New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Elementary proof of QAOA convergence

To cite this article: Lennart Binkowski *et al* 2024 *New J. Phys.* **26** 073001

View the article online for updates and enhancements.

You may also like

- Quantum approximate optimization algorithm in non-Markovian quantum systems
Bo Yue, Shibei Xue, Yu Pan et al.
- Convergence of digitized-counterdiabatic QAOA: circuit depth versus free parameters
Mara Vizzuso, Gianluca Passarelli, Giovanni Cantele et al.
- To quantum or not to quantum: towards algorithm selection in near-term quantum optimization
Charles Moussa, Henri Calandra and Vedran Dunjko

This content was downloaded from IP address 134.130.130.179 on 28/01/2025 at 06:03

IOP Publishing New J. Phys. 26 (2024) 073001 https://doi.org/10.1088/1367-2630/ad59bb

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics



OPEN ACCESS

RECEIVED
7 December 2023

REVISED
12 June 2024

ACCEPTED FOR PUBLICATION
19 June 2024

PUBLISHED
1 July 2024

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Elementary proof of QAOA convergence

Lennart Binkowski[1], **Gereon Koßmann**[2], **Timo Ziegler**[1,3] and **René Schwonnek**[1,*]

[1] Institut für Theoretische Physik, Leibniz Universität Hannover, Hannover, Germany
[2] Institute for Quantum Information, RWTH Aachen University, Aachen, Germany
[3] Volkswagen AG, Berliner Ring 2, 38440 Wolfsburg, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: r.schwonnek@gmail.com, lennart.binkowski@itp.uni-hannover.de, kossmann@physik.rwth-aachen.de and timo.ziegler@volkswagen.de

Keywords: quantum computing, QAOA, VQA, adiabatic evolution

Abstract

The quantum alternating operator ansatz (QAOA) and its predecessor, the quantum approximate optimization algorithm, are one of the most widely used quantum algorithms for solving combinatorial optimization problems. However, as there is yet no rigorous proof of convergence for the QAOA, we provide one in this paper. The proof involves retracing the connection between the quantum adiabatic algorithm and the QAOA, and naturally suggests a refined definition of the 'phase separator' and 'mixer' keywords.

1. Introduction

In the current era of gate-based noisy quantum computers, the class of *variational quantum algorithms* (VQA) is at the center of research. First and foremost, the *quantum approximate optimization algorithm* [1] receives enormous scientific as well as industrial attention. Like many other VQAs, it is developed for the purpose of solving *combinatorial optimization problems* (COP) (maximize $f : \{0,1\}^n \to \mathbb{R}$ subject to some constraints) on quantum computers with the aid of classical optimizers. It is, to some extend, a discretized and gate-based version of the *quantum adiabatic algorithm* (QAA, [2]) which is itself a continuous-time algorithm. The QAA and the closely related *quantum annealing* [3] rely on slowly evolving a quantum system (resp. some external parameters) in order to transition a well-known initial state into some state representing an optimal solution. Due to their analog structure, they are not executable on gate-based architectures, but on *quantum annealers* (see [4] for an overview) which constitute the second large family of quantum computer architectures.

In its original formulation, the quantum approximate optimization algorithm is only suited for unconstrained problems. A common technique for enlarging its scope to constrained problems is *softcoding* the constraints. That is, the constraints enter the objective function as additional terms, penalizing infeasible inputs. However, for several instances, this approach was observed to produce unfavorable output distributions which suffer from poor optimization quality or feasibility violation (see, e.g. [5–7]). In order to improve the treatment of constrained problems Hadfield *et al* extended the quantum approximate optimization algorithm to the *quantum alternating operator ansatz* (QAOA, [8]) which also allows for *hardcoding* the constraints. That is, the objective function is left unchanged and feasibility preservation is instead enforced strictly.

In a nutshell, a QAOA-circuit consists of parametrized *phase separator* gates U_P and *mixer* gates U_M. Both types of gates should preserve feasibility such that—in an ideal setting—feasible states are mapped to feasible states again. The mixer typically is concerned with spreading the amplitude across different computational basis states while the phase separator—often the exponentiated cost Hamiltonian—marks the state according to their classical objective value. Classically and iteratively optimizing the circuit parameters then should yield a good approximation of an optimal solution. This heuristic argument goes through only if every feasible state can be reached: The QAOA-circuit is, given the right parameter values, able to (approximately) produce every feasible state. Typically, the reachability of feasible states only depends on the properties of the mixer U_M.

© 2024 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Physics and Deutsche Physikalische Gesellschaft

A first sketch of a proof why the quantum approximate optimization algorithm should converge for (unconstrained) COPs was already given in [1]—although restricted to the special case of problems having only one optimal solution. That derivation, in turn, builds on the close connection to the QAA and the underlying principle of adiabatic evolution/quantum annealing (see [9] for mathematical treatment). However, neither is the argument carried out in great mathematical detail, nor does it attempt to be as general as possible. Moreover, since the QAOA comprises similar principles as the quantum approximate optimization algorithm, it stands to reason to extend this result, once suitably formalized, to the QAOA; a task that, surprisingly, has not yet been tackled. With this paper we address this issue and come up with refined definitions for the phase separator and mixer gates which make the connection to the quantum approximate optimization algorithm more visible.

First, we prove the convergence of the QAA with suitable initial Hamiltonian and initial state in section 3. The proof is built on the aforementioned proof sketch in [1]. We extract the underlying principles and already obtain a precise definition for a *mixer Hamiltonian*. However, by invoking a version of the adiabatic theorem without *gap condition*, we obtain a more general result which does not require the considered optimization problem to have only one single optimal solution.

Second, we prove the convergence of the QAOA with suitable initial state in section 4. For this, we generalize all the properties of the original mixer proposed in the quantum approximate optimization algorithm. We define our versions of *simultaneous* and *sequential mixers* which directly make use of the just generalized properties. The convergence proof is then built on the convergence of the QAA instance which admits the respective mixer Hamiltonian as initial Hamiltonian. The underlying idea is again due to Farhi *et al* but suitably generalized for constrained problems and sequential mixers.

2. Preliminaries

2.1. Combinatorial optimization problems

In the following, we restrict to maximization problems, as minimization tasks may be considered analogously. This choice of the optimization direction simply allows us to state the convergence proofs more compactly. A generic COP of *size* N is of the form

$$\max_{\boldsymbol{z}\in S} f(\boldsymbol{z})\,, \quad S \subseteq Z(N)\,, \tag{1}$$

where $Z(N)$ denotes the set of bit strings of length N, $f : Z(N) \to \mathbb{R}$ is the *objective function*, and $S \subseteq Z(N)$ is the set of *feasible bit strings* or the *solution set*. The problem is called *unconstrained* if $S = Z(N)$. Moreover, we denote the set of all solutions maximizing f by $S_{\max}$.

2.2. Problem encoding on quantum computers

In order to treat a COP with the help of quantum computers, the problem first has to be translated into a quantum-mechanical language. The standard encoding procedure identifies each bit string $\boldsymbol{z}$ with a computational basis state $|\boldsymbol{z}\rangle$ of the N-qubit Hilbert space $\mathcal{H} := \mathbb{C}^{2^N}$ and translates functions $Z(N) \to \mathbb{C}$ to linear operators on $\mathcal{H}$, $\mathcal{L}(\mathcal{H})$. Namely, classical objective function f is considered as an objective Hamiltonian C via

$$C := \sum_{\boldsymbol{z}\in Z(N)} f(\boldsymbol{z})\,|\boldsymbol{z}\rangle\langle\boldsymbol{z}|. \tag{2}$$

In this setting, the (optimal) solution bit strings span the (*optimal*) *solution space*

$$\mathcal{S} := \operatorname{span}\{|\boldsymbol{z}\rangle \,:\, \boldsymbol{z}\in S\} \subseteq \mathcal{H} \tag{3}$$

$$\mathcal{S}_{\max} := \operatorname{span}\{|\boldsymbol{z}\rangle \,:\, \boldsymbol{z}\in S_{\max}\} \subseteq \mathcal{S}. \tag{4}$$

Note that C both spaces $\mathcal{S}$ and $\mathcal{S}_{\max}$ invariant. Especially, one can consider the restriction of C to $\mathcal{S}$

$$C|_{\mathcal{S}} := \sum_{\boldsymbol{z}\in S} f(\boldsymbol{z})\,|\boldsymbol{z}\rangle\langle\boldsymbol{z}| \tag{5}$$

The maximization task is now equivalent to finding a computational basis state in $\mathcal{S}_{\max}$. By construction, $\mathcal{S}_{\max}$ is the eigenspace of $C|_{\mathcal{S}}$ (C restricted to $\mathcal{S}$) corresponding to its largest eigenvalue. In the following, we will slightly relax the quantum optimization task as we will consider any highest energy state of $C|_{\mathcal{S}}$ an optimal solution.

2.3. Quantum adiabatic algorithm

In a nutshell, the continuous-time *quantum adiabatic algorithm* (QAA, [2]) tackles the eigenstate search via quasi-adiabatic evolution of an initial state $|\iota\rangle$ with respect to a time-dependent Hamiltonian $H(t)$ which interpolates between an initial Hamiltonian H_I and the objective Hamiltonian C. In case of a maximization task, $|\iota\rangle$ should be a highest energy state of H_I. The interpolating Hamiltonian is typically given by the convex combination [10]

$$H(t) = H_{\mathrm{lin}(H_\mathrm{I},C)}(t) := (1-t)\,H_\mathrm{I} + tC, \quad t \in [0,1]\,. \tag{6}$$

The evolution speed is controlled via a parameter $T > 0$: The actual time evolution is with respect to $H(s/T)$, $s \in [0, T]$. The intuition behind the QAA is that evolving a highest energy state of $H(0)$ sufficiently slowly yields a highest energy state of C if the energy levels stay separated. Mathematical rigor is granted by the adiabatic theorems (see section 3).

2.4. Quantum approximate optimization algorithm

The *quantum approximate optimization algorithm* [1] can, in some sense, be seen as a discrete version of the QAA with fixed initial state

$$|+\rangle := \frac{1}{\sqrt{2^N}} \sum_{\boldsymbol{z} \in Z(N)} |\boldsymbol{z}\rangle \tag{7}$$

and initial Hamiltonian

$$B_\mathrm{QAA} := \sum_{n=1}^{N} \sigma_x^{(n)}. \tag{8}$$

Note that $|+\rangle$ is the nondegenerate highest energy state of B. B and C are incorporated into parametrized gates:

$$U_{B_\mathrm{QAA}}(\beta) := \mathrm{e}^{-\mathrm{i}\beta B_\mathrm{QAA}} = \prod_{n=1}^{N} \mathrm{e}^{-\mathrm{i}\beta\sigma_x^{(n)}} \quad \text{and} \tag{9}$$

$$U_C(\gamma) := \mathrm{e}^{-\mathrm{i}\gamma C}. \tag{10}$$

Specifying a *depth* $p \in \mathbb{N}$, the parametrized trial states are constructed via

$$|\vec{\beta},\vec{\gamma}\rangle := V\left(\vec{\beta},\vec{\gamma}\right)|+\rangle := \left(\prod_{q=1}^{p} U_{B_\mathrm{QAA}}(\beta_q)\,U_C(\gamma_q)\right)|+\rangle. \tag{11}$$

In an iterative process, the parameters are updated by a classical optimization rule in order to maximize the expectation value

$$F_p\left(\vec{\beta},\vec{\gamma}\right) = \langle\vec{\beta},\vec{\gamma}|C|\vec{\beta},\vec{\gamma}\rangle. \tag{12}$$

Measuring the final outcome $|\vec{\beta}_\mathrm{opt},\vec{\gamma}_\mathrm{opt}\rangle$ in the computational basis then yields a distribution of optimal solution approximations.

2.5. Quantum alternating operator ansatz

Building on the ideas of the quantum approximate optimization algorithm, the *quantum alternating operator ansatz* (QAOA, [8]) extends its design to general constrained problems. Given a COP with objective Hamiltonian C and solution space $\mathcal{S}$, the parametrized gate U_{B_QAA} is substituted with problem-specific 'mixer' gates. For simplicity, we will focus on the case where the same mixers are used in every iteration. Thereby, we can collect them again in a single mixer gate $U_\mathrm{M}(\beta)$. It is demanded to fulfill two important properties:

- Feasibility preservation: for all parameter values $\beta \in \mathbb{R}$, $U_\mathrm{M}(\beta)$ should leave $\mathcal{S}$ invariant.
- Full mixing of solutions: for all feasible computational basis states $|\boldsymbol{z}\rangle, |\boldsymbol{z}'\rangle \in \mathcal{S}$, there should exist a power $r \in \mathbb{N}$ and a parameter value $\beta \in \mathbb{R}$ so that $\langle\boldsymbol{z}|U_\mathrm{M}^r(\beta)|\boldsymbol{z}'\rangle \neq 0$.

Furthermore, the parametrized gate $U_C(\gamma)$ could be replaced by a more general 'phase separator' gate $U_P(\gamma)$ which resembles the classical objective function's behavior; at least it should give the same notion of optimality as its classical analogue. In order to be more concrete, we will further assume that $U_M(\beta)$ and $U_P(\gamma)$ are given by (products of) exponentials of Hamiltonians. This assumption is, however, not very restrictive as most applications are characterized via Hamiltonians.

The correct definition of $U_M(\beta)$ follows naturally from the following convergence considerations and is given in section 4. We define the phase separator already now:

Definition 1. Given a COP with solution space $\mathcal{S}$ and optimal solution space $\mathcal{S}_{\max}$, a Hamiltonian H is called a *phase separator Hamiltonian* iff it fulfills the following two conditions:

(i) H is diagonal in the computational basis.
(ii) The eigenspace of $H|_{\mathcal{S}}$ corresponding to its largest eigenvalue is $\mathcal{S}_{\max}$.

The corresponding (*parametrized*) *phase separator* is given by

$$U_P(H,\gamma) := e^{i\gamma H}. \tag{13}$$

3. Convergence Proof for the QAA

We first examine the convergence behavior of the QAA. Although originally stated for unconstrained problems, we can easily extend the idea to a COP with a nontrivial solution space $\mathcal{S}$: The initial Hamiltonian H_I should preserve feasibility, i.e. $H(\mathcal{S}) \subseteq \mathcal{S}$, and the initial state $|\iota\rangle$ should lie within $\mathcal{S}$. In addition, we substitute the objective Hamiltonian C with a more general phase separator Hamiltonian H_P which trivially preserves feasibility. Then, for every $t \in [0,1]$, the time evolution with respect to $H_{\mathrm{lin}(H_I,H_P)}(t)$ applied to $|\iota\rangle$ will give again a feasible state. Thus, we effectively restrict ourselves to the subspace $\mathcal{S} \subseteq \mathcal{H}$.

The underlying concept of the QAA is captured by the adiabatic theorem. For our analysis, we use a more general version than Farhi *et al* did in [2].

Theorem 2 (Adiabatic Theorem, [11]). *Let $\{H(t) : 0 \leqslant t \leqslant 1\} \subseteq \mathcal{L}(\mathcal{H})$ be a family of self-adjoint operators such that $H(\cdot) \in C^2([0,1], \mathcal{L}(\mathcal{H}))$. For $T > 0$, let $\tilde{U}_T$ be the solution of*

$$\frac{\mathrm{d}}{\mathrm{d}s}\tilde{U}_T(s) = -iH(s/T)\,\tilde{U}_T(s), \quad 0 \leqslant s \leqslant T; \quad \tilde{U}_T(0) = \mathbb{1} \tag{14}$$

and set $U_T(t) := \tilde{U}_T(tT)$, $0 \leqslant t \leqslant 1$. Let $\lambda(t)$ be an eigenvalue of $H(t)$, respectively, with corresponding spectral projection $\mathbb{P}(t)$ [12]. Furthermore, let $P \in C^2([0,1], \mathcal{L}(\mathcal{H}))$ such that for every $0 \leqslant t \leqslant 1$, $P(t)$ is a projection with $H(t)P(t) = \lambda(t)P(t)$. In addition, $P(t) = \mathbb{P}(t)$ should hold for almost all $t \in [0,1]$. Then

$$\lim_{T\to\infty} (\mathbb{1} - P(t))\, U_T(t)\, P(0) = 0 \tag{15}$$

uniformly in t in $[0,1]$.

Theorem 2 essentially states that, in the adiabatic limit, starting within (a subspace of) the eigenspace of $H(0)$ corresponding to the eigenvalue $\lambda(0)$, one stays within the eigenspace of $H(t)$ corresponding to the eigenvalue $\lambda(t)$, $0 \leqslant t \leqslant 1$, if one follows the time evolution generated by H, and the curve of spectral projections P can be C^2-continued through all potential level crossings. In contrast, Farhi *et al* used a version of the adiabatic theorem that prohibits any level crossing (see [13]). For the reader's convenience, we provide a proof of theorem 2 in the appendix, closely following Teufel's derivation.

A sketch of a convergence proof for the QAA was given in [1] as an intermediate step to argue the convergence of the quantum approximate optimization algorithm. Besides the adiabatic theorem, the proof is mainly based on the Perron–Frobenius Theorem. First recall the definition of irrecudibility in the context of matrices.

Definition 3. A matrix $A \in \mathbb{C}^{n\times n}$ is called *irreducible* iff there are no proper A-invariant coordinate subspaces of $\mathbb{C}^n$. That is, the only coordinate subspaces left invariant by A are the *trivial* subspaces $\{0\}$ and $\mathbb{C}^n$.

Theorem 4 (Perron-Frobenius). *Let $A \in \mathbb{C}^{n\times n}$ be component-wisely non-negative and irreducible. Then A admits a nondegenerate largest eigenvalue.*

The crucial observation is that the matrix representation of the initial Hamiltonian (8) in the computational basis fulfills both requirements of the Perron–Frobenius Theorem. As this will also play an essential role throughout our convergence proof, we use these very properties for giving a first definition of a *mixer*.

Definition 5. A Hamiltonian $B \in \mathcal{L}(\mathcal{H})$ is called a *mixer* for a COP with solution space $\mathcal{S}$ iff $B(\mathcal{S}) \subseteq \mathcal{S}$ and $B|_{\mathcal{S}} \in \mathcal{L}(\mathcal{S})$ is component-wisely non-negative and irreducible in the computational basis.

The idea is now to apply the Perron–Frobenius Theorem to the linear interpolation $H_{\text{lin}(B,C)}(t)$ at every time $0 \leqslant t < 1$ to conclude the existence of an eigenvalue curve $\lambda_{\max}$ that connects both the largest eigenvalues of $H_{\text{lin}(B,C)}|_{\mathcal{S}}(0) = B|_{\mathcal{S}}$ and $H_{\text{lin}(B,C)}|_{\mathcal{S}}(1) = C|_{\mathcal{S}}$. For this, we need the following immediate result which can be proven quite easily.

Corollary 6. *Let $A \in \mathbb{C}^{n\times n}$ be diagonal and let $B \in \mathbb{C}^{n\times n}$ be irreducible. Then also $A + B$ is irreducible.*

Theorem 7 (Convergence of QAA). *Consider a COP with solution space $\mathcal{S} \subseteq \mathcal{H}$, optimal solution space $\mathcal{S}_{opt} \subseteq \mathcal{S}$, and phase separator Hamiltonian C. If $B \in \mathcal{L}(\mathcal{H})$ is a mixer Hamiltonian in the sense of definition 5 and $|\iota\rangle \in \mathcal{S}$ is a highest energy state of $B|_{\mathcal{S}}$, then*

$$\lim_{T\to\infty} U_T(1)\,|\iota\rangle \in \mathcal{S}_{opt}, \tag{16}$$

where U_T is the quasi-adiabatic evolution w.r.t. to the linear interpolation between B and C.

In the following proof, we directly identify all appearing operators $\mathcal{L}(\mathcal{S})$ with their matrix representation in the computational basis.

Proof. Let $\lambda_{\max}(t)$ denote the largest eigenvalue of $H_{\text{lin}(B,C)}|_{\mathcal{S}}(t)$, for $0 \leqslant t \leqslant 1$, respectively. Let $0 \leqslant t_0 < 1$. Since $B|_{\mathcal{S}}$ is irreducible, so is $(1-t_0)B|_{\mathcal{S}}$. As C is diagonal, also $H_{\text{lin}(B,C)}(t_0)|_{\mathcal{S}} = (1-t_0)B|_{\mathcal{S}} + t_0 C|_{\mathcal{S}}$ is irreducible by corollary 6. W.l.o.g. assume that C has non-negative spectrum [14]. Then, $B|_{\mathcal{S}}$ as well as $C|_{\mathcal{S}}$ are component-wisely non-negative. In summary, $H_{\text{lin}(B,C)}(t_0)|_{\mathcal{S}}$ is component-wisely non-negative and irreducible. According to the Perron–Frobenius Theorem, $\lambda_{\max}(t_0)$ is nondegenerate. Furthermore, the mapping

$$\begin{aligned} H_{\text{lin}(B,C)}|_{\mathcal{S}} : \mathbb{R} &\to \mathcal{L}(\mathcal{S}) \\ t &\mapsto H_{\text{lin}(B,C)}|_{\mathcal{S}}(t) = (1-t)\,B|_{\mathcal{S}} + tC|_{\mathcal{S}} \end{aligned} \tag{17}$$

is analytic and $H_{\text{lin}(B,C)}|_{\mathcal{S}}(t)$ is symmetric for all $t \in \mathbb{R}$. Let L denote the discrete set of level crossings of $H_{\text{lin}(B,C)}|_{\mathcal{S}}$ [15]. According to [16, theorem 6.1], the instantaneous eigenvalues of $H_{\text{lin}(B,C)}|_{\mathcal{S}}(t)$, $t \in \mathbb{R}$, can be sorted as $\{\lambda_m(t) : 1 \leqslant m \leqslant M,\ M \leqslant 2^N\}$, such that $[t \mapsto \lambda_m(t)]$ is analytic on whole $\mathbb{R}$ and, for every $1 \leqslant m \leqslant M$, the corresponding curves of spectral projections $[t \mapsto \mathbb{P}_m(t)]$ are analytic on $\mathbb{R} \setminus L$. Furthermore, the spectral projections have removable singularities in L, i.e. there exist analytic continuations P_m, defined on whole $\mathbb{R}$, such that $P_m(t) = \mathbb{P}_m(t)$ for $t \in \mathbb{R} \setminus L$, for all $1 \leqslant m \leqslant M$. By continuity, these continuations are themselves orthogonal projections with constant rank and fulfill

$$H_{\text{lin}(B,C)}|_{\mathcal{S}}(t)\,P_m(t) = \lambda_m(t)\,P_m(t)$$

for all $t \in \mathbb{R}$. W.l.o.g. assume $\lambda_1(0) = \lambda_{\max}(0)$. Since $\lambda_{\max}(t_0)$ remains nondegenerate for $0 \leqslant t_0 < 1$, it follows that $\lambda_1 \equiv \lambda_{\max}$ on $[0,1)$ and by continuity of λ_1 that $\lambda_1 \equiv \lambda_{\max}$ on $[0,1]$. In addition, the corresponding spectral projection $\mathbb{P}_1$ is well-defined on $[0,1)$. Therefore, its continuation P_1 fulfills all properties necessary to apply theorem 2, i.e. (15) holds. Since $P_1(0) = \mathbb{P}_1(0)$, one especially obtains that

$$\begin{aligned} 0 &= \lim_{T\to\infty} (\mathbb{1} - P_1(1))\,U_T(1)\,P_1(0)\,|\iota\rangle \\ &= \lim_{T\to\infty} (\mathbb{1} - P_1(1))\,U_T(1)\,|\iota\rangle \\ &\Leftrightarrow \lim_{T\to\infty} U_T(1)\,|\iota\rangle = P_1(1) \lim_{T\to\infty} U_T(1)\,|\iota\rangle. \end{aligned}$$

Since $P_1(1)$ is a projection with $H_{\text{lin}(B,C)}|_{\mathcal{S}}(1)P_1(1) = \lambda_{\max}(1)P_1(1)$, one concludes (16). □

Following the above proof, one realizes that the eigenvalue curve $\lambda_{\max}$ does not cross any other eigenvalue curve of $H_{\text{lin}(B,C)}|_{\mathcal{S}}$ except, possibly, at $t = 1$. In [1], even a level crossing at $t = 1$ is avoided by assuming that the COP only has one optimal solution, implying that $\lambda_{\max}(1)$ is nondegenerate. However, by invoking a more general version of the adiabatic theorem, we were able to get rid of this assumption.

4. Convergence Proof for the QAOA

We next examine the convergence behavior of the QAOA which contains the quantum approximate optimization algorithm as a special case. Its ingredients are basically the same as for our generalized version of the QAA. However, the decomposition of the mixer Hamiltonian into local Hamiltonians is extremely valuable from an application-oriented point of view and is also introduced by the QAOA. In the spirit of definition 5, we propose the following adaptation of Hadfield *et al*'s definition.

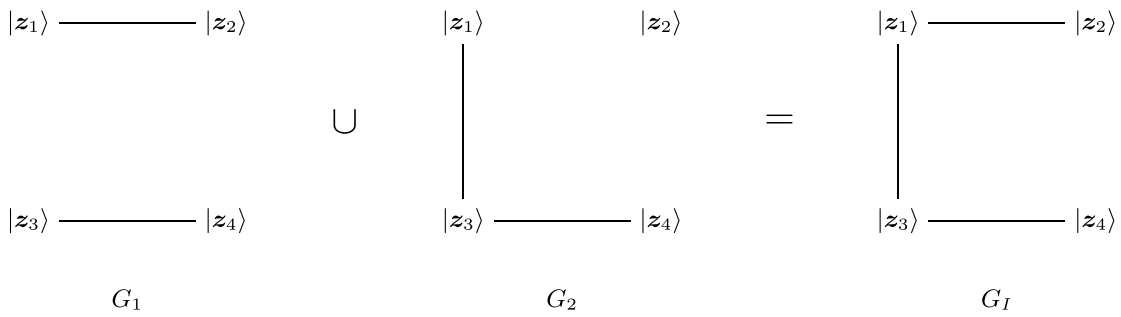


Figure 1. Illustration of the proof of proposition 9. The feasible subspace $\mathcal{S}$ is spanned by $|z_1\rangle, |z_2\rangle, |z_3\rangle$, and $|z_4\rangle$. The proper invariant coordinate subspaces of $B_1|_{\mathcal{S}}$ are span$\{|z_1\rangle, |z_2\rangle\}$ and span$\{|z_3\rangle, |z_4\rangle\}$ while $B_2|_{\mathcal{S}}$ has span$\{|z_1\rangle, |z_3\rangle, |z_4\rangle\}$ and span$\{|z_2\rangle\}$ as its proper invariant coordinate subspaces. Thus, the set of their mutual proper invariant coordinate subspaces is empty, hence $B_I = B_1 + B_2$ is a mixer.

Definition 8. Given a COP with solution space $\mathcal{S}$, a family of Hamiltonians $\{B_i\}_{i\in I} \subset \mathcal{L}(\mathcal{H})$—$I$ being an arbitrary index set—is called a *mixing family* iff for every $i \in I$, $B_i(\mathcal{S}) \subseteq \mathcal{S}$, $B_i|_{\mathcal{S}}$ is component-wise non-negative in the computational basis, and any coordinate subspace of $\mathcal{S}$ that is left invariant under every B_i is trivial.

That definition 8 really is a decomposed version of definition 5 can be argued as follows: Consider the matrix representation of each of the operators $B_i|_{\mathcal{S}}$ in the computational basis as adjacency matrix of a graph G_i whose vertices are identified with feasible computational basis states [17]. Starting from the graph G_1 resembled by B_1, adding another operator B_i corresponds to adding edges represented by nonzero entries of B_i's matrix representation. The actual weights (i.e. values of the entries) are not important, but the condition of component-wise non-negativity implies that no entries are cancelled during the summation, that is, the edge set of the graph G_I with adjacency matrix

$$B_I|_{\mathcal{S}}, \quad B_I := \sum_{i\in I} B_i \tag{18}$$

really is the union of all the edge sets of the graphs G_i with respective adjacency matrix $B_i|_{\mathcal{S}}$, $i \in I$. The imposed condition of triviality of mutual invariant coordinate subspaces then is equivalent to the fact that G_I is fully connected (see figure 1 for an example) which, in turn, is equivalent to its adjacency matrix being irreducible. Thus, we have concluded

Proposition 9. *Given a COP with solution space $\mathcal{S}$, a family of Hamiltonians $\{B_i\}_{i\in I} \subset \mathcal{L}(\mathcal{H})$ is a mixing family iff B_I, constructed as in (18), is a mixer Hamiltonian.*

Utilizing our definition of a mixing family, we now introduce our version of 'simultaneous' and 'sequential' mixers.

Definition 10. Let $\mathsf{H} = \{H_i\}_{i\in I} \subset \mathcal{L}(\mathcal{H})$ be a mixing family for a given COP. The corresponding (*parametrized*) *simultaneous mixer* is defined as

$$U_{\mathrm{M},0}(\mathsf{H}, \beta) := \mathrm{e}^{-\mathrm{i}\beta \sum_{i\in I} H_i}. \tag{19}$$

Specifying a permutation $\sigma \in S(I)$, the corresponding (*parametrized*) *sequential mixer* is defined as

$$U_{\mathrm{M},\sigma}(\mathsf{H}, \beta) := \prod_{i\in I} \mathrm{e}^{-\mathrm{i}\beta H_{\sigma(i)}}. \tag{20}$$

From their definition it immediately follows that both (19) and (20) fulfill the original QAOA demands: feasibility preservation and full mixing of solutions. However, due to our refined definition, we can now extend the sketch of a convergence proof in [1] to the general QAOA setting. The procedure is as follows:

1. discretize the quasi-adiabatic time evolution U_T
2. decompose $H_{\mathrm{lin}(B,C)}$ using a (multivariate) Lie product formula (see [18, theorem 2.11] and [19, problem IX.8.5])
3. exploit the convergence of the corresponding QAA instance

We start with a simple statement about the distance of products of operators with factors being close together.

Lemma 11. *Let $||\cdot||$ be any norm on $\mathcal{H}$. For $\varepsilon > 0$ and $m \in \mathbb{N}$, let $\{V_j\}_{j=1}^m$, $\{W_j\}_{j=1}^m, \subset \mathcal{L}(\mathcal{H})$ be families of unitary operators so that*

$$\left\|V_j - W_j\right\| < \varepsilon \tag{21}$$

holds for all $j \in [m]$. Then the following estimate is valid:

$$\left\|\prod_{j=1}^{m} V_j - \prod_{j=1}^{m} W_j\right\| < m\varepsilon. \tag{22}$$

Proof. Equation (22) clearly holds for $m = 1$. Therefore, it remains to show that if (22) holds for an $m \in \mathbb{N}$, then it also holds for $m+1$:

$$\begin{aligned}
\left\|\prod_{j=1}^{m+1} V_j - \prod_{j=1}^{m+1} W_j\right\| &= \left\|\left(\prod_{j=1}^{m+1} V_j\right) - V_{m+1}\left(\prod_{j=1}^{m} W_j\right) + V_{m+1}\left(\prod_{j=1}^{m} W_j\right) - \prod_{j=1}^{m+1} W_j\right\| \\
&\leqslant \left\|V_{m+1}\left(\prod_{j=1}^{m} V_j - \prod_{j=1}^{m} W_j\right)\right\| + \left\|(V_{m+1} - W_{m+1})\prod_{j=1}^{m} W_j\right\| \\
&\leqslant \left\|\prod_{j=1}^{m} V_j - \prod_{j=1}^{m} W_j\right\| + \|V_{m+1} - W_{m+1}\| \\
&< m\varepsilon + \varepsilon = (m+1)\varepsilon.
\end{aligned}$$

Hence, by induction, the claim is true for all $m \in \mathbb{N}$. □

Theorem 12 (Convergence of QAOA). *Consider a COP with solution space $\mathcal{S} \subseteq \mathcal{H}$, optimal solution space $\mathcal{S}_{opt} \subseteq \mathcal{S}$, phase separator Hamiltonian C, and mixing family $\{B_i\}_{i\in I}$. Let U_P and U_M be the corresponding phase separator and (simultaneous or sequential) mixer. Furthermore, let $|\iota\rangle \in \mathcal{S}$ be a highest energy state of $B_I|_{\mathcal{S}}$. Then, for every $\varepsilon > 0$, one can choose finitely many parameters $\vec{\beta}$ and $\vec{\gamma}$ such that*

$$\operatorname{dist}\left(|\vec{\beta},\vec{\gamma}\rangle, \mathcal{S}_{\text{opt}}\right) < \varepsilon, \tag{23}$$

where dist is the usual distance function in $\mathcal{H}$ w.r.t. to the norm $||\cdot||$ and

$$|\vec{\beta},\vec{\gamma}\rangle := V\left(\vec{\beta},\vec{\gamma}\right)|\iota\rangle := \left(\prod_{q} U_{\text{M}}\left(\beta_q\right) U_{\text{P}}\left(\gamma_q\right)\right)|\iota\rangle. \tag{24}$$

Proof. Let U_T, $T > 0$, denote the quasi-adiabatic evolution w.r.t. $H_{\text{lin}(B_I,C)}$. By proposition 9, B_I is a mixer Hamiltonian in the sense of definition 5. Therefore, for any $\varepsilon > 0$, theorem 7 implies the existence of a $T > 0$ so that

$$||(\mathbb{1} - P_1(1))\, U_T(1)\,|\iota\rangle||_{\mathcal{L}(\mathcal{S})} < \frac{\varepsilon}{2},$$

where $||\cdot||_{\mathcal{L}(\mathcal{S})}$ is the usual operator norm on $\mathcal{L}(\mathcal{H})$ and P_1 is the C^2-continuation of the curve of spectral projections onto the highest energy eigenspaces of $H_{\text{lin}(B_I,C)}|_{\mathcal{S}}$. W.l.o.g. assume that $\dim(\mathcal{S}) > 1$ as the statement would be trivial otherwise. Then, $\alpha := ||\mathbb{1} - P_1(1)||_{\mathcal{L}(\mathcal{S})} > 0$ since $P_1(1)$ has rank one by continuity. Following [20], discretizing the quasi-adiabatic time evolution $U_T(1)$ yields the existence of an $m \in \mathbb{N}$ such that

$$\left\|\prod_{j=1}^{m} \mathrm{e}^{-\mathrm{i}H_{\text{lin}(B_I,C)}\left(j\frac{T}{m}\right)\frac{T}{m}} - U_T(1)\right\|_{\mathcal{L}(\mathcal{S})} < \frac{\varepsilon}{4\alpha}. \tag{25}$$

In the following, set

$$W_j := \mathrm{e}^{-\mathrm{i}H_{\text{lin}(B_I,C)}\left(j\frac{T}{m}\right)\frac{T}{m}}$$

and distinguish between the two possibilities to choose a mixer.

Simultaneous mixer: The Lie product formula [18, theorem 2.11] implies that for all $j \in [m]$, there exist $n_j \in \mathbb{N}$ such that for all $\tilde{n} \geqslant n_j$ it holds that

$$\left\| \left(e^{-i\frac{1-j\frac{T}{m}}{\tilde{n}}\frac{T}{m}B_I} e^{-i\frac{j\left(\frac{T}{m}\right)^2}{\tilde{n}}C} \right)^{\tilde{n}} - W_j \right\|_{\mathcal{L}(\mathcal{S})} < \frac{\varepsilon}{4m\alpha}, \tag{26}$$

respectively. Taking $n := \max\{n_j : j \in [m]\}$, this estimate holds for all $j \in [m]$ and (especially) $\tilde{n} = n$.

Sequential mixer: W.l.o.g. choose the identity permutation $\sigma = \mathrm{id}_I$. The multivariate Lie product formula [19, problem IX.8.5] implies that for all $j \in [m]$, there exist $n_j \in \mathbb{N}$ so that for all $\tilde{n} \geqslant n_j$ it holds that

$$\left\| \left(\left(\prod_{i\in I} e^{-i\frac{\left(1-j\frac{T}{m}\right)}{\tilde{n}}\frac{T}{m}B_i} \right) e^{-i\frac{j\left(\frac{T}{m}\right)^2}{\tilde{n}}C} \right)^{\tilde{n}} - W_j \right\|_{\mathcal{L}(\mathcal{S})} < \frac{\varepsilon}{4m\alpha}. \tag{27}$$

In both these cases, choose $p = n^m$ parameter values for $\vec{\beta} = (\beta_{11}, \ldots, \beta_{1k}, \ldots, \beta_{1n}, \beta_{21}, \ldots, \beta_{mn})$ as well as for $\vec{\gamma} = (\gamma_{11}, \ldots, \gamma_{1k}, \ldots, \gamma_{1n}, \gamma_{21}, \ldots, \gamma_{mn})$ according to

$$\beta_{jk} = \frac{1-j\frac{T}{m}}{n}\frac{T}{m}, \quad \gamma_{jk} = \frac{j\left(\frac{T}{m}\right)^2}{n}$$

for all $k \in [n]$ and all $j \in [m]$. Then, by construction, (26) and (27) translate into

$$||V\left(\vec{\beta}_j, \vec{\gamma}_j\right) - W_j||_{\mathcal{L}(\mathcal{S})} < \frac{\varepsilon}{4m\alpha}.$$

Thus, by (25) and lemma 11, it follows that

$$\begin{aligned} \left\| V\left(\vec{\beta}, \vec{\gamma}\right) - U_T(t) \right\|_{\mathcal{L}(\mathcal{S})} &= \left\| \prod_{j=1}^{m} V\left(\vec{\beta}_j, \vec{\gamma}_j\right) - U_T(t) \right\|_{\mathcal{L}(\mathcal{S})} \\ &\leqslant \left\| \prod_{j=1}^{m} V\left(\vec{\beta}_j, \vec{\gamma}_j\right) - \prod_{j=1}^{m} W_j \right\|_{\mathcal{L}(\mathcal{S})} + \left\| \prod_{j=1}^{m} W_j - U_T(t) \right\|_{\mathcal{L}(\mathcal{S})} \\ &< \frac{\varepsilon}{4\alpha} + \frac{\varepsilon}{4\alpha} = \frac{\varepsilon}{2\alpha}. \end{aligned}$$

In summary, it follows that

$$\begin{aligned} \left\| \left(\mathbb{1} - P_1(1)\right) |\vec{\beta}, \vec{\gamma}\rangle \right\| &= ||\left(\mathbb{1} - P_1(1)\right) V\left(\vec{\beta}, \vec{\gamma}\right) |\iota\rangle|| \\ &\leqslant ||\left(\mathbb{1} - P_1(1)\right) \left(V\left(\vec{\beta}, \vec{\gamma}\right) - U_T(1)\right) |\iota\rangle|| \\ &\quad + ||\left(\mathbb{1} - P_1(1)\right) U_T(1) |\iota\rangle|| \\ &< ||\mathbb{1} - P_1(1)||_{\mathcal{L}(\mathcal{S})} \frac{\varepsilon}{2\alpha} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

The image of $P_1(1)$ lying in $\mathcal{S}_{\mathrm{opt}}$ proves the assertion. □

5. Conclusion and outlook

In this paper we presented an elementary proof for the convergence of the QAOA. This proof can be regarded as a discretized and carefully extended version of the Adiabatic Theorem, building on the ideas of Farhi *et al* Beside another core theorem (Perron–Frobenius), this extension is merely based on elementary matrix inequalities. Most importantly, our proof builds on fewer assumptions (multiple optimal solutions are allowed) and extends to nontrivial feasibility structures ($\mathcal{S} \subsetneq \mathcal{H}$).

Furthermore, the proof canonically gave rise to refined definitions of the QAOA-mixer and QAOA-phase separator concepts. Most notably, exactly the same notions arise when properly recreating classical feasibility symmetries within the framework of QAOA (see [21]). This strongly indicates that the definitions we gave in this paper optimally capture the overall principle the QAOA is based on.

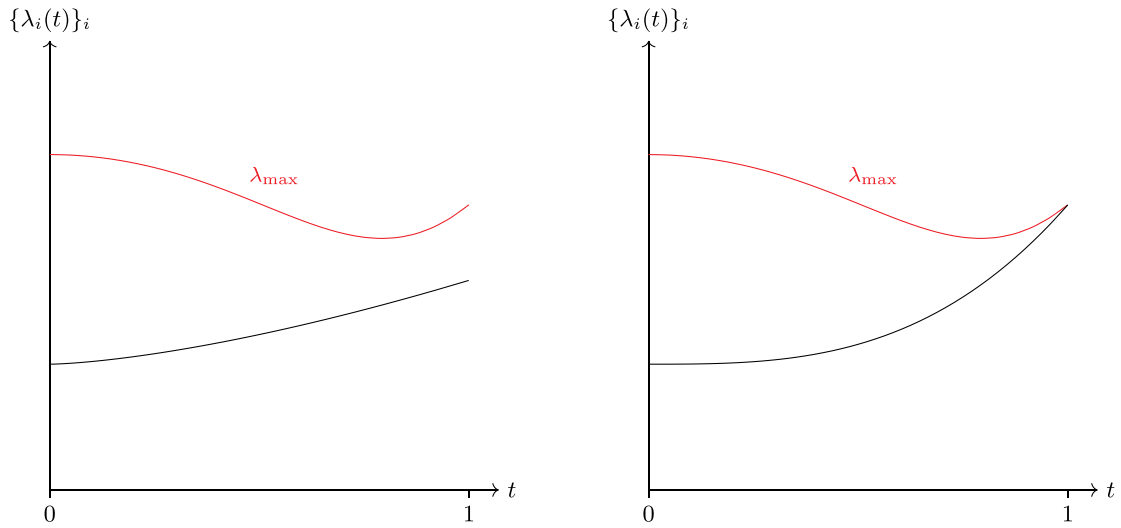


Figure 2. The eigenvalue curve $\lambda_{\max}$ (red line) stays separated from all the other eigenvalue curves (represented by a single black line) for $0 \leqslant t < 1$. If the corresponding COP has exactly one optimal solution the separation extends to $t = 1$ (left plot). However, if the COP has multiple optimal solutions $\lambda_{\max}$ intersects with at least one other eigenvalue curve at $t = 1$ (right plot).

We essentially showed that irreducibility and component-wise non-negativity of the mixer Hamiltonian B, restricted to the feasible subspace $\mathcal{S}$, are sufficient criteria for the convergence of the QAA and the QAOA. Moreover, one can readily verify that irreducibility is also a necessary condition in the following sense: Given an arbitrary initial state $|\iota\rangle$ and the existence of a nontrivial $B|_{\mathcal{S}}$-invariant coordinate subspace, there always exists an objective Hamiltonians C such that the QAA and the QAOA will not be able to approximate any state in $\mathcal{S}_{\max}$ to arbitrary precision. On the other hand, the condition that $B|_{\mathcal{S}}$ should be component-wise non-negative is not necessary. In our convergence proof, we imposed this condition in order to apply the Perron–Frobenius Theorem. However, there also exist more general versions of this theorem (see, e.g. [22]) which substitute this condition and irreducibility with the more general properties of preserving a given cone and permuting its faces, respectively. Unfortunately, the cones in questions are merely given by all the orthants in $\mathcal{S}$ since every coordinate subspace of $\mathcal{S}$ should be represented on their faces. This yields again, up to some additionally allowed matrix signatures, the same conditions. Therefore, we do not see much possibilities for relaxing the assumptions, made in theorems 7 and 12.

An interesting and still remaining question is whether one can also characterize the rate of convergence of the QAA for the case of multiple optimal solutions. In this case, the spectral gap is necessarily vanishing for $t \to 1$ (see figure 2), but stays finite throughout the interval $[0, 1)$. That is, even though a level crossing occurs, it only happens once and at a predictable time. There are some results on the rate of convergence in the Adiabatic Theorem which are valid for all kinds of (allowed) level crossings (see, e.g. [11, 23]). Fine-tuning these results with respect to the particular situation of the QAA promises to be an insightful future project.

Data availability statement

No new data were created or analysed in this study.

Acknowledgment

We thank Tim Heine, Lauritz van Luijk, Tobias J Osborne, Christoph Pohl, Antonio Rotundo, and Martin Steinbach for helpful discussions. L B acknowledges financial support by the Quantum Valley Lower Saxony. L B and T Z acknowledge financial support by the BMBF project QuBRA. G K acknowledges financial support by the DAAD and IIT Indore (Kapil Ahuja) for a guest stay. RS acknowledges financial support by the Quantum Valley Lower Saxony, by the Cluster of Excellence 'Quantum Fron- tiers' (DFG 390837967), and by the BMBF Projects ATIQ and CBQD.

Appendix. Proof of theorem 2

For fixed $T > 0$, let V_T be the solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} V_T(t) = -\mathrm{i}T H_T(t)\, V_T(t)\,,\ 0 \leqslant t \leqslant 1;\ V_T(0) = \mathbb{1} \tag{A1}$$

with the time-dependent Hamiltonian

$$H_T(t) := H(t) + \frac{\mathrm{i}}{T}\left[\dot{P}(t), P(t)\right], \tag{A2}$$

parametrized by T. Using $U_T^\dagger(0) = V_T(0) = U_T(t)U_T^\dagger(t) = \mathbb{1}$, we obtain

$$\begin{aligned} U_T(t) - V_T(t) &= -U_T(t)\left(U_T^\dagger(t)\,V_T(t) - U_T^\dagger(0)\,V_T(0)\right) = -U_T(t)\int_0^t \mathrm{d}s \frac{\mathrm{d}}{\mathrm{d}s}\left(U_T^\dagger(s)\,V_T(s)\right) \\ &= -U_T(t)\int_0^t \mathrm{d}s\, U_T^\dagger(s)\left(-\mathrm{i}TH(s) + \left[\dot{P}(s), P(s)\right] + \mathrm{i}TH(s)\right)V_T(s) \\ &= -U_T(t)\int_0^t \mathrm{d}s\, U_T^\dagger(s)\left[\dot{P}(s), P(s)\right]V_T(s). \end{aligned} \tag{A3}$$

Since $H(t)$ is hermitian for all $t \in [0,1]$, $H(t) - (\lambda(t) - \mathrm{i}\Delta)\mathbb{1}$ is always invertible as long as $\Delta > 0$. Therefore, we can consider the resolvent $R(\lambda(t) - \mathrm{i}\Delta, t) := (H(t) - (\lambda(t) - \mathrm{i}\Delta)\mathbb{1})^{-1}$. Note that its adjoint is given by

$$R(\lambda(t) - \mathrm{i}\Delta, t)^\dagger = \left(H(t)^\dagger - \left(\overline{\lambda(t) - \mathrm{i}\Delta}\right)\mathbb{1}\right)^{-1} = (H(t) - (\lambda(t) + \mathrm{i}\Delta))^{-1} = R(\lambda(t) + \mathrm{i}\Delta, t).$$

Furthermore, note that by $P(t) = P(t)^\dagger$ implies that also $\dot{P}(t)^\dagger = \dot{P}(t)$. For a fixed $\Delta > 0$, we define

$$\begin{aligned} A_\Delta(t) &:= R(\lambda(t) - \mathrm{i}\Delta, t)\,\dot{P}(t)\,P(t) \\ X_\Delta(t) &:= A_T(t) + A_T(t)^\dagger = R(\lambda(t) - \mathrm{i}\Delta, t)\,\dot{P}(t)\,P(t) + P(t)\,\dot{P}(t)\,R(\lambda(t) + \mathrm{i}\Delta, t) \\ Y_\Delta(t) &:= \mathrm{i}\Delta A_T(t) + \mathrm{i}\Delta A_T(t)^\dagger = \mathrm{i}\Delta R(\lambda(t) - \mathrm{i}\Delta, t)\,\dot{P}(t)\,P(t) + \mathrm{i}\Delta P(t)\,\dot{P}(t) R(\lambda(t) + \mathrm{i}\Delta, t). \end{aligned}$$

For the next step, we make use of the fact that $H(t)P(t) = P(t)H(t) = \lambda(t)P(t)$ as well as

$$H(t)\,R(\lambda(t) \pm \mathrm{i}\Delta, t) = R(\lambda(t) \pm \mathrm{i}\Delta, t)\,H(t) = \mathbb{1} + (\lambda(t) \pm \mathrm{i}\Delta)\,R(\lambda(t) \pm \mathrm{i}\Delta, t)$$

hold for all $t \in [0,1]$. Visually hiding the dependency on t, we obtain

$$\begin{aligned} [H, A_\Delta] &= R(\lambda - \mathrm{i}\Delta)\left[H, \dot{P}\right]P = R(\lambda - \mathrm{i}\Delta)\,H\dot{P}P - R(\lambda - \mathrm{i}\Delta)\,\dot{P}HP = \dot{P}P + \lambda A_\Delta - \mathrm{i}\Delta A_\Delta - \lambda A_\Delta = \dot{P}P - \mathrm{i}\Delta A_\Delta \\ \left[H, A_\Delta^\dagger\right] &= \left[H^\dagger, A_\Delta^\dagger\right] = [H, A_\Delta]^\dagger = P\dot{P} + \mathrm{i}\Delta A_\Delta^\dagger. \end{aligned}$$

In summary, we obtain the commutator equation

$$[H, X_\Delta] = \left[H, A_\Delta + A_\Delta^\dagger\right] = \dot{P}P - P\dot{P} - \mathrm{i}\Delta A_\Delta - \mathrm{i}\Delta A_\Delta^\dagger = \left[\dot{P}, P\right] - Y_\Delta \iff \left[\dot{P}, P\right] = [H, X_\Delta] + Y_\Delta. \tag{A4}$$

Substituting the right-hand side of (A4) into the integrand in (A3), we find that

$$U_T^\dagger\left[\dot{P}, P\right]V_T = U_T^\dagger H X_\Delta V_T - U_T^\dagger X_\Delta H V_T + U_T^\dagger Y_\Delta V_T.$$

This can be further rewritten since

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\left(U_T^\dagger X_\Delta V_T\right) = \mathrm{i}TU_T^\dagger H X_\Delta V_T + U_T^\dagger \dot{X}_\Delta V_T - \mathrm{i}TU_T^\dagger X_\Delta H V_T + U_T^\dagger X_\Delta\left[\dot{P}, P\right]V_T \\ \implies &U_T^\dagger\left[\dot{P}, P\right]V_T = -\frac{\mathrm{i}}{T}\frac{\mathrm{d}}{\mathrm{d}s}\left(U_T^\dagger X_\Delta V_T\right) + \frac{\mathrm{i}}{T}\left(U_T^\dagger X_\Delta\left[\dot{P}, P\right]V_T + U_T^\dagger \dot{X}_\Delta V_T\right) + U_T^\dagger Y_\Delta V_T. \end{aligned} \tag{A5}$$

Using (A5) within (A3), taking the operator norm on both sides, applying triangle and Minkowki's inequality, and finally taking the supremum over all $t \in [0,1]$ yield

$$\begin{aligned} \sup_{t\in[0,1]} ||U_T(t) - V_T(t)||_{\mathcal{L}(\mathcal{H})} &\leqslant \sup_{t\in[0,1]} \frac{\|X_\Delta(t)\|_{\mathcal{L}(\mathcal{H})}}{T} \\ &\quad + \int_0^1 \mathrm{d}s \left(\frac{\|X_\Delta(s)\left[\dot{P}(s), P(s)\right]\|_{\mathcal{L}(\mathcal{H})} + \|\dot{X}_\Delta(s)\|_{\mathcal{L}(\mathcal{H})}}{T} + \|Y_\Delta(s)\|_{\mathcal{L}(\mathcal{H})}\right) \\ &\leqslant \sup_{t\in[0,1]} \frac{\left(1 + 2\|\dot{P}(t)\|_{\mathcal{L}(\mathcal{H})}\right)\|X_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} + \|\dot{X}_\Delta(t)\|_{\mathcal{L}(\mathcal{H})}}{T} + \int_0^1 \mathrm{d}s \|Y_\Delta(s)\|_{\mathcal{L}(\mathcal{H})}, \end{aligned} \tag{A6}$$

where we already used in the first step that the unitary operators $U_T^\dagger(s)$ and $V_T(s)$ are norm-preserving. For the second inequality we used norm inequality together with the fact that the projections $P(s)$ are

norm-nonincreasing. Next, using the fact that $P(\cdot)$ and $\dot{P}(\cdot)$ are continuous on the compact set $[0,1]$ and that the distance between $\lambda(t) \pm i\Delta$ to purely real spectrum of $H(t)$ is always Δ, we obtain the bounds

$$\sup_{t\in[0,1]} \|X_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} \leqslant \frac{C}{\Delta} \text{ and } \sup_{t\in[0,1]} \|Y_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} \leqslant \frac{C}{\Delta} \tag{A7}$$

for some constant $C < \infty$. Furthermore, as

$$\frac{\mathrm{d}}{\mathrm{d}s} R(\lambda(t) - \mathrm{i}\Delta, t) = -R(\lambda(t) - \mathrm{i}\Delta, t)\left(\dot{H}(t) - \dot{\lambda}(t)\mathbb{1}\right) R(\lambda(t) - \mathrm{i}\Delta, t),$$

we obtain via triangle and norm inequality together with $P(\cdot)$, $\dot{P}(\cdot)$, $\ddot{P}(\cdot)$, $\dot{H}(\cdot)$, and $\dot{\lambda}(\cdot)$ being continuous on the compact set $[0,1]$ that

$$\sup_{t\in[0,1]} \|\dot{X}(t)\|_{\mathcal{L}(\mathcal{H})} \leqslant 2\frac{C_1}{\Delta^2} + 2\frac{C_2}{\Delta} \leqslant \frac{C}{\Delta^2}, \tag{A8}$$

where we can choose C the same as in (A7). Additionally, observe that since $P(t) = P^2(t)$, differentiating yields

$$\begin{aligned}
&\dot{P}(t) = \dot{P}(t)P(t) + P(t)\dot{P}(t) = \left(\dot{P}(t)P(t) + P(t)\dot{P}(t)\right)P(t) + P(t)\dot{P}(t) \\
&= \dot{P}(t)P(t) + P(t)\dot{P}(t) + P(t)\dot{P}(t)P(t) \\
\implies & P(t)\dot{P}(t)P(t) = 0 \\
\implies & \dot{P}(t) = \dot{P}(t)P(t) + P(t)\dot{P}(t) - 2P(t)\dot{P}(t)P(t) = \left(\mathbb{1} - P(t)\right)\dot{P}(t)P(t) + P(t)\dot{P}(t)\left(\mathbb{1} - P(t)\right) \qquad \text{(A9)} \\
\implies & \dot{P}(t)P(t) = (\mathbb{1} - P(t))\dot{P}(t)P(t). \qquad \text{(A10)}
\end{aligned}$$

Note that $H(t) - (\lambda(t) - \mathrm{i}\Delta)\mathbb{1}$ remains invertible on the orthogonal complement of the eigenspace corresponding to $\lambda(t)$ even for $\Delta \to 0$. By assumptions, for almost all $t \in [0,1]$ it holds that $P(t)$ is the projection onto that eigenspace, i.e. $(\mathbb{1} - P(t))$ projects onto the orthogonal complement. Therefore, by (A10),

$$\sup_{\Delta\in(0,1]} \|A_\Delta(t)|\psi\rangle\|_{\mathcal{H}} = \sup_{\Delta\in(0,1]} \|R(\lambda(t) - \mathrm{i}\Delta, t)(\mathbb{1} - P(t))\dot{P}(t)P(t)|\psi\rangle\|_{\mathcal{H}} < \infty$$

for all $|\psi\rangle \in \mathcal{H}$ and almost all $t \in [0,1]$. By continuity of $A_\Delta(\cdot)$, we can take the supremum over all $t \in [0,1]$ and remain finite:

$$\sup_{\substack{t\in[0,1]\\ \Delta\in(0,1]}} \|A_\Delta(t)|\psi\rangle\|_{\mathcal{H}} < \infty, \text{for all } |\psi\rangle \in \mathcal{H}.$$

By the uniform boundedness principle, this yields that

$$\sup_{\substack{t\in[0,1]\\ \Delta\in(0,1]}} \|A_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} \implies \sup_{\substack{t\in[0,1]\\ \Delta\in(0,1]}} \|Y_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} \tag{A11}$$

Additionally, we infer from (A10) that

$$\begin{aligned}
\lim_{\Delta\to 0} \|\mathrm{i}\Delta A_\Delta(t)|\psi\rangle\|_{\mathcal{H}} &= \lim_{\Delta\to 0} \|\mathrm{i}\Delta R(\lambda(t) - \mathrm{i}\Delta, t)|\phi\rangle\|_{\mathcal{H}} \\
&= \lim_{\Delta\to 0} \int_{\mathbb{R}} \mu_{|\phi\rangle}(\mathrm{d}r) \frac{\Delta^2}{(r - \lambda(t))^2 + \Delta^2} = \mu_{|\phi\rangle}(\lambda(t)),
\end{aligned}$$

where $|\phi\rangle := \dot{P}(t)P(t)|\psi\rangle$ and $\mu_{|\phi\rangle}$ denotes the spectral measure of $H(t)$ for $|\phi\rangle$. Now, for almost all $t \in [0,1]$ we have that $P(t) = \mathbb{P}(t)$ and therefore $\mu_{|\phi\rangle}(\lambda(t)) = 0$ as $|\phi\rangle$ lies in the orthogonal complement of the eigenspace of $H(t)$ associated to $\lambda(t)$. We therefore deduce that

$$\lim_{\Delta\to 0} \|i\Delta A_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} = 0 \implies \lim_{\Delta\to 0} \|Y_\Delta(t)\|_{\mathcal{L}(\mathcal{H})} = 0. \tag{A12}$$

By choosing a sequence $\Delta = \Delta(T)$ which obeys

$$\lim_{T\to\infty} \Delta(T) = 0 \text{ and } \lim_{T\to\infty} T\Delta^2(T) = \infty,$$

Equations (A7) and (A8) imply that the first summand on the right-hand side of (A6) vanishes for $T \to \infty$. Second, (A11) and (A12) allow for the application of dominated convergence to the remaining summand and show that it also vanishes for $T \to \infty$. In total, we obtain that

$$\lim_{T\to\infty} \|U_T(t) - V_T(t)\|_{\mathcal{L}(\mathcal{H})} = 0 \quad \text{uniformly in } t \in [0,1]\,. \tag{A13}$$

Furthermore, note that by (A9) and the fact that H_T and $P(t)$ commute by assumption, it holds that

$$\dot{P}(t) = \dot{P}(t)P(t) - P(t)\dot{P}(t)P(t) + P(t)\dot{P}(t) - P(t)\dot{P}(t)P(t) = \left[\left[\dot{P}(t), P(t)\right], P(t)\right] = -iT[H_T(t), P(t)]\,.$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(V_T^\dagger(t)P(t)V_T(t)\right) = \mathrm{i}TV_T^\dagger(t)H_T(t)P(t)V_T - \mathrm{i}TV_T^\dagger(t)[H_T(t), P(t)]V_T(t) - \mathrm{i}TV_T^\dagger P(t)H_T(t)V_T(t) = 0.$$

Hence the identity $V_T^\dagger(t)P(t)V_T(t) = P(0)$ holds true for all $t \in [0,1]$ as it is clearly fulfilled for $t = 0$ and the derivative of the left-hand side vanishes. Finally, this allows to deduce the claim from the stronger result (A13).

ORCID iDs

Lennart Binkowski https://orcid.org/0009-0001-3723-7699
René Schwonnek https://orcid.org/0000-0002-9055-5266

References

[1] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm (arXiv:1411.4028 [quant-ph])
[2] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 Quantum computation by adiabatic evolution (arXiv:0001106 [quant-ph])
[3] Kadowaki T and Nishimori H 1998 *Phys. Rev.* E **58** 5355
[4] Hauke P, Katzgraber H G, Lechner W, Nishimori H and Oliver W D 2020 *Rep. Prog. Phys.* **83** 054401
[5] Baker J S and Radha S K 2022 Wasserstein solution quality and the quantum approximate optimization algorithm: a portfolio optimization case study (arXiv:2202.06782 [quant-ph])
[6] de la Grand'rive P D and Hullo J-F 2019 Knapsack problem variants of QAOA for battery revenue optimisation (arXiv:1908.02210 [quanth-ph])
[7] van Dam W, Eldefrawy K, Genise N and Parham N 2021 Quantum optimization heuristics with an application to knapsack problems (arXiv:2108.08805 [quant-ph])
[8] Hadfield S, Wang Z, O'Gorman B, Rieffel E, Venturelli D and Biswas R 2019 *Algorithms* **12** 34
[9] Morita S and Nishimori H 2008 *J. Math. Phys.* **49** 125210
[10] There exist also more sophisticated convex combinations, where the coefficients in front of H_I and C are nonlinear functions of t (see [24]). Allowing for such problem-specific coefficients can improve the convergence rate of the QAA significantly.
[11] Teufel S 2001 *Lett. Math. Phys.* **58** 261–6
[12] The spectral projection $\mathbb{P}$ corresponding to an eigenvalue λ of a given matrix operator A is defined as the projection of the eigenspace of A associated with λ.
[13] Messiah A 1976 *Quantum Mechanics* vol II (Wiley)
[14] Otherwise consider $\tilde{B} := B - \lambda_{\min}\mathbb{1}$ and $\tilde{C} = C - \lambda_{\min}\mathbb{1}$, where $\lambda_{\min} < 0$ is the smallest eigenvalue of C. Then $H_{\mathrm{lin}}(\tilde{B}, \tilde{C}) = H_{\mathrm{lin}(B,C)} - \lambda_{\min}\mathbb{1}$ generates the same time evolution as $H_{\mathrm{lin}(B,C)}$ up to a global phase.
[15] Since the eigenvalue curves of $H_{\mathrm{lin}(B,C)|_S}$ are analytic according to [16, theorem 6.1] they can only mutually intersect on a discrete set. Since there are only finitely many eigenvalue curves the set of all level crossings $L := \{t \in [0,1] : \lambda_{m_1}(t) = \lambda_{m_2}(t), m_1 \neq m_2$ is also discrete.
[16] Kato T 1995 *Pertubation Theory for Linear Operators* (Springer)
[17] Given a graph $G = (V, E)$ and an enumeration $v_1, \ldots, v_n$ of its vertices, the corresponding adjacency matrix is defined via $a_i j = 1$ if $(v_i, v_j) \in E$ and $a_i j = 0$ otherwise.
[18] Hall B C 2015 *Lie Groups, Lie Algebras and Representations* (Springer)
[19] Bhatia R 1997 *Matrix Analysis* (Springer)
[20] Jerome J W and Polizzi E 2014 *Appl. Anal.* **93** 2574
[21] Koßmann G, Binkowski L, Tutschku C and Schwonnek R 2022 Open-shop scheduling with hard constraints (arXiv:2211.05822 [quant-ph])
[22] Schneider H and Vidyasagar M 1970 *SIAM J. Numer. Anal.* **7** 508
[23] Avron J E and Elgart A 1999 *Commun. Math. Phys.* **203** 445
[24] Roland J and Cerf N J 2002 *Phys. Rev.* A **65** 042308