^{(\infty)}_{j_1}\right)^\dagger\right)|i_1\rangle\langle j_1|_E\otimes|i_2\rangle\langle j_2|_{E'}$$

can be written as

$$P^{(\infty)}_{EE'} = \mathbb{1}_A\otimes|\varphi\rangle\langle\varphi|_{BC}\otimes\mathbb{1}_D,$$

where $|\varphi\rangle=\sum_i\sqrt{\xi_i}|ii\rangle$, each of the subsystems A,B,C,D is isomorphic to $\mathbb{C}^D$, and $E=AB$, $E'=CD$.

Proof. All eigenvalues of a CPU map are contained in the closed unit disc in the complex plane. By assumption $\mathcal{T}$ has only one eigenvalue on the boundary and this eigenvalue is 1. Those eigenvalues of $\mathcal{T}^n$, which are contained in the open unit disc decay with increasing n, while 1 is an eigenvalue of $\mathcal{T}^n$ for any n. Hence, $\lim_{n\to\infty}\mathcal{T}^n$ simply converges to the projector onto the eigenvector $\mathbb{1}$ corresponding to the eigenvalue 1 of $\mathcal{T}$. The fact that $A^{(\infty)}_{(pq)}=\sqrt{\xi_q}|p\rangle\langle q|$ is then straight forward since the dual map $(\mathcal{T}^*)^\infty$ acts as $(\mathcal{T}^*)^\infty(X)=\mathrm{tr}(X)\Xi$.

It follows from the first assertion of the lemma and the fact that $\{A^{(\infty)}_i\}_{i=1,\dots,D^2}$ span the entire matrix algebra that the vectors $|\mu^{(\infty)}(X)\rangle=\sum_{i_1i_2}^{D^2}\mathrm{tr}(XA^{(\infty)}_{i_1}A^{(\infty)}_{i_2})|i_1i_2\rangle$ span the image of $\rho^{(\infty)}_{EE'}$. Furthermore they can be written as

$$|\mu^{(\infty)}(X)\rangle = (\mathbb{1}\otimes\sqrt{\Xi}X)_{AD}|\omega\rangle_{AD}|\varphi\rangle_{BC},$$

where $|\omega\rangle_{AD}=\sum_i|ii\rangle_{AD}$. Observe that $P^{(\infty)}_{EE'}$ as defined in the lemma has rank D^2 and $P^{(\infty)}_{EE'}|\mu^{(\infty)}(X)\rangle=|\mu^{(\infty)}(X)\rangle$. Therefore $P^{(\infty)}_{EE'}$ projects onto the image of $\rho^{(\infty)}_{EE'}$. □

C. The core argument

In this subsection we consider the stability of the spectral gap of a parent Hamiltonian with nearest neighbour interaction $H_\Lambda=\sum_k h_{k,k+1}$. We prove that the spectral gap is stable under the assumption that at each site $\{A_i\}_{i=1,\dots,d}$ span the entire algebra of $D\times D$ matrices. In the following subsections we extend this argument to show that stability holds more generally for generic MPS in the sense of (G1).

Proof of stability (Theorem 2) under strong assumptions. We show that at large scale the parent Hamiltonian H_Λ is a perturbation of a classical model and apply Theorem 1 to obtain the perturbation result. For this we divide Λ into blocks Λ_k of length L and block the terms of H_Λ into Hamiltonians $H_{\Lambda_k\cup\Lambda_{k+1}}:=\sum_{j:\{j,j+1\}\subset\Lambda_k\cup\Lambda_{k+1}}h_{j,j+1}$ acting locally on $\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}$ such that

$$H_\Lambda=\frac{1}{2}\sum_k\left(H_{\Lambda_k\cup\Lambda_{k+1}}+h_{kL,kL+1}\right).$$

For notational convenience we shall abbreviate $H_{k,k+1}:=\frac{1}{2}\left(H_{\Lambda_k\cup\Lambda_{k+1}}+h_{kL,kL+1}\right)$. Clearly it holds that

$$\mathrm{Kern}\,H_{k,k+1}=\mathrm{Kern}\,H_{\Lambda_k\cup\Lambda_{k+1}}$$

and that

$$H_{k,k+1} \geq \frac{1}{2} H_{\Lambda_k \cup \Lambda_{k+1}}.$$

We introduce the density matrix

$$\rho_{\Lambda_k \cup \Lambda_{k+1}} := \frac{1}{D} \sum_{\substack{i_1 \dots i_{2L} \\ j_1 \dots j_{2L}}}^{d} \operatorname{tr}(A_{i_1} \cdot \ldots \cdot A_{i_{2L}} A^\dagger_{j_{2L}} \cdot \ldots \cdot A^\dagger_{j_1}) |i_1 \ldots i_{2L}\rangle\langle j_1 \ldots j_{2L}|.$$

By assumption the matrices $\{A_i\}_{i=1,\ldots,d}$ span the entire matrix algebra. Hence, for any L the image of $\rho_{\Lambda_k \cup \Lambda_{k+1}}$ is spanned by the D^2-dimensional manifold of vectors

$$|\mu(X)\rangle = \sum_{i_1 \ldots i_{2L}}^{d} \operatorname{tr}(X A_{i_1} \cdot \ldots \cdot A_{i_{2L}}) |i_1 \ldots . i_{2L}\rangle,$$

where X is a $D \times D$ matrix with complex entries (see Section II B). On the other hand these vectors exactly span the kernel of $H_{\Lambda_k \cup \Lambda_{k+1}}$ (see Section II C and [12]) and we obtain

$$\operatorname{Im} \rho_{\Lambda_k \cup \Lambda_{k+1}} = \operatorname{Kern} H_{\Lambda_k \cup \Lambda_{k+1}}.$$

The local Hamiltonians $H_{\Lambda_k \cup \Lambda_{k+1}}$ have a positive spectral gap (see Section II C). Let $G_{\Lambda_k \cup \Lambda_{k+1}}$ denote the projector onto Kern $H_{\Lambda_k \cup \Lambda_{k+1}}$ then there is a $\gamma > 0$ that does not depend on L such that

$$H_{\Lambda_k \cup \Lambda_{k+1}} \geq \gamma(\mathbb{1} - G_{\Lambda_k \cup \Lambda_{k+1}}). \tag{5}$$

An application of Lemma 7 shows that there is a unitary U_{Λ_k} acting non-trivially on $\mathcal{H}_{\Lambda_k}$ only, with the property that

$$U_{\Lambda_k} \otimes U_{\Lambda_{k+1}} \, \rho_{\Lambda_k \cup \Lambda_{k+1}} \, U^\dagger_{\Lambda_k} \otimes U^\dagger_{\Lambda_{k+1}} = \left(\begin{array}{c|c} \rho^{(L)}_{EE'} & 0 \\ \hline 0 & 0 \end{array} \right), \tag{6}$$

where

$$\rho^{(L)}_{EE'} := \frac{1}{D} \sum_{\substack{i_1 i_2 \\ j_1 j_2}}^{\min\{D^2, d^L\}} \operatorname{tr}\left(A^{(L)}_{i_1} A^{(L)}_{i_2} (A^{(L)}_{j_2})^\dagger (A^{(L)}_{j_1})^\dagger \right) |i_1\rangle\langle j_1|_E \otimes |i_2\rangle\langle j_2|_{E'}$$

and the matrices $A^{(L)}_{i_j}$ are as in Lemma 7. The matrix $U_{\Lambda_k} \otimes U_{\Lambda_{k+1}} \, \rho_{\Lambda_k \cup \Lambda_{k+1}} \, U^\dagger_{\Lambda_k} \otimes U^\dagger_{\Lambda_{k+1}}$ acts on a space that is isomorphic to $(\mathbb{C}^d)^{\otimes L} \otimes (\mathbb{C}^d)^{\otimes L}$ but only the action on a $\left(\min\{D^2, d^L\}\right)^2$ dimensional subspace is non-zero. In the sequel we shall assume that L is chosen large such that $\rho^{(L)}_{EE'}$ acts on a $(D^2)^2$ dimensional space. For any given L we fix this space and define the matrix $\rho^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}}$ by replacing $\rho^{(L)}_{EE'}$ in that space by $\rho^{(\infty)}_{EE'}$ i.e.

$$\rho^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}} = \left(\begin{array}{c|c} \rho^{(\infty)}_{EE'} & 0 \\ \hline 0 & 0 \end{array} \right).$$

We denote by $G^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}}$ the projector onto the image of $\rho^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}}$. Note that since the orientation of the $(D^2)^2$ dimensional subspace in $(\mathbb{C}^d)^{\otimes L} \otimes (\mathbb{C}^d)^{\otimes L}$ can depend on L it follows that $\rho^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}}$

and $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ can depend on L.

We will now discuss the asymptotic properties of the matrices $\rho^{(L)}_{\Lambda_k\cup\Lambda_{k+1}}$. We will prove that with a suitable unitary transformation acting locally on the spaces $\mathcal{H}_{\Lambda_k}$ and with L chosen large the operators $\rho^{(L)}_{\Lambda_k\cup\Lambda_{k+1}}$ and $\rho^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ can be made arbitrarily close. This will provide us with an explicit unitary acting locally on (sufficiently large) spaces $\mathcal{H}_{\Lambda_k}$ that transforms the kernel of $H_{\Lambda_k\cup\Lambda_{k+1}}$ into a shape determined by $\rho^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$.

Let us consider the CPU map $\mathcal{T}$ associated with the MPS $|\Psi\rangle$ and let λ_2 denote its largest in magnitude subdominant eigenvalue. We note that $\sup_{k\geq 0}\left\|\mathcal{T}^k\right\|_{CB}=1$ i.e. $\mathcal{T}$ is power-bounded with respect to the CB-norm and constant 1. Hence, the discussion in [15] applies and yields that there is C that does not depend on L such that

$$\|\mathcal{T}^L-\mathcal{T}^\infty\|_{CB}\leq C|\lambda_2|^L.$$

By Lemma 7 this is equivalent to

$$\|\mathcal{T}^{(L)}-\mathcal{T}^{(\infty)}\|_{CB}\leq C|\lambda_2|^L,$$

where the maps $\mathcal{T}^{(L)}$ are defined as in the lemma. We apply Lemma 4 to conclude that there is a unitary V_E such that

$$\|V_E\otimes V_{E'}\,\rho^{(L)}_{EE'}\,V_E^\dagger\otimes V_{E'}^\dagger-\rho^{(\infty)}_{EE'}\|_\infty\leq 4D^4\sqrt{C}|\lambda_2|^{L/2}.$$

By Lemma 6 it holds for L chosen sufficiently large that

$$\|V_E\otimes V_{E'}\,P^{(L)}_{EE'}\,V_E^\dagger\otimes V_{E'}^\dagger-P^{(\infty)}_{EE'}\|_\infty\leq\frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu-4D^4\sqrt{C}|\lambda_2|^{L/2})^2},\tag{7}$$

where μ is the smallest non-zero eigenvalue of $\rho^{(\infty)}_{EE'}$. A straight forward computation shows that in fact μ equals the smallest eigenvalue of the fixed point matrix Λ.

Taken together, the inequalities (7) and (6) imply that the projectors onto the images of $\rho_{\Lambda_k\cup\Lambda_{k+1}}$ and $\rho^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ can be made exponentially close with a local unitary operation: There is a unitary W_{Λ_k} such that

$$\|W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}G_{\Lambda_k\cup\Lambda_{k+1}}W_{\Lambda_k}^\dagger\otimes W_{\Lambda_{k+1}}^\dagger-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\|_\infty\leq\frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu-4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\tag{8}$$

In terms of the Hamiltonians $H_{\Lambda_k\cup\Lambda_{k+1}}$ this means that we have achieved to construct a unitary acting locally on spaces $\mathcal{H}_{\Lambda_k}$ that on sufficiently large scale transforms the ground state space of $H_{\Lambda_k\cup\Lambda_{k+1}}$ into a a certain subspace determined by $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$. In the next step we a construct a classical Hamiltonian with this ground state subspace. For each L the structure of the operators $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ is known from Lemma 8. We have that

$$G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}=\left(\begin{array}{c|c}\mathbb{1}_A\otimes|\varphi\rangle\langle\varphi|_{BC}\otimes\mathbb{1}_D & 0\\ \hline 0 & 0\end{array}\right)$$

with $|\varphi\rangle=\sum_i\sqrt{\xi_i}|ii\rangle$. Thus $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ induces a natural decomposition of $\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}$ into a subspace $\mathcal{H}_X$ on which $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ acts as the zero operator and a subspace which is isomorphic to $\mathbb{C}^{D^2}\otimes\mathbb{C}^{D^2}$.

The latter can further be decomposed according to the structure of $G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}$ into $\mathbb{C}^{D^2}\otimes\mathbb{C}^{D^2}\cong \mathbb{C}^D_A\otimes\mathbb{C}^D_B\otimes\mathbb{C}^D_C\otimes\mathbb{C}^D_D$. By an additional decomposition of $\mathcal{H}_X$ and choosing L even we achieve the decomposition

$$\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}\cong(\mathbb{C}^D_A\oplus\mathcal{H}_{X_A})\otimes(\mathbb{C}^D_B\oplus\mathcal{H}_{X_B})\otimes(\mathbb{C}^D_C\oplus\mathcal{H}_{X_C})\otimes(\mathbb{C}^D_D\oplus\mathcal{H}_{X_D}).$$

Here the spaces $\mathcal{H}_{X_A},...,\mathcal{H}_{X_D}$ are chosen to have dimension $d^{L/2}-D$. In the decomposition of $\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}$ we identify the "half-shifted" spaces $\mathcal{H}^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}:=(\mathbb{C}^D_B\oplus\mathcal{H}_{X_B})\otimes(\mathbb{C}^D_C\oplus\mathcal{H}_{X_C})$. Note that $\mathcal{H}^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}\cong\mathcal{H}_{\Lambda_k}$ and that the following inclusions hold:

$$\mathcal{H}^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}\subset\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}\subset\mathcal{H}^{\mathrm{HS}}_{\Lambda_{k-1}\cup\Lambda_k}\otimes\mathcal{H}^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}\otimes\mathcal{H}^{\mathrm{HS}}_{\Lambda_{k+1}\cup\Lambda_{k+2}}.$$

Let $H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}$ denote the projector in $\mathcal{H}^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}$ onto the orthogonal complement of $|\varphi\rangle$. The above inclusions translate into the estimates

$$H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}\leq\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\leq H^{\mathrm{HS}}_{\Lambda_{k-1}\cup\Lambda_k}+H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}+H^{\mathrm{HS}}_{\Lambda_{k+1}\cup\Lambda_{k+2}}. \tag{9}$$

Consider the operator

$$H^{\mathrm{CL}}_\Lambda:=3L\sum_k H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}.$$

This operator is classical in the sense of Theorem 1 with respect to the half-shifted spaces $H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}$. We claim that for L chosen large enough $(\bigotimes_k W_{\Lambda_k})H_\Lambda(\bigotimes_k W_{\Lambda_k})^\dagger$ is a perturbation of H^{CL}_Λ satisfying the assumptions of Theorem 1. We construct this perturbation explicitly. It consists of a bounded part

$$\begin{aligned}\phi^{(b)}_{k,k+1}:=&W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})H_{k,k+1}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}\\&-(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}H_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})\end{aligned}$$

and a relatively bounded part

$$\begin{aligned}\phi^{(r)}_{k,k+1}:=&(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}H_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})\\&-L\,(H^{\mathrm{HS}}_{\Lambda_{k-1}\cup\Lambda_k}+H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}+H^{\mathrm{HS}}_{\Lambda_{k+1}\cup\Lambda_{k+2}}).\end{aligned}$$

Taking both together yields

$$(\bigotimes_k W_{\Lambda_k})H_\Lambda(\bigotimes_k W_{\Lambda_k})^\dagger=H^{\mathrm{CL}}_\Lambda+\sum_k\phi^{(b)}_{k,k+1}+\sum_k\phi^{(r)}_{k,k+1}.$$

First we estimate

$$\begin{aligned}\|\phi^{(b)}_{k,k+1}\|_\infty=&\Big\|W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})H_{k,k+1}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}\\&-(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}H_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})\Big\|_\infty\\\leq&\Big\|H_{k,k+1}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}-H_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})\Big\|_\infty\\&+\Big\|W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}(\mathbb{1}-G_{\Lambda_k\cup\Lambda_{k+1}})H_{k,k+1}-(\mathbb{1}-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}H_{k,k+1}\Big\|_\infty\\\leq&2\,\Big\|W\otimes WG_{\Lambda_k\cup\Lambda_{k+1}}W^\dagger\otimes W^\dagger-G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\Big\|_\infty\,\|H_{k,k+1}\|_\infty\\\leq&\frac{32LD^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu-4D^4\sqrt{C}|\lambda_2|^{L/2})^2}.\end{aligned}$$

The last inequality makes use of (8) and the fact that $\|H_{k,k+1}\|_\infty \leq L$. Thus we have shown that the norm of $\phi^{(b)}_{k,k+1}$ decays exponentially fast with increasing size of the blocks Λ_k.
To verify that $\phi^{(r)}_x$ is in accordance with the conditions of Theorem 1 we need to estimate $|\sum_{x\in I}\phi^{(r)}_x|$ for any $I \subset \{1,...,N/L\}$. The maximum is attained when $I = \Lambda$ since

$$\begin{aligned}\phi^{(r)}_{k,k+1} &\leq L\,(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}) - L\,(H^{\mathrm{SUB}}_{\Lambda_{k-1}\cup\Lambda_k} + H^{\mathrm{SUB}}_{\Lambda_k\cup\Lambda_{k+1}} + H^{\mathrm{SUB}}_{\Lambda_{k+1}\cup\Lambda_{k+2}})\\ &\leq 0,\end{aligned}$$

where the second inequality makes use of (9). A lower bound on $\phi^{(r)}_{k,k+1}$ follows from the gappedness of $H_{k,k+1}$ (5):

$$\begin{aligned}&(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}H_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}) \geq\\ &\frac{\gamma}{2}\,(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}})W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}(\mathbb{1} - G_{\Lambda_k\cup\Lambda_{k+1}})W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}) \geq\\ &\frac{\gamma}{2}\,\left(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\right)\left(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}} - \frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu - 4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\cdot\mathbb{1}\right)\left(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\right) \geq\\ &\frac{\gamma}{2}\,\left(1 - \frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu - 4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\right)\,\left(\mathbb{1} - G^{(\infty)}_{\Lambda_k\cup\Lambda_{k+1}}\right) \geq \frac{\gamma}{2}\,\left(1 - \frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu - 4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\right)\,H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}}.\end{aligned}$$

We sum the terms $\phi^{(r)}_{k,k+1}$ to conclude that

$$\begin{aligned}&\sum_{k,k+1}\phi^{(r)}_{k,k+1} \geq\\ &\sum_{k,k+1}\left(\frac{\gamma}{2}\,\left(1 - \frac{16D^4\sqrt{C}|\lambda_2|^{L/2}}{(\mu - 4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\right)\,H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}} - L\,(H^{\mathrm{HS}}_{\Lambda_{k-1}\cup\Lambda_k} + H^{\mathrm{HS}}_{\Lambda_k\cup\Lambda_{k+1}} + H^{\mathrm{HS}}_{\Lambda_{k+1}\cup\Lambda_{k+2}})\right)\\ &= \left(-1 + \frac{\gamma}{6L} - \frac{8\gamma D^4\sqrt{C}|\lambda_2|^{L/2}}{3L(\mu - 4D^4\sqrt{C}|\lambda_2|^{L/2})^2}\right)\,H^{\mathrm{CL}}_\Lambda.\end{aligned}$$

Thus for Theorem 1 we have that

$$\left|\sum_{k,k+1}\langle\psi|\phi^{(r)}_{k,k+1}|\psi\rangle\right| \leq \alpha\,\langle\psi|H^{\mathrm{CL}}_\Lambda|\psi\rangle$$

with $\alpha = (1 - \frac{\gamma}{6L} + \mathcal{O}(|\lambda_2|^{\frac{L}{2}}))$ and $\beta = \delta\,(\frac{\gamma}{6L} - \mathcal{O}(|\lambda_2|^{\frac{L}{2}}))^{2\kappa}$, where the constants δ and κ still have to be chosen appropriately. As long as γ decays sub-exponentially fast with L, for L sufficiently large $\|\phi^{(b)}_{k,k+1}\|_\infty \leq \beta$ holds. For parent Hamiltonians, which have a constant local gap this is certainly the case.

Applying Theorem 1 we could recover the well-known fact that H_Λ has a gapped ground state. However, the conditions of Theorem 1 are "open" in the sense that adding sufficiently small bounded perturbation to $\phi^{(b)}_{k,k+1}$ still results in a total perturbation, which is within the range where Theorem 1 can be applied. This provides us with a perturbation result for Hamiltonians in the neighbourhood of H_Λ. More precisely, let $\Phi' := \sum_{k,k+1}\phi'_{k,k+1}$ be a finite range interaction with $\|\phi'_{k,k+1}\|_\infty \leq \beta'$ and $\beta' > 0$ small enough. We analyse the spectral gap of $H'_\Lambda = H_\Lambda + \Phi'$. Suppose for the moment that $\phi'_{k,k+1}$ acts exactly on $\mathcal{H}_{\Lambda_k\cup\Lambda_{k+1}}$ and let

$$\phi''_{k,k+1} := W_{\Lambda_k}\otimes W_{\Lambda_{k+1}}\phi'_{k,k+1}W^\dagger_{\Lambda_k}\otimes W^\dagger_{\Lambda_{k+1}}.$$

Consider the Hamiltonian

$$(\bigotimes_k W_{\Lambda_k})H_\Lambda(\bigotimes_k W_{\Lambda_k})^\dagger + \sum_k \phi''_{k,k+1} = (\bigotimes_k W_{\Lambda_k})(H_\Lambda + \Phi')(\bigotimes_k W_{\Lambda_k})^\dagger.$$

If $\beta' > 0$ is chosen sufficiently small Theorem 1 applies and proves the stability of the spectral gap of $H_\Lambda + \Phi'$. In general, though, we want to allow an arbitrary (finite) interaction range for $\phi_{k,k+1}$. If $\phi_{k,k+1}$ acts nontrivially on a subchain of $\Lambda_k \cup \Lambda_{k+1}$ only it is possible to group the $\phi_{k,k+1}$ terms in such a way that in total one gets a finite range interaction on $\Lambda_k \cup \Lambda_{k+1}$. Choosing β' we make sure that the grouped perturbation is sufficiently small for an application of Theorem 1. On the other hand if the perturbation has interaction range exceeding the subchain $\Lambda_k \cup \Lambda_{k+1}$ one simply chooses L larger and the previous discussion applies to the larger subchains. □

D. Proof of Theorem 2 and Corollary 3

Proof of Theorem 2. The proof is a simple upgrade of the restricted discussion of the previous subsection. By condition (G1) there is finite P_0 such that the matrices $\{A_{i_1} \cdot ... \cdot A_{i_{P_0}}\}$ span the whole algebra of $D \times D$ matrices. Hence, $H_\Lambda = \sum_i \tau^i(h_{G_P})$ has a unique ground state for any $P > P_0$, see Section II C. We proceed as in the proof of the theorem and divide Λ into chains Λ_k of length L. In addition we assume that the chains are sufficiently large to support $h_{\mathcal{G}_P}$ i. e. $L \geq P$. We define the operators

$$H_{\Lambda_k \cup \Lambda_{k+1}} := \sum_{i:\{i+1,...,i+P\} \subset \Lambda_k \cup \Lambda_{k+1}} \tau^i(h_{\mathcal{G}_P}),$$

which are sums of all the translates of $h_{\mathcal{G}_P}$ that act locally on $\Lambda_k \cup \Lambda_{k+1}$. There are $P-1$ terms in the above Hamiltonian that partially act on block Λ_k and partially on Λ_{k+1}. We define the operators $H_{k,k+1}$ by adding these terms to $H_{\Lambda_k \cup \Lambda_{k+1}}$. Formally

$$H_{k,k+1} = \frac{1}{2} H_{\Lambda_k \cup \Lambda_{k+1}} + \frac{1}{2} \sum_{i:(i+1 \in \Lambda_k \ \wedge \ i+P \in \Lambda_{k+1})} \tau^i(h_{\mathcal{G}_P}).$$

As before, we have the properties

$$\begin{aligned} &H_{k,k+1} \geq H_{\Lambda_k \cup \Lambda_{k+1}}, \\ &\mathrm{Kern}(H_{k,k+1}) = \mathrm{Kern}(H_{\Lambda_k \cup \Lambda_{k+1}}) \end{aligned}$$

and

$$H_\Lambda = \sum_k H_{k,k+1}.$$

The kernel of $H_{\Lambda_k \cup \Lambda_{k+1}}$ is given by the image (see also [12, Section 4.1.1]) of

$$\rho_{\Lambda_k \cup \Lambda_{k+1}} = \sum_{\substack{i_1 ... i_{2L} \\ j_1 ... j_{2L}}}^{d} \mathrm{tr}(A_{i_1} \cdot ... \cdot A_{i_{2L}} A^\dagger_{j_{2L}} \cdot ... \cdot A^\dagger_{j_1}) |i_1 ... i_{2L}\rangle\langle j_1 ... j_{2L}|.$$

As before, the spectral gap of $H_{\Lambda_k \cup \Lambda_{k+1}}$ can be lower bounded by some constant. With $G_{\Lambda_k \cup \Lambda_{k+1}}$ and $G^{(\infty)}_{\Lambda_k \cup \Lambda_{k+1}}$ defined as in Subsection IV C the derivation follows the same lines as before. Hence, stability follows under condition (G1), which completes the proof of Theorem 2. □

Proof of Corollary 3. As before we choose $L \geq P$ and divide Λ into subchains of length L. The restrictions of $\hat{H}_\Lambda = \sum_j h_{j,j+1}$ and $H_\Lambda = \sum_i \tau^i(h_{G_P})$ to $\Lambda_k \cup \Lambda_{k+1}$ are given by $\hat{H}_{\Lambda_k \cup \Lambda_{k+1}} = \sum_{\{j,j+1\} \subset \Lambda_k \cup \Lambda_{k+1}} h_{j,j+1}$ and $H_{\Lambda_k \cup \Lambda_{k+1}} = \sum_{i:\{i+1,\ldots,i+P\} \subset \Lambda_k \cup \Lambda_{k+1}} \tau^i(h_{\mathcal{G}_P})$. The condition $c_1\, h_{G_P} \leq \sum_{j=1}^{P-1} h_{j,j+1} \leq c_2\, h_{G_P}$ implies that

$$c_1\, H_{\Lambda_k \cup \Lambda_{k+1}} \leq \sum_{i:\{i+1,\ldots,i+P\} \subset \Lambda_k \cup \Lambda_{k+1}} \tau^i \left(\sum_{j=1}^{P-1} h_{j,j+1} \right) \leq c_2\, H_{\Lambda_k \cup \Lambda_{k+1}}.$$

It follows that $\sum_{i:\{i+1,\ldots,i+P\} \subset \Lambda_k \cup \Lambda_{k+1}} \tau^i \left(\sum_{j=1}^{P-1} h_{j,j+1} \right)$ has the same kernel as $H_{\Lambda_k \cup \Lambda_{k+1}}$. Thus the kernels of $H_{\Lambda_k \cup \Lambda_{k+1}}$ and $\hat{H}_{\Lambda_k \cup \Lambda_{k+1}}$ are identical and Corollary 3 follows from the derivation of Theorem 2. □

Acknowledgments

We acknowledge financial support from the QCCC programme of the Elite Network of Bavaria, the CHIST-ERA/BMBF project CQC and the Alfried Krupp von Bohlen und Halbach-Stiftung.

[1] I. Affleck, T. Kennedy, E. Lieb, and H. Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.*, 59(7):799–802, 1987.

[2] C. Albanese. Unitary dressing transformations and exponential decay below the threshold for quantum spin systems. *Comm. in Math. Phys.*, 134:1–27, 237–272, 1990.

[3] R. Bhatia. *Matrix Analysis*. Springer, 1996.

[4] I. Cirac, S. Michalakis, D. Péres-García, and N. Schuch. Robustness in Projected Entangled Pair States. 2013. arXiv: 1306.4003.

[5] M. Fannes, B. Nachtergaele, and R. Werner. Finitely correlated states on quantum spin chains. *Comm. in Math. Phys.*, 144(3):443–490, 1992.

[6] T. Kennedy and H. Tasaki. Hidden symmetry breaking in the Haldane phase S=1 quantum spin chains. *Comm. in Math. Phys.*, 147:431–484, 1992.

[7] T. Kennedy and H. Tasaki. Hidden $Z_2 \times Z_2$ symmetry breaking in Haldane gap antiferromagnets. *Phys. Rev. B*, 45:304, 1992.

[8] D. Kretschmann, D. Schlingemann, M. Wolf, and R. Werner. The information-disturbance tradeoff and the continuity of Stinespring's theorem. *IEEE Trans. on Inf. Theory*, 54(4):1708–1717, 2006.

[9] S. Michalakis and J. Pytel. Stability of Frustration-Free Hamiltonians. *Comm. in Math. Phys.*, 322(2):277–302, 2013.

[10] B. Nachtergaele. The spectral gap for some quantum spin chains with discrete symmetry breaking. *Comm. in Math. Phys.*, 175:565–606, 1996.

[11] V. I. Paulsen. *Completely bounded maps and operator algebras*. Cambridge University Press, 2002.

[12] D. Pérez-García, F. Verstaete, M. Wolf, and J. Cirac. Matrix product state representations. *Quantum Inf. Comput.*, 7:401–430, 2007.

[13] M. Sanz, D. Péres-García, M. Wolf, and C. J.I. A quantum version of Wielandt's inequality. *IEEE Trans. on Inf. Theory*, 56:9:4668–4673, 2010.

[14] W. Spitzer and S. Starr. Improved bounds on the spectral gap above frustration free ground states of quantum spin chains. *Lett. Math. Phys.*, 63:165–177, 2002.

[15] O. Szehr, D. Reeb, and M. Wolf. Spectral convergence bounds for classical and quantum Markov processes. 2013. arXiv: 1301.4827.

[16] F. Verstraete, J. I. Cirac, J. I. Latorre, E. Rico, and M. M. Wolf. Renormalization-group transformations on quantum states. *Phys. Rev. Lett.*, 94:140601, Apr 2005.

[17] G. Vidal. *Phys. Rev. Lett.*, 91:147902, 2003.

[18] J. Watrous. Notes on super-operator norms induced by schatten norms. arXiv: 0411077v1, Nov. 2004.
[19] S. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863, 1992.
[20] K. Wilson. The renormalization group: Critical phenomena and the Kondo problem. *Rev. Mod. Phys.*, 47:773–840, 1975.
[21] D. Yarotsky. Perturbations of ground states in weakly interacting quantum spin systems. *J. Math. Phys.*, 45:2134–2152, 2004.
[22] D. Yarotsky. Uniqueness of the ground state in weak perturbations of non-interacting gapped quantum lattice systems. *J. Stat. Phys.*, 118:119–144, 2005.
[23] D. A. Yarotsky. Ground states in relatively bounded quantum perturbations of classical lattice systems. *Comm. in Math. Phys.*, 261(3):799–819, 2004.