

The 2-divergence and mixing times of quantum Markov processes

K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete

Citation: *Journal of Mathematical Physics* **51**, 122201 (2010); doi: 10.1063/1.3511335

View online: <http://dx.doi.org/10.1063/1.3511335>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jmp/51/12?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

Accurate and highly efficient calculation of the highly excited pure OH stretching resonances of O (1 D) HCl , using a combination of methods

J. Chem. Phys. **121**, 4467 (2004); 10.1063/1.1779577

Converged quantum calculations of HO 2 bound states and resonances for J=6 and 10

J. Chem. Phys. **120**, 9583 (2004); 10.1063/1.1711811

Fully relativistic coupled-cluster static dipole polarizabilities of the positively charged alkali ions from Li + to 119 +

J. Chem. Phys. **116**, 172 (2002); 10.1063/1.1420747

Ga + , In + , and Tl + impurities in alkali halide crystals: Distortion trends

J. Chem. Phys. **113**, 8680 (2000); 10.1063/1.1317520

A coupled cluster study of the 1 1 A 1g and 1 1 B 2u states of benzene

J. Chem. Phys. **108**, 3987 (1998); 10.1063/1.475801

An advertisement for the journal 'Computing: Science & Engineering'. The background shows a row of computer monitors in a library or office setting, each displaying the journal's cover. The cover features a colorful, abstract image of a galaxy or nebula. In the bottom right corner, the journal's logo 'computing SCIENCE & ENGINEERING' is displayed. Below the monitors, the text 'AIP's JOURNAL OF COMPUTATIONAL TOOLS AND METHODS. AVAILABLE AT MOST LIBRARIES.' is written in large, white, bold letters.

The χ^2 -divergence and mixing times of quantum Markov processes

K. Temme,^{1,a)} M. J. Kastoryano,^{2,3} M. B. Ruskai,⁴ M. M. Wolf,² and F. Verstraete¹

¹*Faculty of Physics, University of Vienna, Boltzmannngasse 5, A-1090 Vienna, Austria*

²*Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*

³*Niels Bohr International Academy, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*

⁴*Department of Mathematics, Tufts University, Medford, Massachusetts 02155, USA*

(Received 20 June 2010; accepted 13 October 2010; published online 14 December 2010)

We introduce quantum versions of the χ^2 -divergence, provide a detailed analysis of their properties, and apply them in the investigation of mixing times of quantum Markov processes. An approach similar to the one presented in the literature for classical Markov chains is taken to bound the trace-distance from the steady state of a quantum processes. A strict spectral bound to the convergence rate can be given for time-discrete as well as for time-continuous quantum Markov processes. Furthermore, the contractive behavior of the χ^2 -divergence under the action of a completely positive map is investigated and contrasted to the contraction of the trace norm. In this context we analyze different versions of quantum detailed balance and, finally, give a geometric conductance bound to the convergence rate for unital quantum Markov processes. © 2010 American Institute of Physics. [doi:10.1063/1.3511335]

I. INTRODUCTION

The mixing time of a classical Markov chain is the time it takes for the chain to be close to its steady state distribution, starting from an arbitrary initial state. The ability to bound the mixing time is important, for example in the field of computer science, where the bound can be used to give an estimate for the running time of some probabilistic algorithm such as the Monte Carlo algorithm. The mixing time for a classical Markov process P_{ij} , with $\sum_i P_{ij} = 1$ on the space of probability measures \mathcal{S} is commonly defined in terms of the one norm, $\|p\|_1 = \sum_i |p_i|$. Let π denote the fixed point of the classical Markov process, i.e., $P\pi = \pi$, then the mixing time is defined as

$$t_{mix}(\epsilon) = \min \{n \mid \forall q \in \mathcal{S}, \|P^n q - \pi\|_1 < \epsilon\}. \quad (1)$$

A large set of tools has emerged over the years that allows to investigate the convergence rate of classical Markov chains.⁴ One of the most prominent approaches¹⁻³ to bounding the mixing time of a Markov chain is based on the χ^2 -divergence⁵. This divergence is defined for two probability distributions $p, q \in \mathcal{S}$ as

$$\chi^2(p, q) = \sum_i \frac{(p_i - q_i)^2}{q_i}. \quad (2)$$

The usefulness of the χ^2 -divergence for finding bounds to the mixing time of classical Markov chains arises from the fact that it serves as an upper bound to the one norm difference between two probability distributions,

$$\|p - q\|_1^2 \leq \chi^2(p, q) \quad (3)$$

^{a)}Author to whom correspondence should be addressed. Electronic mail: kristan.temme@univie.ac.at.

and allows for an easier access to the spectral properties of the Markov chain. The χ^2 -divergence is intimately related to the Kullback–Leibler divergence, or relative entropy, $H(p, q) = \sum_i p_i (\log p_i - \log q_i)$. In fact, it can be obtained directly from the relative entropy as the approximating quadratic form, i.e., as the Hessian, of the latter:

$$\chi^2(p, q) = -\frac{\partial^2}{\partial_\alpha \partial_\beta} H(q + \alpha(p - q), q + \beta(p - q)) \Big|_{\alpha=\beta=0}. \quad (4)$$

The χ^2 divergence was first introduced by Karl Pearson in the context of statistical inference tests, the most widely used of which is the “Pearson’s χ^2 test.” Its computational simplicity and its clear relation to other distance measures have made it one of the most studied divergence measures in the literature.

In this paper, we find convergence bounds for arbitrary quantum Markov chains, also called quantum channels, with a technique that can be seen as a generalization of the work of ^{1–3} to noncommutative probability spaces. A prototypical example of mixing time in physics is the decoherence time of the underlying quantum process, i.e., the time in which quantum states decohere to an (often classical) mixture given a specific underlying noise model. The ability to bound the mixing time for quantum processes also turns out to be relevant when one seeks to give bounds on the runtime of quantum algorithms that are based on quantum Markov chains.^{6,7} Other applications of such bounds can be found in the framework of matrix product states,^{8,9} where the correlation length of the quantum state is connected to the convergence of the corresponding transfer operator that can be interpreted as a quantum channel. In this article, we introduce the mathematical framework necessary to extend the classical mixing time results to the quantum setting. In particular, we introduce a new divergence measure—the quantum χ^2 -divergence—for quantum states and use it to obtain some basic convergence bounds that mirror existing classical ones. Furthermore, we extend the classical concept of detailed balance to the quantum setting and discuss its relevance in general terms.

The paper is organized as follows: the remainder of Sec. I is devoted to setting the notation and to recalling the framework of quantum channels. Then in Sec. II, we introduce the quantum χ^2 -divergence, and prove some basic properties relating it to other divergence and distance measures. In particular, we focus on a specific subfamily of interest. In Sec. III, we consider contraction of the χ^2 -divergence under the action of a channel, and relate it to trace-norm contraction. Furthermore, we prove some fundamental quantum mixing time results, whose classical analogues are well known. In Sec. IV, we study quantum detailed balance, and in Sec. V, we extend an important classical geometric mixing bound (Cheeger’s inequality) to the quantum setting. Conclusions are drawn in Sec. VI.

A. Formal setting and notation

Throughout this paper we will consider linear maps from the complex d -dimensional matrix algebra \mathcal{M}_d to itself. The states are density matrices $\rho \in \mathcal{S}_d$, where $\mathcal{S}_d = \{\rho \in \mathcal{M}_d \mid \rho = \rho^\dagger, \rho \geq 0, \text{tr}[\rho] = 1\}$, acting on $\mathfrak{H} = \mathbb{C}^d$. The set of pure states is denoted \mathcal{S}_d^1 , while the set of positive definite states is denoted \mathcal{S}_d^+ . Note that \mathcal{M}_d itself becomes a Hilbert space when equipped with the standard Hilbert–Schmidt scalar product $\langle A \mid B \rangle \equiv \text{tr}[A^\dagger B]$; this Hilbert space is naturally isomorphic to $\mathcal{M}_d \simeq \mathbb{C}^{d^2}$. The eigenvalues and singular values of T are understood in terms of the matrix representation $\hat{T} \in \mathcal{M}_{d^2}$ of T acting on \mathbb{C}^{d^2} . The matrix representation of a quantum operation, which will always be written with a hat (ex. \hat{T}), is given in terms of some complete orthonormal basis $\{F_i\}_{i=1\dots d^2}$ of \mathcal{M}_d , where its matrix elements are $\hat{T}_{ij} = \langle F_i \mid T \mid F_j \rangle$. Unless otherwise specified, we consider the basis of matrix units. The distance between states $\rho_1, \rho_2 \in \mathcal{S}_d$ will be measured in terms of the trace distance $\|\rho_1 - \rho_2\|_1$ induced by the trace norm $\|A\|_1 = \text{tr}[\sqrt{A^\dagger A}] = \sum_i s_i(A)$ for $s_i(A)$ the singular values of $A \in \mathcal{M}_d$. Time discrete quantum Markovian dynamics are described by completely positive, trace-preserving maps (cpt-maps, or quantum channels)¹⁰ $T : \mathcal{M}_d \mapsto \mathcal{M}_d$. Due to the Kraus representation theorem, every cp-map can be expressed in terms of the Kraus

operators $A_\mu \in M_d$, as

$$T(\rho) = \sum_{\mu} A_\mu \rho A_\mu^\dagger \quad \text{with,} \quad \sum_{\mu} A_\mu^\dagger A_\mu = \mathbb{1}. \quad (5)$$

The dual map T^* can be seen as the hermitian conjugate of T with respect to the Hilbert–Schmidt scalar product. In the above matrix representation of the map, this corresponds exactly to taking the hermitian conjugate \hat{T}^\dagger . A quantum channel is called unital, or doubly stochastic, when the dual map T^* is also trace preserving. This immediately implies that T has a fixed point $\sigma = \mathbb{1}/\text{tr}[\mathbb{1}]$. We will also consider time continuous quantum Markov processes, described by a one-parameter semi-group

$$T_t(\rho) = e^{t\mathcal{L}}(\rho). \quad (6)$$

The Liouvillian \mathcal{L} , obeys $\mathcal{L}^*(\mathbb{1}) = 0$,¹¹ and its matrix representation $\hat{\mathcal{L}}$ is obtained as in the time-discrete case. In this paper, we shall often consider *primitive* quantum channels, i.e., channels with a unique maximal rank fixed point, and a unique eigenvalue of magnitude 1 (Ref. 12).

II. THE QUANTUM χ^2 -DIVERGENCE

We want to define a generalization of the classical χ^2 -divergence to the case when we are working on spaces with noncommuting density matrices. We shall require that any generalization to the setting of density matrices satisfies the condition that when the inputs are diagonal, the classical χ^2 -divergence is recovered. The first observation we make, reading straight off from 2, is that the classical χ^2 -divergence can be seen as an inner product on the probability space weighted with the inversion of the distribution q_i . Due to the noncommutative nature of density matrices there is no unique generalization of this inversion. Consider for instance a generalization for two density matrices $\rho, \sigma \in S_d$, where for now we assume σ to be full rank, that is given by

$$\chi_\alpha^2(\rho, \sigma) = \text{tr}[(\rho - \sigma)\sigma^{-\alpha}(\rho - \sigma)\sigma^{\alpha-1}] = \text{tr}[\rho\sigma^{-\alpha}\rho\sigma^{\alpha-1}] - 1. \quad (7)$$

This gives rise to an entire family of χ^2 -divergences with (as we see below) special properties, for every $\alpha \in [0, 1]$. The natural question of whether there exists a classification of all possible inversions of σ , was investigated in a series of papers by Morozova and Chentsov³¹ Petz,^{14–16} in the context of information geometry. They considered the characterization of monotone Riemannian metrics on matrix spaces. Their general definition is based on the modular operator formalism of Araki,^{19,20} which we will also consider here. In order to classify the valid inversions, we first need to define the following set of functions, each of which gives rise to a possible inversion:

$$\mathcal{K} = \{k | -k \text{ is operator monotone, } k(w^{-1}) = wk(w), \text{ and } k(1) = 1\}. \quad (8)$$

Now, we define left and right multiplication operators as $L_Y(X) = YX$ and $R_Y(X) = XY$ respectively. The *modular operator* is defined as

$$\Delta_{\rho, \sigma} = L_\rho R_\sigma^{-1}, \quad (9)$$

for all $\rho, \sigma \in S_d$, $\sigma > 0$. Note that R_σ and L_ρ commute and inherit hermiticity and positivity from ρ, σ . The above should be read as follows: acting on some $A \in M_d$, $\Delta_{\rho, \sigma}(A) = \rho A \sigma^{-1}$. When manipulating the modular operators it is often convenient to write them in matrix form, in which case they read: $\hat{\Delta}_{\rho, \sigma}|A\rangle = \rho \otimes \bar{\sigma}^{-1}|A\rangle$, where $|A\rangle = A \otimes \mathbb{1}|I\rangle$, and $|I\rangle = \sum_{i=1}^d |ii\rangle$ corresponds to d times the maximally mixed state. This formalism gives rise to a more general quantum χ^2 -divergence.

Definition 1: For $\rho, \sigma \in S_d$, and $k \in \mathcal{K}$ we define the the quantum χ^2 -divergence

$$\chi_k^2(\rho, \sigma) = \langle \rho - \sigma, \Omega_\sigma^k(\rho - \sigma) \rangle, \quad (10)$$

when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and infinity otherwise. The inversion of σ is defined only when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and given by

$$\Omega_\sigma^k = R_\sigma^{-1}k(\Delta_{\sigma, \sigma}). \quad (11)$$

The functions $k_\alpha(w) = \frac{1}{2}(w^{-\alpha} + w^{\alpha-1})$, with $k_\alpha \in \mathcal{K}$, yield the family of χ_α^2 -divergences given in (7) which we call the mean α -divergences to distinguish them from the well-known family of Wigner-Yanase-Dyson (WYD) α -divergences, which are described in Appendix A, along with several other families. Although we focus on the family (7), most of our results hold for any divergence given by (10) with $k \in \mathcal{K}$ with the exceptions of Theorem 14.

A. Monotone Riemannian metrics and generalized relative entropies

This definition of the χ^2 -divergence stems from the analysis of monotone Riemannian metrics. By Riemannian metric, we mean a positive definite bilinear form $M_\sigma(A, B)$ on the hermitian tangent hyperplane $\mathcal{TP} = \{A \in \mathcal{M}_d : A = A^\dagger, \text{tr}[A] = 0\}$. The metric is monotone if for all quantum channels $T : \mathcal{M}_d \mapsto \mathcal{M}_d$, states $\sigma \in \mathcal{S}_d^+$ and $A \in \mathcal{TP}$, $M_{T(\sigma)}(T(A), T(A)) \leq M_\sigma(A, A)$. Petz showed that there is a one-to-one correspondence between the above metrics and a special class of convex operator functions, which correspond to $1/k$ in our notation. He could furthermore relate several generalized relative entropies (which he defined much earlier¹³ and referred to as quasi-entropies) to monotone Riemannian metrics.¹⁵⁻¹⁷ The reverse implication that every monotone Riemannian metric stems from a generalized relative entropy was first proved by Lesniewski and Ruskai.¹⁸ Taking advantage of the well-known integral representations of operator monotone and convex functions²¹ one can express the χ^2 -divergences as well as the relative entropies explicitly. We shall briefly repeat the key points of the analysis that are necessary for our understanding of the mixing-time and contraction analysis for cpt-maps.

We need to consider the class of functions \mathcal{G} by which we denote the set of continuous operator convex functions from \mathbb{R}^+ to \mathbb{R} that satisfy $g(1) = 0$. Note that these functions can all be classified in terms of the integral representation:

$$g(w) = a(w-1) + b(w-1)^2 + c \frac{(w-1)^2}{w} + \int_0^\infty \frac{(w-1)^2}{w+s} d\nu(s), \quad (12)$$

where $a, b, c > 0$ and the integral of the positive measure $d\nu(s)$ on $(0, \infty)$ is bounded. The generalized relative entropy for states $\rho, \sigma \in \mathcal{S}_d^+$ was first defined in Refs. 13 and 22.

Definition 2: Let $g \in \mathcal{G}$. The generalized quantum relative entropy is given by

$$H_g(\rho, \sigma) = \text{tr}[\rho^{1/2} g(\Delta_{\sigma, \rho})(\rho^{1/2})] \quad (13)$$

when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and infinity otherwise, and where $\Delta_{\rho, \sigma}$ is again the modular operator.

We now recall without proof a theorem^{15,16,18} relating the relative entropy and the monotone Riemannian metric, mirroring the classical result (4):

Theorem 3: For every $k \in \mathcal{K}$, there is a $g \in \mathcal{G}$ such that for a given $\sigma \in \mathcal{S}_d$, and A, B hermitian traceless, we get

$$\begin{aligned} M_\sigma^k(A, B) &= - \frac{\partial^2}{\partial \alpha \partial \beta} H_g(\sigma + \alpha A, \sigma + \beta B) \Big|_{\alpha=\beta=0} \\ &= \langle A, \Omega_\sigma^k(B) \rangle. \end{aligned} \quad (14)$$

and, k is related to g by

$$k(w) = \frac{g(w) + wg(w^{-1})}{(w-1)^2} \quad (15)$$

From this theorem follows a convenient integral representation of the inversion Ω_σ^k , which is equivalent to (11) (Ref. 18),

$$\Omega_\sigma^k = \int_0^\infty \left(\frac{1}{sR_\sigma + L_\sigma} + \frac{1}{R_\sigma + sL_\sigma} \right) N_g(s) ds, \quad (16)$$

where N_g denotes the singular measure $N_g(s)ds = (b_g + c_g)\delta(s)ds + d\nu_g(s)$. Note that the relationship between k and g is not one-to-one. Indeed, by setting $\hat{g}(w) = wg(w^{-1})$, we get back the

above relation. However, there is a one-to-one correspondence between each k and a symmetric $g_s(w) = g(w) + wg(w^{-1})$, and hence between each metric and a symmetric relative entropy.

Note that the α -subfamily of (7) has the associated symmetric relative entropy: $g^{\text{sym}}(x) = \frac{(1-w)^2}{2}(w^{\alpha-1} + w^{-\alpha})$, so that

$$H_\alpha^{\text{sym}}(\rho, \sigma) = \frac{1}{2}(H_\alpha(\rho, \sigma) + H_\alpha(\sigma, \rho)), \quad (17)$$

where

$$H_\alpha(\rho, \sigma) = \text{tr}[\rho^{2-\alpha}\sigma^{\alpha-1} + \rho^{1+\alpha}\sigma^{-\alpha} - 2\rho^\alpha\sigma^{1-\alpha}].$$

The integral representation (16) of the inversion Ω_σ^k allows for a partial ordering of different monotone Riemannian metrics that follows from the set of inequalities:

$$\frac{2}{x+1} \leq \frac{1+s}{2} \left(\frac{1}{s+x} + \frac{1}{sx+1} \right) \leq \frac{x+1}{2x}, \quad (18)$$

for $s \in [0, 1]$ and $x \in \mathbb{R}^+$. We therefore see that there exists a partial ordering for the inversions, with a lowest and highest element in the hierarchy. The lowest element gives rise to the so called Bures metric. Thus,

$$\Omega_\sigma^{\text{Bures}} = 2(R_\sigma + L_\sigma)^{-1} \leq \Omega_\sigma^k \leq (L_\sigma^{-1} + R_\sigma^{-1})/2 = \Omega_\sigma^{\alpha=0}. \quad (19)$$

The χ^2 -divergence is recovered from the metric upon setting $\chi_k^2(\rho, \sigma) \equiv M_\sigma^k(\rho - \sigma, \rho - \sigma)$. We are therefore left with a partial order for all possible χ^2 -divergences with a smallest and largest element according to

$$\chi_{\text{Bures}}^2(\rho, \sigma) \leq \chi_k^2(\rho, \sigma) \leq \chi_{\alpha=0}^2(\rho, \sigma). \quad (20)$$

The defining attribute of the above set of metrics is their monotonicity under the action of quantum channels. This was first shown by Petz,¹⁶ and later a proof based on the integral representation of Ω_σ^k (16), and on Schwarz-type inequalities, was provided by Ruskai and Lesniewski.¹⁸ Due to its importance for the mixing time analysis we shall repeat it here.

Theorem 4: For all $\sigma \in \mathcal{S}_d$, M_σ^k is monotone under the action of a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ for all $k \in \mathcal{K}$ and $A \in \mathcal{M}_d$, i.e.,

$$M_\sigma^k(A, A) \geq M_{T(\sigma)}^k(T(A), T(A)). \quad (21)$$

Proof: The monotonicity follows immediately from the integral representation of the inversion Ω_σ^k in (16), and an argument proved in the appendix, Theorem 24. The theorem states that for every channel T and for arbitrary A , we have

$$\text{tr} \left[A^\dagger \frac{1}{R_\sigma + sL_\sigma} A \right] = \text{tr} \left[T \left(A^\dagger \frac{1}{R_\sigma + sL_\sigma} A \right) \right] \geq \text{tr} \left[T(A)^\dagger \frac{1}{R_{T(\sigma)} + sL_{T(\sigma)}} T(A) \right]. \quad (22)$$

□

B. Properties of the quantum χ^2 -divergence

The fact that the quantum χ_k^2 -divergence can be used to bound the mixing time lies in the following lemma, that upper bounds the trace distance which is the relevant distance measure in the mixing time definition.

Lemma 5: For every pair of density operators $\rho, \sigma \in \mathcal{S}_d$, we have that

$$\|\rho - \sigma\|_1^2 \leq \chi_k^2(\rho, \sigma). \quad (23)$$

Proof: If the support of ρ is not contained in the support of σ , then the right-hand side is ∞ . We can therefore assume w.l.o.g. that $\sigma > 0$ by restricting the analysis to the support space of σ . The

trace norm $\|A\|_1$ of some matrix $A \in \mathcal{M}_d$ can be expressed as²³ $\|A\|_1 = \max_{U \in U(d)} |\text{tr}[UA]|$, where the maximum is taken over all unitaries acting on the d -dimensional Hilbert space. Thus, for any inversion Ω_σ^k :

$$\begin{aligned} \|A\|_1^2 &= \max_{U \in U(d)} |\text{tr}[UA]|^2 = \max_{U \in U(d)} \left| \text{tr} \left[U [\Omega_\sigma^k]^{-1/2} \circ [\Omega_\sigma^k]^{1/2} (A) \right] \right|^2 \\ &= \max_{U \in U(d)} \left| \text{tr} \left[[\Omega_\sigma^k]^{-1/2} (U) [\Omega_\sigma^k]^{1/2} (A) \right] \right|^2 \\ &\leq \text{tr} \left[A^\dagger \Omega_\sigma^k (A) \right] \max_{U \in U(d)} \text{tr} \left[U^\dagger [\Omega_\sigma^k]^{-1} (U) \right]. \end{aligned} \tag{24}$$

Let us consider the Bures inversion given by $\Omega_\sigma^{\text{Bures}} = 2[L_\sigma + R_\sigma]^{-1}$. Clearly, its inverse is $[\Omega_\sigma^{\text{Bures}}]^{-1} = \frac{1}{2} [L_\sigma + R_\sigma]$. Therefore, for any unitary U ,

$$\text{tr} \left[U^\dagger [\Omega_\sigma^{\text{Bures}}]^{-1} (U) \right] = \frac{1}{2} (\text{tr} [U^\dagger \sigma U] + \text{tr} [U^\dagger U \sigma]) = 1. \tag{25}$$

Setting $A = \rho - \sigma$ and observing that $\chi_{\text{Bures}}^2 \leq \chi_k^2$ for all $k \in \mathcal{K}$ completes the proof. \square

We have already stated that the family of χ_α^2 -divergences defined in 7 can be cast into the general framework of monotone Riemannian metrics. Because of its computational simplicity, and its special symmetry when $\alpha = 1/2$, we consider its properties more specifically. It is possible for instance to show monotonicity of this subfamily using arguments from joint convexity. As the proof is interesting in its own right, we give it here.

Proposition 6: χ_α^2 is jointly convex in its arguments for $\alpha \in [0, 1]$. Moreover, it is monotone w.r.t. completely positive trace-preserving maps, i.e.,

$$\chi_\alpha^2(\rho, \sigma) \geq \chi_\alpha^2(T(\rho), T(\sigma)), \tag{26}$$

for every quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$.

Proof: A direct application of Corollary 2.1 in Ref. 28 guarantees that $\chi_\alpha^2(\rho, \sigma)$ is jointly convex in its arguments for any $\alpha \in [0, 1]$. This in turn implies monotonicity w.r.t. cp-maps by a standard argument: let us represent T as $T(\rho) = \text{tr}_E[U(\rho \otimes \psi)U^\dagger]$ where ψ is a pure state (i.e., rank-one projection), U a unitary and tr_E the partial trace over an “environmental” system of dimension m . If we take a unitary operator basis $\{V_i\}_{i=1, \dots, m^2}$ in \mathcal{M}_m (orthonormal w.r.t. the Hilbert–Schmidt inner product), we can write

$$T(\rho) \otimes \mathbb{1}_m/m = \frac{1}{m^2} \sum_{i=1}^{m^2} (\mathbb{1} \otimes V_i) U(\rho \otimes \psi) U^\dagger (\mathbb{1} \otimes V_i^\dagger). \tag{27}$$

However, since $\chi_\alpha^2(T(\rho), T(\sigma)) = \chi_\alpha^2(T(\rho) \otimes \tau, T(\sigma) \otimes \tau)$, in particular for $\tau = \mathbb{1}_m/m$, we can now apply joint convexity. With the help of the fact that for any unitary W it holds that $(W \cdot W^\dagger)^\alpha = W(\cdot)^\alpha W^\dagger$ we obtain the claimed result. \square

Furthermore, we note that this subfamily also has a natural ordering.

Proposition 7: For every $\rho, \sigma \in \mathcal{S}_d$, χ_α^2 is convex in α , and reaches a minimum for $\alpha = 1/2$.

Proof: First note that $\chi_{\alpha=0}^2(\rho, \sigma) = \chi_{\alpha=1}^2(\rho, \sigma)$. That the minimum is reached for $\alpha = 1/2$ follows directly from the Cauchy–Schwarz inequality. Applied to our problem, we get

$$\begin{aligned} \text{tr}[\rho \sigma^{-1/2} \rho \sigma^{-1/2}]^2 &= \text{tr}[\rho \sigma^{(\alpha-1)/2} \sigma^{-\alpha/2} \rho \sigma^{(\alpha-1)/2} \sigma^{-\alpha/2}]^2 \\ &\leq \text{tr}[\rho \sigma^{-\alpha} \rho \sigma^{\alpha-1}]^2. \end{aligned} \tag{28}$$

To see convexity, consider the second partial derivative of χ_α^2 with respect to α :

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} \chi_\alpha^2(\rho, \sigma) &= \text{tr} \sigma^{\alpha-1} \rho \sigma^{-\alpha} (\rho \log^2 \sigma + \log^2 \sigma \rho - 2 \log \sigma \rho \log \sigma) \\ &= \sum_{kl} \mu_k^{\alpha-1} \mu_l^{-\alpha} (\log \mu_k - \log \mu_l)^2 |\langle k | \rho | l \rangle|^2 \geq 0, \end{aligned} \quad (29)$$

where we used $\sigma = \sum_k \mu_k |k\rangle \langle k|$. \square

Finally, we point out a bound on the relative entropy in terms of the α -subfamily of χ^2 -divergences:

Theorem 8: For every pair of density operators ρ and σ and every $\alpha \in (0, 1]$, we have that

$$\chi_\alpha^2(\rho, \sigma) \geq S(\rho, \sigma), \quad (30)$$

where $S(\rho, \sigma) = \text{tr} \rho (\log \rho - \log \sigma)$ is the usual relative entropy.

Proof: It was shown in Ref. 27 that for $\gamma \in (0, 1]$, the following holds:

$$S(\rho, \sigma) \leq \frac{1}{\gamma} (\text{tr} \rho^{1+\gamma} \sigma^{-\gamma} - 1). \quad (31)$$

Then consider,

$$\begin{aligned} \chi_\alpha^2(\rho, \sigma) - S(\rho, \sigma) &\geq \text{tr} \rho \sigma^{-1/2} \rho \sigma^{-1/2} - 2 \text{tr} \rho^{3/2} \sigma^{-1/2} + 1 \\ &= \text{tr} (\rho^{1/2} \sigma^{-1/2} \rho^{1/2} - \rho^{1/2})^2 \geq 0, \end{aligned} \quad (32)$$

where the first inequality comes from taking $\gamma = 1/2$ in (II B), and $\alpha = 1/2$ for χ_α^2 , and the last line is obtained from rearranging terms. \square

III. MIXING TIME BOUNDS AND CONTRACTION OF THE χ^2 -DIVERGENCE UNDER CPT MAPS

A. Mixing time Bounds

The χ^2 -divergence is an essential tool in the study of Markov chain mixing times, because on the one hand it bounds the trace distance, and on the other it allows easy access to the spectral properties of the map. The subsequent analysis can be seen as a generalization of the work presented in Refs. 1 and 2 to the noncommutative setting.

Theorem 9: Let $T : M_d \mapsto M_d$ be an ergodic quantum channel with fixed point $\sigma \in S_d$, for any $\rho \in S_d$ and any $k \in \mathcal{K}$, we can bound

$$\|T^n(\rho) - \sigma\|_1 \leq (s_1^k)^n \sqrt{\chi_k^2(\rho, \sigma)}. \quad (33)$$

Here s_1^k denotes the second largest singular value (the largest being 1) of the map

$$Q_k = [\Omega_\sigma^k]^{1/2} \circ T \circ [\Omega_\sigma^k]^{-1/2}. \quad (34)$$

Before we prove Theorem 9, we would like to point out an important fact that regards the singular values of Q_k . The monotonicity of the χ^2 -divergence ensures that the singular values s_i^k of Q_k are always contained in $[0, 1]$ irrespective of the choice of $k \in \mathcal{K}$. Let us therefore prove the following.

Lemma 10: The spectrum of the map $S_k \equiv Q_k^* \circ Q_k = [\Omega_\sigma^k]^{-1/2} \circ T^* \circ \Omega_\sigma^k \circ T \circ [\Omega_\sigma^k]^{-1/2}$ is contained in $[0, 1]$.

Proof: Let us first note that the map S_k is hermitian and positive by construction. Furthermore, the monotonicity of the χ^2 -divergence, as stated in Theorem 4, ensures that the Rayleigh–Ritz

quotient is bounded by 1. This holds, since $\forall B$

$$\begin{aligned} \langle B, S_k(B) \rangle &= \langle A, T^* \circ \Omega_\sigma^k \circ T(A) \rangle = M_{T(\sigma)}^k(T(A), T(A)) \\ &\leq M_\sigma^k(A, A) = \langle A, \Omega_\sigma^k(A) \rangle = \langle B, B \rangle, \end{aligned} \tag{35}$$

where we defined the intermediate state $A = [\Omega_\sigma^k]^{-1/2}(B)$. Note that we made use of the fact that $\sigma = T(\sigma)$ is the fixed point of the map. Therefore

$$\lambda_{max} = \max_{B \in \mathcal{M}_d} \frac{\langle B, S_k(B) \rangle}{\langle B, B \rangle} \leq 1 \tag{36}$$

and the maximum is attained for $\lambda_{max} = 1$ and $B_{max} = [\Omega_\sigma^k]^{1/2}(\sigma)$. □

With the bound on the spectrum at hand, it is now straightforward to prove Theorem 9.

Proof: Define $e(n) \in \mathcal{M}_d$, as $e(n) = T^n(\rho - \sigma)$. By Lemma 5, we get $\|e(n)\|_1^2 \leq \chi_k^2(T^n(\rho), T^n(\sigma)) \equiv \chi_k^2(n)$. In the matrix representation, $|e(n)\rangle = e(n) \otimes \mathbb{1}|I\rangle$, we can rewrite $\chi_k^2(n) = \langle e(n)| \hat{\Omega}_\sigma^k |e(n)\rangle$. Also note that $|e(n+1)\rangle = \hat{T}|e(n)\rangle$ and so,

$$\chi_k^2(n) - \chi_k^2(n+1) = \langle e(n)| \hat{\Omega}_\sigma^k |e(n)\rangle - \langle e(n)| \hat{T}^\dagger \hat{\Omega}_\sigma^k \hat{T} |e(n)\rangle \tag{37}$$

$$= \langle e(n)| [\hat{\Omega}_\sigma^k]^{1/2} (\mathbb{1} - \hat{Q}_k^\dagger \hat{Q}_k) [\hat{\Omega}_\sigma^k]^{1/2} |e(n)\rangle. \tag{38}$$

Due to Lemma 10 we know that the spectrum of $\hat{S}_k = \hat{Q}_k^\dagger \hat{Q}_k$, which is equal to the square of the singular values of \hat{Q}_k , is contained in the interval $[0, 1]$. Hence,

$$\langle e(n)| [\hat{\Omega}_\sigma^k]^{1/2} (\mathbb{1} - \hat{S}_k) [\hat{\Omega}_\sigma^k]^{1/2} |e(n)\rangle \tag{39}$$

$$\geq (1 - s_1^2) \langle e(n)| [\hat{\Omega}_\sigma^k]^{1/2} \sum_{\alpha \neq 0} P_\alpha [\hat{\Omega}_\sigma^k]^{1/2} |e(n)\rangle. \tag{40}$$

The sum is taken over spectral projectors P_α^k of $\hat{S}_k = \sum_\alpha (s_\alpha^k)^2 P_\alpha$, apart from P_0^k which projects onto $[\hat{\Omega}_\sigma^k]^{-1/2}|\sigma\rangle$. In particular, $P_0^k = [\hat{\Omega}_\sigma^k]^{-1/2}|\sigma\rangle\langle I|[\hat{\Omega}_\sigma^k]^{-1/2}$, so that $\langle e(n)|[\hat{\Omega}_\sigma^k]^{1/2} P_0^k [\hat{\Omega}_\sigma^k]^{1/2} |e(n)\rangle = \langle e(n)|\sigma\rangle \text{tr}[T^n(\rho - \sigma)] = 0$, by trace preservation of T . We can write

$$\chi_k^2(n) - \chi_k^2(n+1) \geq (1 - (s_1^k)^2) \chi_k^2(n). \tag{41}$$

Rearranging terms completes the theorem. □

Remark: The fact that the singular values of Q_k are always smaller or equal to 1 justifies the use of the generalized χ^2 -divergence as the appropriate distance measure to bound the convergence of an arbitrary channel. It is tempting to use the Hilbert–Schmidt inner product to give an upper bound to the trace norm. This can always be done at the cost of a dimension dependent prefactor, since on finite dimensional spaces all norms are equivalent. However, when doing so a problem arises if one tries to bound the convergence in terms of the spectral properties of the map $S_{HS} = T^* \circ T$. It is in general not ensured that the spectrum will be bounded by 1. In fact, for every nonunital channel T , S_{HS} will have an eigenvalue greater than 1 (Ref. 32). The similarity transformation of the channel T with $[\Omega_\sigma^k]^{1/2}$ alters the singular values, but of course leaves the spectrum invariant. Furthermore, it is a well-known fact²³ that the singular values of a square matrix log-majorize the absolute value of the eigenvalues. As the spectrum of Q_k is bounded by 1 (and equal that of \hat{T} by similarity), we conclude that its second largest eigenvalue is always smaller or equal to its second largest singular value. One can also give a general bound in terms of the second largest eigenvalue of T (Ref. 33), but one is then confronted with a potentially severe dimensional prefactor.

For some instances of the inversion Ω_σ^k it becomes immediately evident that the symmetrization S_k has the desired spectral properties without making use of the monotonicity of the χ_k^2 -divergence. It can occur that S_k is again similar to a quantum channel that is of the form $T_s^k = [\Omega_\sigma^k]^{-1/2} \circ S_k \circ [\Omega_\sigma^k]^{1/2}$. A possible example of such an inversions is $\Omega_\sigma^{\alpha=1/2} = L_\sigma^{-1/2} R_\sigma^{-1/2}$. This is however not the generic case, most inversions will lead to maps that are not completely positive any longer. It would be very desirable to find other such examples, as they mirror the classical situation where the

symmetrized maps are always probability transition matrices, and because these specific inversions allow for clean contraction bounds as seen in Sec. III.B.

It is clear from the discussion above that the singular values of Q_k play a crucial role in the mixing time analysis presented here. This seems to contradict the general understanding that the convergence is determined by the spectral properties of the channel T in the asymptotic limit. This can however be understood as follows: the matrix \hat{Q}_k is similar to \hat{T} , i.e., $\hat{Q}_k = [\hat{\Omega}_\sigma^k]^{1/2} \cdot \hat{T} \cdot [\hat{\Omega}_\sigma^k]^{-1/2}$, so the spectra of Q_k and T coincide. The following lemma establishes a relation between the singular values and the eigenvalues in the asymptotic limit (for a proof, see e.g., Ref. 24, p. 180).

Lemma 11: *Let $\hat{Q}_k \in \mathcal{M}_{d^2}$ be given, and let $s_0(\hat{Q}_k) \geq \dots \geq s_{d^2-1}(\hat{Q}_k)$ and $\{\lambda_i(\hat{Q}_k)\}_{i=0\dots d^2-1}$ denote its singular values and eigenvalues, respectively with $|\lambda_0(\hat{Q}_k)| \geq \dots \geq |\lambda_{d^2-1}(\hat{Q}_k)|$. Then*

$$\lim_{n \rightarrow \infty} [s_i(\hat{Q}_k^n)]^{1/n} = |\lambda_i(\hat{Q}_k)| \quad \forall i = 0 \dots d^2 - 1. \tag{42}$$

In the limit of $n \rightarrow \infty$ applications of the quantum channel, we can start blocking the channel in m subsequent applications $T^{(m)} \equiv T^m$ and bound the convergence rate as a function of the singular values of the corresponding $\hat{Q}_k^{(m)}$, which indeed converge to the eigenvalues of the original cp-map. Convergence following the eigenvalue is therefore only guaranteed in the limit of $n \rightarrow \infty$, and this would indeed be the case, when for example the eigenstructure of the original cp-maps contains a Jordan block associated with the second largest eigenvalue. Note that convergence in the above lemma goes typically as $1/n$, which is very slow. Hence for finite n , convergence is governed by the singular values of \hat{Q}_k as opposed to the eigenvalues. The bound derived in (9) is an absolute bound for finite n and clearly leads to a strictly monotonic decay. Note that in the case that the second largest singular value is also equal to 1, this can then always be cured by blocking the cp-maps together. Finally, it is worth mentioning that the convergence can in fact be much more rapid if one starts in a state “closer” to the fixed point. In particular, if the initial state is such that $\rho - \sigma \propto Y_k$, $k \geq 2$, where Y_k is the eigenvector corresponding to λ_k , then the convergence will be governed by the magnitude of λ_k . Furthermore, if instead of a single fixed point, we have a fixed subspace, or a collection of fixed subspaces (with or without rotating points), then the convergence to this fixed subspace will be governed by the largest eigenvalue whose magnitude is strictly smaller than 1.

Thus far we have only considered the time-discrete case, it is however straightforward to give a similar bound for time-continuous Markov processes that are described by a one parameter semigroup. The following lemma bounds the trace-distance as a function of $t \in \mathbb{R}_0^+$: The proof of the following lemma is very similar to the proof of the time discrete case, we will therefore omit it here.

Lemma 12: *Let \mathcal{L} denote the generator of a time continuous Markov process, described by the master equation $\partial_t \rho = \mathcal{L}(\rho)$, with solution $\rho(t) \in \mathcal{S}_d \forall t \in [0, +\infty)$. Furthermore, let $\sigma \in \mathcal{S}_d^+$ denote the fixed point $\mathcal{L}(\sigma) = 0$, then*

$$\|\rho(t) - \sigma\|_1^2 \leq e^{l_1^k t} \chi_k^2(\rho(0), \sigma). \tag{43}$$

Here, $l_1^k \leq 0$ refers to the second largest eigenvalue of

$$\Lambda_k = [\Omega_\sigma^k]^{1/2} \circ \mathcal{L}^* \circ [\Omega_\sigma^k]^{-1/2} + [\Omega_\sigma^k]^{-1/2} \circ \mathcal{L} \circ [\Omega_\sigma^k]^{1/2}. \tag{44}$$

The symmetrization for the generator of the time continuous Markov process is additive as would be expected. Furthermore, we note that the monotonicity of the χ^2 -divergence ensures that the spectrum of Λ_k is never positive, based on a similar reasoning as given in lemma (10).

B. Contraction Coefficients

In the following we study the contraction of the χ^2 -divergences under quantum channels, and its relation to the trace norm contraction. We consider general contraction rather than contraction to the fixed point because analytic results are more readily available, and because these bounds

are in a sense the *most* stringent one can require. We focus primarily on the mean α -subfamily of χ^2 -divergences.

Let us define the following contraction coefficients which we call the χ^2 - and trace norm-contraction respectively:

$$\eta_\chi^\alpha(T) = \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{\chi_\alpha^2(T(\rho), T(\sigma))}{\chi_\alpha^2(\rho, \sigma)} \tag{45}$$

and

$$\eta_{\text{tr}}(T) = \sup_{\rho, \sigma \in \mathcal{S}_d} \frac{\|T(\rho - \sigma)\|_1}{\|\rho - \sigma\|_1} = \sup_{\phi, \psi \in \mathcal{S}_d^+, \langle \phi | \psi \rangle = 0} \frac{1}{2} \|T(\psi) - T(\phi)\|_1, \tag{46}$$

where $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ is a quantum channel, and the last equality is seen simply by convexity of the trace norm.

We first upper bound the trace-norm contraction in terms of the χ^2 contraction, which is a generalization of a result in Ref. 26.

Theorem 13: For all $\alpha \in (0, 1]$, and a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,

$$\eta_{\text{tr}}(T) \leq \sqrt{\eta_\chi^\alpha(T)}. \tag{47}$$

Proof: From Lemma 5, we have that $\|T(\rho - \sigma)\|_1^2 \leq \chi_\alpha^2(T(\rho), T(\sigma))$, for all $\rho, \sigma \in \mathcal{S}_d$. Let N be traceless and hermitian, and note that it can be written as $N = N_+ - N_-$, where N_+, N_- are positive definite and orthogonal in their support. Now let $P = |N|/||N||_1$ and recall that $|N| = N_+ + N_-$, then we get $\text{tr}[NP^{-\alpha}NP^{\alpha-1}] = ||N||_1^2$, for every $\alpha \in (0, 1]$. Also,

$$\frac{\|T(N)\|_1^2}{||N||_1^2} \leq \frac{\text{tr}[T(N)T(P)^{-\alpha}T(N)T(P)^{\alpha-1}]}{\text{tr}[NP^{-\alpha}NP^{\alpha-1}]}, \tag{48}$$

where the inequality is in the numerator, and the denominators are equal, by the previous observation. Taking the supremum over all traceless hermitian N on the left hand side and identifying $\rho - \sigma = N/||N||_1, P = \sigma$ then gives the desired result. \square

We now provide a lower bound to the trace norm contraction for primitive channels.

Theorem 14: Given a quantum channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$,

$$\eta_\chi^{\alpha=1/2}(T) \leq \eta_{\text{tr}}(T). \tag{49}$$

First we introduce an eigenvalue type min-max characterization of the χ^2 -contraction, and then show that this eigenvalue must be smaller than the trace norm-contraction.

Let $P > 0$, and consider the following eigenvalue equation:

$$\hat{\Gamma}|A\rangle \equiv \hat{\Omega}_P^{-1} \hat{T}^\dagger \hat{\Omega}_{T(P)} \hat{T}|A\rangle = \lambda|A\rangle, \tag{50}$$

where $\Omega_\chi \equiv \Omega_\chi^{\alpha=1/2}$. If T has a nontrivial kernel, then $\Omega_{T(P)}$ should be understood in terms of the pseudoinverse. First note that Γ is a quantum channel, so its spectrum is bounded by 1, and that it reaches 1 for $A = P$. Also note that Γ is similar to a hermitian operator, so it has all real eigenvalues, so we can take the eigenvectors to be hermitian. Then rewriting (50) as $\hat{T}^\dagger \hat{\Omega}_{T(P)} \hat{T}|A\rangle = \lambda \hat{\Omega}_P|A\rangle$, we can express the second largest eigenvalue as

$$\begin{aligned} \lambda_1(T, P) &= \sup_{\langle N | \Omega_P(P) \rangle = 0, N = N^\dagger} \frac{\langle N | \hat{T}^\dagger \hat{\Omega}_{T(P)} \hat{T} | N \rangle}{\langle N | \hat{\Omega}_P | N \rangle} \\ &= \sup_{\text{tr} N = 0, N = N^\dagger} \frac{\text{tr}[T(N)T(P)^{-1/2}T(N)T(P)^{-1/2}]}{\text{tr}[NP^{-1/2}NP^{-1/2}]}. \end{aligned} \tag{51}$$

Clearly, by maximizing over all P , one recovers $\eta_\chi^{1/2}(T)$. We now prove the above theorem.

Proof: Let N_1 be the eigenvector for which λ_1 satisfies the eigenvalue equation (50), and recall that N_1 is Hermitian and traceless. Then,

$$\lambda_1 \|N_1\|_1 = \|\Gamma(T(N_1))\|_1 \leq \|T(N_1)\|_1 \tag{52}$$

because Γ is a channel, and

$$\lambda_1 \leq \frac{\|T(N_1)\|_1}{\|N_1\|_1} \leq \sup_{\text{tr} N=0, N^\dagger=N} \frac{\|T(N)\|_1}{\|N\|_1} = \eta_{tr}, \tag{53}$$

taking the supremum over positive P completes the proof. □

Remark: Theorem 14 gives a computable lower bound to the trace norm contraction. A key subtlety in the argument is that $[\Omega_P(A)]^{-1} = \sqrt{P}A\sqrt{P}$ is a completely positive (CP) map (with a single Kraus operator \sqrt{P}), which implies that Γ is a quantum channel. In general, Ω_P is not even positivity preserving. Another exception is the monotone metric associated with the usual logarithmic relative entropy for which $k(w) = \frac{\log w}{w-1}$. It is well known^{13,18,38} that $\Omega_P^{\log}(A)$ can be written as

$$\Omega_P^{\log}(A) = \int_0^\infty \frac{1}{P+xI} A \frac{1}{P+xI} dx, \tag{54}$$

which is clearly CP. An analogous lower bound was shown in Ref. 18 for this map using a similar argument. Clearly, this can be extended to any monotone metric for which Ω_P is CP; however, we do not know of any other examples.

Very little is known about the ordering of the general η_k contraction coefficients. In particular, we do not know whether η_χ^{\log} is smaller or larger than $\eta^{\alpha=1/2}_\chi$. However, it is known¹⁸ that η_k are not all identical for different $k \in \mathcal{K}$; because examples can be constructed using nonunital qubit channels. Theorem 13 can readily be extended to any metric associated with $k \in \mathcal{K}$. However, it seems unlikely that Theorem 14 holds in general. Thus, we can conclude

$$\max\{\eta_\chi^{\alpha=1/2}(T), \eta_\chi^{\log}(T)\} \leq \eta_{tr}(T) \leq \inf_{k \in \mathcal{K}} \sqrt{\eta_\chi^k(T)}. \tag{55}$$

Note that if instead of maximizing over all P , we only consider contraction of the map to the steady state and denote it $\bar{\eta}(T) = \eta(T)_{P=\sigma}$, then from the above arguments one immediately gets

$$\bar{\eta}_\chi^\alpha(T) \leq \bar{\eta}_{tr}(T) \leq \eta_{tr}(T) \leq 1. \tag{56}$$

Combing this with the previous bounds above, we have

$$\lambda_1 \leq s_1^{\alpha=1/2} = \bar{\eta}_\chi^{\alpha=1/2} \leq \eta_\chi^{\alpha=1/2} \leq \eta_{tr} \leq \sqrt{\eta_\chi^{\alpha=1/2}}. \tag{57}$$

Moreover, $k(w) = \sqrt{w}$ on the right can be replaced by any $k \in \mathcal{K}$, and that on the left by $k(w) = (w-1)^{-1} \log w$. It is very tempting to conjecture that $\bar{\eta}_{tr}^2 \leq \bar{\eta}_\chi^\alpha$, and/or that $\eta_{tr} \leq \sqrt{\bar{\eta}_\chi^{\alpha=1/2}}$, but simple numerical counterexamples show these to be false.

IV. QUANTUM DETAILED BALANCE

The detailed balance condition is often crucial in the analysis of classical Markov chain mixing times, as it ensures several convenient properties of the Markov chain. In particular, it implies that the classical probability distribution with respect to which the stochastic map is detailed balanced is a fixed point of the chain. Furthermore, detailed balanced stochastic maps have a real spectrum. In this section we generalize the notion of classical detailed balance to quantum Markov chains. Alternative definitions of quantum detailed balance have been given in the literature^{29,34,36,37} and references therein. Central to our approach is the operator Q_k as previously introduced in Lemma 9.

In the literature, for classical Markov chains, an analogous matrix exists and is often referred to as the discriminant.

Definition 15: For a channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and a state $\sigma \in \mathcal{S}_d^+$ with corresponding inversion Ω_σ^k as defined in (11), we define the quantum discriminant of T as

$$Q_k = [\Omega_\sigma^k]^{1/2} \circ T \circ [\Omega_\sigma^k]^{-1/2}. \tag{58}$$

We recall that the convergence of an arbitrary quantum Markov process can be bounded by the singular values of \hat{Q}_k . Classical detailed balanced Markov chains have the property that the corresponding discriminant becomes symmetric. We shall therefore define the quantum generalization for a quantum detailed balanced process

$$Q_k^* = Q_k. \tag{59}$$

This immediately allows to make a statement about the spectrum of quantum detailed balanced maps. Due to the hermicity of the matrix representation of the map (58), we can immediately deduce, just as for classically case, that the quantum channel T has a real spectrum. For detailed balanced maps, the second largest eigenvalue in magnitude coincides with the second largest singular value. Furthermore, we would like to point out that this is actually not just a single condition for quantum detailed balance but a whole family. Hence every different inversion Ω_σ^k gives rise to a different condition for detailed balance. We therefore define as the quantum generalization of detailed balance.

Definition 16: For a channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ and a state $\sigma \in \mathcal{S}_d^+$, we say that T obeys k -detailed balance with respect to σ with $k \in \mathcal{K}$, when

$$[\Omega_\sigma^k]^{-1} \circ T^* = T \circ [\Omega_\sigma^k]^{-1}. \tag{60}$$

A consequence of this definition is that σ is a fixed point of T .

Lemma 17: Let $\sigma \in \mathcal{S}_d$ be a state and T be a channel that satisfies the detailed balance definition 16 with respect to Ω_σ^k , then σ is a steady state of T .

Proof: Recall that the inverse is given by $[\Omega_\sigma^k]^{-1} = R_\sigma f(\Delta_{\sigma,\sigma})$, where $f(w) = 1/k(w)$. Hence, since $k(1) = f(1) = 1$, we have

$$[\Omega_\sigma^k]^{-1}(\mathbb{1}) = R_\sigma f(\Delta_{\sigma,\sigma})\mathbb{1} = R_\sigma \mathbb{1} = \sigma. \tag{61}$$

Now, since $T^*(\mathbb{1}) = \mathbb{1}$, we have that

$$T(\sigma) = T \circ [\Omega_\sigma^k]^{-1}(\mathbb{1}) = [\Omega_\sigma^k]^{-1} \circ T^*(\mathbb{1}) = [\Omega_\sigma^k]^{-1}(\mathbb{1}) = \sigma. \tag{62}$$

□

Given a probability distribution on some set of states, it is desirable to have a simple criterium to check whether a completely positive map obeys detailed balance with respect to the state generated from the distribution. This criterium may then serve to set up a Markov chain that actually converges to the desired steady state.

Proposition 18: Let $\{|i\rangle\}_i$ be a complete orthonormal basis of \mathfrak{H} and let $\{\mu_i\}_i$ be a probability distribution on this basis. Furthermore, assume that a quantum channel T obeys

$$\frac{\mu_n}{k(\mu_m/\mu_n)} \langle i | T(|n\rangle\langle m|) | j \rangle = \frac{\mu_i}{k(\mu_j/\mu_i)} \langle m | T(|j\rangle\langle i|) | n \rangle, \tag{63}$$

then $\sigma = \sum_i \mu_i |i\rangle\langle i|$ and T obey the detailed balance condition with respect to Ω_σ^k .

Proof: Note that $\{|i\rangle\langle j|\}_{ij}$ forms a complete and orthonormal basis in the space \mathcal{M}_d with respect to the Hilbert–Schmidt scalar product. We can therefore express (60) on this basis. The individual

entries are equal due to

$$\begin{aligned} \text{tr}[(|m\rangle\langle n|)^\dagger [\Omega_\sigma^k]^{-1} \circ T^*(|j\rangle\langle i|)] &= \mu_n k^{-1} (\mu_m/\mu_n) \text{tr}[T(|m\rangle\langle n|)^\dagger (|j\rangle\langle i|)] = \\ \mu_n k^{-1} (\mu_m/\mu_n) \langle i| T(|n\rangle\langle m|) |j\rangle &= \mu_i k^{-1} (\mu_j/\mu_i) \langle m| T(|j\rangle\langle i|) |n\rangle = \\ \mu_i k^{-1} (\mu_j/\mu_i) \text{tr}[(|m\rangle\langle n|)^\dagger T(|j\rangle\langle i|)] &= \text{tr}[(|m\rangle\langle n|)^\dagger T \circ [\Omega_\sigma^k]^{-1} (|j\rangle\langle i|)]. \end{aligned} \tag{64}$$

□

Remark: We note that the different quantum detailed balance conditions coincide for classical channels, i.e., for stochastic processes that are included in the framework of quantum channels. Define the following “classical” Kraus operators:

$$A_{ij}^{cl} = \sqrt{P_{ij}} |i\rangle\langle j| \text{ and a state, } \sigma = \sum_i \mu_i |i\rangle\langle i|. \tag{65}$$

In this case, the condition of Proposition 18 reduces to the classical condition. This can be seen when considering the channel $T^{cl}(\rho) = \sum_{ij} A_{ij}^{cl} \rho A_{ij}^{cl\dagger}$ and checking for detailed balance with respect to sigma, since

$$\frac{\mu_m}{k(\mu_n/\mu_m)} \langle i| T^{cl}(|n\rangle\langle m|) |j\rangle = \frac{\mu_m}{k(\mu_n/\mu_m)} \delta_{nm} \delta_{ij} P_{in}$$

and

$$\frac{\mu_i}{k(\mu_j/\mu_i)} \langle i| T^{cl}(|n\rangle\langle m|) |j\rangle = \frac{\mu_i}{k(\mu_j/\mu_i)} \delta_{nm} \delta_{ij} P_{ni}. \tag{66}$$

However since $k(1) = 1$ we are just left with the classical detailed balance condition $\mu_i P_{ni} = \mu_n P_{in}$ for all pairs i, n .

A natural question to ask is, therefore, whether the different detailed balance condition are all identical. To see that this is not the case, consider the example given by the Kraus operators of a single qubit, i.e., $\mathfrak{H} = \mathbb{C}^2$,

$$A_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } A_2 = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}. \tag{67}$$

This channel has the unique fixed point

$$\sigma = \frac{1}{6} \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}. \tag{68}$$

From this channel it is now possible to construct a channel that obeys detailed balance with respect to the inversion given by choosing $k(w) = w^{-1/2}$, that is the inversion reads $\Omega_\sigma^{\alpha=1/2} = L_\sigma^{-1/2} R_\sigma^{-1/2}$. We consider therefore the symmetrized map,

$$T_s = [\Omega_\sigma^{\alpha=1/2}]^{-1} \circ T^* \circ \Omega_\sigma^{\alpha=1/2} \circ T. \tag{69}$$

For the specific instance where $\Omega_\sigma^{\alpha=1/2}$ is given as above, we are assured that the map T_s is again a quantum channel, because one immediately finds the Kraus representation for $T_s(\rho) = \sum_{ij} B_{ij} \rho B_{ij}^\dagger$ as $B_{ij} = \sqrt{\sigma} A_i^\dagger [\sqrt{\sigma}]^{-1} A_j$. The individual Kraus operators read

$$\begin{aligned} B_{11} &= \frac{3}{5} \begin{pmatrix} 1 & 1 \\ 1/2 & 1/2 \end{pmatrix} \text{ and } B_{12} = \frac{\sqrt{2}}{5} \begin{pmatrix} 1 & -1 \\ 1/2 & -1/2 \end{pmatrix}, \\ B_{21} &= \frac{\sqrt{2}}{20} \begin{pmatrix} 3 & 3 \\ -1 & -1 \end{pmatrix} \text{ and } B_{22} = \frac{1}{5} \begin{pmatrix} 3 & -3 \\ -1 & 1 \end{pmatrix}. \end{aligned} \tag{70}$$

The channel T_s satisfies detailed balance with respect to $\Omega_\sigma^{\alpha=1/2}$ by construction. This channel however does not satisfy detailed balance with respect to the inversion $\Omega_\sigma^{Bures} = 2[L_\sigma + R_\sigma]^{-1}$ as can be seen directly by evaluating the detailed balance condition in terms of the matrix representations,

$$[\hat{\Omega}_\sigma^{Bures}]^{-1} \cdot \hat{T}_s^\dagger - \hat{T}_s \cdot [\hat{\Omega}_\sigma^{Bures}]^{-1} = \frac{7}{600} [\mathbb{1} \otimes Y + Y \otimes \mathbb{1}], \tag{71}$$

where

$$Y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{72}$$

The family of quantum detailed balance conditions is therefore much richer than the classical counterpart.

V. QUANTUM CHEEGER'S INEQUALITY

In the context of classical stochastic processes a very powerful formalism has been developed, often referred to as the conductance bound or Cheeger's inequality, to bound convergence rates of stochastic processes. We will generalize this to the quantum setting in this section. Similar results have appeared in Ref. 30. The gap of the map S_k is defined as the difference between the largest and second largest eigenvalue, $\Delta = 1 - \lambda_1$. The gap can be characterized in a variational fashion.²³

Proposition 19: *The gap of the map $S_k = [\Omega_\sigma^k]^{-1/2} \circ T^* \circ \Omega_\sigma^k \circ T \circ [\Omega_\sigma^k]^{-1/2}$ is given by*

$$\Delta = \min_{X \in \mathcal{M}_d} \frac{\langle X, (\text{id} - S_k)X \rangle}{\frac{1}{2} \|(X \otimes \sqrt{\sigma} - \sqrt{\sigma} \otimes X)\|_{HS}^2}, \tag{73}$$

where $\|A\|_{HS}^2 = \text{tr}[A^\dagger A]$ denotes the standard Hilbert–Schmidt norm and $\langle \cdot, \cdot \rangle$ the corresponding Hilbert–Schmidt scalar product.

Proof: The eigenvector that corresponds to the eigenvalue $\lambda_0 = 1$ of S_k is given by $\sqrt{\sigma}$. The gap can therefore be written as²³

$$\begin{aligned} \Delta &= \min_{X \in \mathcal{M}_d, \text{tr}[X\sqrt{\sigma}] = 0} 1 - \frac{\text{tr}[X^\dagger S(X)]}{\text{tr}[X^\dagger X]} \\ &= \min_{X \in \mathcal{M}_d, \text{tr}[X\sqrt{\sigma}] = 0} \frac{\text{tr}[X^\dagger (X - S(X))]}{\text{tr}[X^\dagger X] - \text{tr}[X\sqrt{\sigma}]^2} \\ &= \min_{X \in \mathcal{M}_d} \frac{\text{tr}[X^\dagger (X - S(X))]}{\frac{1}{2} \|(X \otimes \sqrt{\sigma} - \sqrt{\sigma} \otimes X)\|_{HS}^2}. \end{aligned} \tag{74}$$

Note that the constrained $\text{tr}[X\sqrt{\sigma}] = 0$ can be dropped in the last line. Suppose that $\text{tr}[X\sqrt{\sigma}] = c$, we can then define $X' = X - c\sqrt{\sigma}$ and vary X' since the equation is invariant under such shifts. \square

Throughout the remainder of this section we consider unital quantum channels, i.e., maps which obey $T(\mathbb{1}) = \mathbb{1}$. For this case it is ensured that already the simple map $S = T^* \circ T$ has a spectrum that is contained in $[0, 1]$, since all Ω_σ^k coincide and correspond to the identity map. The χ^2 -divergence just reduces to the standard Hilbert–Schmidt inner product times a prefactor given by $d = \dim(\mathfrak{H})$. In the case of a detailed balanced stochastic map, it even suffices to just consider the map itself. In either case we will denote the corresponding map as S from now on. The variational characterization of the gap Δ now allows us to give an upper as well as a lower bound to the second largest eigenvalue of S .

Lemma 20: *Let $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$ be a unital quantum channel. Then the second largest eigenvalue λ_1 of its symmetrization $S = T^* \circ T$, is bounded by*

$$1 - 2h \leq \lambda_1 \leq 1 - \frac{1}{2}h^2, \tag{75}$$

where h is Cheeger's constant defined as

$$h = \min_{\Pi_A, \text{tr}[\Pi_A] \leq d/2} \frac{\text{tr}[(\mathbb{1} - \Pi_A)S(\Pi_A)]}{\text{tr}[\Pi_A]}. \tag{76}$$

The minimum is to be taken over all projectors Π_A on the space $A \subset \mathfrak{H}$, so that $\text{tr}[\Pi_A] \leq d/2$.

Proof: An upper bound to the gap is immediately found by choosing $X = \Pi_A$. Due proposition (19) we can write

$$\begin{aligned} \Delta &\leq \frac{\text{tr}[\Pi_A(\text{id} - S)(\Pi_A)]}{\text{tr}[\Pi_A^2] - \frac{1}{d}\text{tr}[\Pi_A]^2} \\ &= \frac{\text{tr}[(\mathbb{1} - \Pi_A)S(\Pi_A)]}{\frac{1}{d}\text{tr}[(\mathbb{1} - \Pi_A)]\text{tr}[\Pi_A]} \leq 2h, \end{aligned} \tag{77}$$

where in the last line we have used that $\text{tr}[\mathbb{1} - \Pi_A] \geq d/2$.

For the lower bound, we can restrict the minimization in (76) to diagonal projections. Furthermore, when considering only unital quantum channels, it is possible to reduce the problem of bounding the gap Δ to that of bounding the gap of a classical stochastic process. To see this, let us work on the basis where the eigenvector $X_1 \in \mathcal{M}_d$ corresponding to λ_1 is diagonal. We shall assume wlog that $X_1^\dagger = X_1$. Accordingly, we can write $X = \sum x_i |i\rangle\langle i|$. The numerator then becomes

$$\begin{aligned} \text{tr}[X^\dagger(X - S(X))] &= \sum_{ij} x_i x_j (\text{tr}[|i\rangle\langle i||j\rangle\langle j|] - \text{tr}[|i\rangle\langle i|S(|j\rangle\langle j|)]) \\ &= \sum_i x_i^2 - \sum_{ij} x_i x_j P_{ij} = \frac{1}{2} \sum_{ij} P_{ij} (x_i - x_j)^2. \end{aligned} \tag{78}$$

We introduced the matrix $P_{ij} = \langle i|S(|j\rangle\langle j|)|i\rangle$, which is a symmetric nonnegative matrix which obeys $P_{ij} \geq 0$, $\sum_i P_{ij} = 1$ and $P^T = P$. Hence P is doubly stochastic. Performing the same reduction in the denominator, we obtain

$$\frac{1}{2d} \|(X \otimes \mathbb{1} - \mathbb{1} \otimes X)\|_{HS}^2 = \frac{1}{2d} \sum_{ij} (x_i - x_j)^2. \tag{79}$$

Hence, we arrive at the classical version of Mihail’s identity,³

$$\Delta = \min_{\{x_i\}} \frac{\sum_{ij} P_{ij} (x_i - x_j)^2}{1/d \sum_{ij} (x_i - x_j)^2}. \tag{80}$$

Given the classical version of Mihail’s identity, the proof of the lower bound is the same as in the classical case. For completeness we repeat it here. First, we define, $z_i \equiv |x_i|x_i$ and write

$$\begin{aligned} \sum_{ij} P_{ij} |z_i - z_j| &= \sum_{ij} P_{ij} ||x_i|x_i - |x_j|x_j| \leq \sum_{ij} \sqrt{P_{ij}}\sqrt{P_{ij}}(|x_i| + |x_j|)(x_i - x_j) \\ &\leq \sqrt{\sum_{ij} P_{ij} (x_i - x_j)^2} \sqrt{\sum_{ij} P_{ij} (|x_i| + |x_j|)^2}, \end{aligned} \tag{81}$$

where we used Cauchy–Schwartz in the last step. Consider now,

$$\sum_{ij} P_{ij} (|x_i| + |x_j|)^2 = 2(\sum_i x_i^2 + \sum_{ij} |x_i|P_{ij}|x_j|) \leq 4 \sum_i |x_i|^2. \tag{82}$$

Furthermore, note that we can bound

$$1/d \sum_{ij} (x_i - x_j)^2 \leq 2/d \sum_{ij} x_i^2 = 2 \sum_i |z_i|. \tag{83}$$

We are therefore left with a lower bound to Mihail’s identity, which holds for all choices of $\{x_i\}$

$$\frac{1}{2} \left(\frac{\sum_{ij} P_{ij} |z_i - z_j|}{2 \sum_i |z_i|} \right)^2 \leq \frac{\sum_{ij} P_{ij} (x_i - x_j)^2}{1/d \sum_{ij} (x_i - x_j)^2}. \tag{84}$$

We shall now assume that $x_i \geq 0$ everywhere and we can hence drop the absolute values in the definition for the z_i . This assumption is valid since we are free to add an arbitrary constant $x_i \rightarrow$

$x_i + c$ to make all x_i positive. Note that we therefore are left with a lower bound to the gap of the form

$$\Delta \geq \frac{1}{2} \left(\frac{\sum_{ij} P_{ij} |x_i^2 - x_j^2|}{2 \sum_i x_i^2} \right)^2. \tag{85}$$

Let us focus on the right side of the inequality. Since

$$2 \sum_{i,j : x_i \geq x_j} P_{ij} (x_i^2 - x_j^2) = 4 \sum_{i,j : x_i \geq x_j} P_{ij} \int_{x_j}^{x_i} t \, dt = 4 \int_0^\infty t \sum_{ij : x_i > t \geq x_j} P_{ij} \, dt, \tag{86}$$

and furthermore,

$$\sum_{ij : x_i > t \geq x_j} P_{ij} = \sum_{i \in A(t)} \sum_{j \in A^c(t)} P_{ij} \quad \text{where, } A(t) \equiv \{i | x_i \geq t\}, \tag{87}$$

we can bound

$$4 \int_0^\infty t \sum_{ij : x_i > t \geq x_j} P_{ij} \, dt \geq h \, 4 \int_0^\infty t \sum_{i \in A(t)} \Theta(t - x_i) \, dt = 2 h \left(\sum_i x_i^2 \right), \tag{88}$$

where we defined h in a similar fashion as above. We have therefore found the desired lower bound for the spectral gap of the map S . □

A. Example: Conductance bound for unital qubit channels

A convenient basis for the matrix space \mathcal{M}_2 associated with the Hilbert space $\mathfrak{H} = \mathbb{C}^2$ is given in terms of the Pauli basis $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$. On this basis a density matrix $\rho \in \mathcal{S}_2$ can be parametrized in terms of its Bloch vector $\mathbf{r} \in \mathbb{R}^3$. In the Bloch representation the density matrix reads $\rho = \frac{1}{2} (\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_x, \sigma_y, \sigma_z)$. It is also straightforward to determine the matrix representation of a quantum channel $T : \mathcal{M}_2 \rightarrow \mathcal{M}_2$ with respect to the Pauli basis. A general channel can be written as a matrix $\hat{T} \in \mathcal{M}_4$

$$\hat{T} = \begin{pmatrix} 1 & 0 \\ \mathbf{t} & \mathbf{L} \end{pmatrix}. \tag{89}$$

The channel acts on a density matrix via $T(\rho) = T(\frac{1}{2} (\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\Sigma})) = \frac{1}{2} (\mathbb{1} + (\mathbf{t} + \mathbf{L}\mathbf{r}) \cdot \boldsymbol{\Sigma})$. It can be shown that the map T is unital if and only if $\mathbf{t} = 0$. Let us now consider the optimization for Cheeger’s constant h as given in lemma (20). Given the constraint, we have to vary all one-dimensional projectors $\Pi_A = |\psi\rangle\langle\psi|$ with $\| |\psi\rangle \|_2 = 1$, so that

$$h = \min_{|\psi\rangle \in \mathbb{C}^2} \text{tr} [(\mathbb{1} - |\psi\rangle\langle\psi|) S(|\psi\rangle\langle\psi|)]. \tag{90}$$

The symmetrized map S of the unital channel T , with $\mathbf{t} = 0$, now assumes the matrix representation,

$$\hat{S} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{L}^\dagger \mathbf{L} \end{pmatrix}. \tag{91}$$

Furthermore note that any projector $|\psi\rangle\langle\psi| \in \mathcal{S}_2$ can be parametrized via a Bloch vector $\mathbf{a} \in \mathbb{R}^3$ that obeys $\| \mathbf{a} \|_2 = 1$. The minimization for Cheeger’s constant reduces therefore to

$$h = \min_{\| \mathbf{a} \|_2 = 1} 1 - \langle \mathbf{a} | \mathbf{L}^\dagger \mathbf{L} | \mathbf{a} \rangle, \tag{92}$$

where $\langle \mathbf{a} | \mathbf{b} \rangle$ denotes the canonical scalar product in \mathbb{R}^3 . The minimum is attained when \mathbf{a} is the eigenvector associated with the largest eigenvalue s_1^2 of the matrix $\mathbf{L}^\dagger \mathbf{L}$. Hence for an arbitrary single qubit unital channel, Cheeger’s constant is given by $h = 1 - s_1^2$, where s_1 is the largest singular value of the matrix \mathbf{L} and hence the second largest singular value of the channel T . We see that the conductance bound as stated in (20) is indeed satisfied, since

$$2s_1^2 - 1 \leq s_1^2 \leq \frac{1}{2}(1 + s_1^2). \tag{93}$$

VI. CONCLUSION AND DISCUSSION

We have seen that by generalizing the χ^2 -divergence to the quantum setting, many of the classical results for the convergence of Markov processes can be recovered. The general perception that the convergence should be governed by the spectral properties of the quantum channel could be verified in the asymptotic limit. The fact that we were working with noncommuting probabilities gave rise to a larger set of possibilities of defining an inversion of the fixed point density matrix, all of which give rise to a valid upper bound for the trace distance. An interesting question is how the different singular values s_i^k of the corresponding quantum discriminant relate to each other. The generalization of the χ^2 -divergence also led to the definition of detailed balance for quantum channels. Again, not only a single condition for quantum detailed balance exists but an entire family of conditions each determined by a different function $k \in \mathcal{K}$, all of which coincide in the case when we consider classical stochastic processes on a commuting subspace. The quantum concept of detailed balance therefore appears to be richer and allows for a wider set of channels to obey this definition. The conductance bound that was derived could only be shown for unital quantum channels. However we would like to point out that it is possible to give conductance bounds for classical maps when the Markov chain is not doubly stochastic. The fact that in general we may not assume that the fixed point of an arbitrary channel commutes with the eigenvector associated with the second largest eigenvalues seems to hinder a generalization for nonunital channels. Moreover, the classical conductance bound has a nice geometrical interpretation in terms of the cut set analysis and the maximal flow on the graph associated with the stochastic matrix P_{ij} . When investigating general quantum channels such a nice geometric interpretation seems to be lacking. For unital quantum channels Cheeger's constant can also be viewed in terms of the minimal probability flow of one subspace to its complement.

ACKNOWLEDGMENTS

Part of this work was done during a workshop at the Erwin Schrödinger Institute for Mathematical Physics. Wolf and Kastoryano thank David Reeb for helpful discussions. Temme was supported by the FWF doctoral program CoQuS. Verstraete is supported by the FWF grant FoQuS and the European grants QUEVADIS and QUERG. Kastoryano was supported in part by the NBIA. Wolf acknowledges the financial support by the Danish Research Council (FNU), QUANTOP and the European projects QUEVADIS and COQUIT. Ruskai was partially supported by the US National Science Foundation Grant DMS-0604900.

APPENDIX A: FAMILIES OF DIVERGENCES

The most widely used family of divergences, often called α -divergence (see Ref. 38, Chapter 7), is associated with the functions

$$k_\alpha^{\text{WYD}}(w) = \frac{(1-w^\alpha)(1-w^{1-\alpha})}{\alpha(1-\alpha)(1-w)^2} \quad \text{for} \quad \alpha \in [-1, 2]. \quad (\text{A1})$$

This family is sometimes called the WYD divergences, because it arises from an extension of the Wigner–Yanase–Dyson entropy^{18,41} associated with the (unsymmetrized) function $g(w) = \frac{1}{\alpha(1-\alpha)}(w-w^\alpha)$. In the limit $\alpha \rightarrow 1$ this yields³⁹ the familiar (asymmetric) relative entropy $H(\rho, \sigma) = \text{Tr} \rho(\log \rho - \log \sigma)$ and Ω_p^{\log} given by (54). Like the family of divergences introduced here, the minimal WYD divergence occurs for $\alpha = 1/2$, it is convex in α , symmetric around $\alpha = 1/2$ and yields the maximal $\frac{1+w}{2w}$ when $\alpha = -1$ or 2. However, $\alpha = 1/2$ gives $k_{1/2}^{\text{WYD}}(w) = 4(1 + \sqrt{w})^{-2}$ which is quite different from $k_{1/2}^{\text{mean}}(w) = w^{-1/2}$. The WYD family is often studied only for $\alpha \in (0, 1)$; it was first observed by Hasegawa in⁴⁰ that it yields a monotone metric if and only if $\alpha \in [-1, 2]$.

The metrics associated with $k_\alpha^{\text{mean}}(w)$ and $k_\alpha^{\text{WYD}}(w)$ both give increasing families for $\alpha \geq \frac{1}{2}$ and both yield the maximal metric $k(w) = (1+w)(2w)$ for α the maximal values of 1 and 2, respectively. However, neither reduces to the minimal metric $k(w) = 2/(1+w)$. The measure $\delta(s-a)$ in (16) leads to the family $k_a(w) = \frac{(1+a)^2}{2} \frac{(1+w)}{(1+wa)(w+a)}$ for $a \in [0, 1]$ which reduces to the the maximal and

minimal functions for $a = 0, 1$. However, this family is neither increasing nor decreasing. Hansen⁴¹ has found families of functions which increase monotonically from the smallest to the largest of which we mention only

$$k_a(w) = w^{-a} \left(\frac{1+w}{2} \right)^{2a-1} \quad \text{for } a \in [0, 1]. \quad (\text{A2})$$

APPENDIX B: PROOF OF A KEY INEQUALITY

The proof of the contractivity of a general Riemannian metric is based on the following theorem first proved in Ref. 18.

Theorem 21: For a channel $T : \mathcal{M}_d \rightarrow \mathcal{M}_d$, we have that

$$\begin{aligned} \text{tr} \left[A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right] &= \text{tr} \left[T \left(A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right) \right] \geq \\ &\text{tr} \left[T(A)^\dagger \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \end{aligned} \quad (\text{B1})$$

Proof: Let $\sigma > 0$, then $\text{tr}[A^\dagger \sigma A] \geq 0$, and $\text{tr}[A^\dagger A \sigma] \geq 0$ so that L_σ as well as R_σ are both positive semidefinite super operators on the matrix space. Therefore we infer that for a positive $\rho > 0$ the operator $R_\sigma + sL_\rho$ is also positive. We define a matrix $X = [R_\sigma + sL_\rho]^{-1/2}(A) + [R_\sigma + sL_\rho]^{1/2}T^*(A)$ and furthermore $B = [R_{T(\sigma)} + sL_{T(\rho)}]^{-1}T(A)$. Since $\text{tr}[X^\dagger X] \geq 0$, we have that

$$\text{tr} \left[A^\dagger \frac{1}{R_\sigma + sL_\rho} A \right] - \text{tr} [T^*(B^\dagger)A] - \text{tr} [A^\dagger T^*(B)] + \text{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] \geq 0. \quad (\text{B2})$$

Furthermore note that

$$-\text{tr} [A^\dagger T^*(B)] - \text{tr} [T^*(B^\dagger)A] = -2\text{tr} \left[T(A^\dagger) \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \quad (\text{B3})$$

It therefore suffices to show that we are able to bound the last term in (B2) by the right side of the inequality (B1). Note that

$$\begin{aligned} \text{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] &= \text{tr} [T^*(B^\dagger)T^*(B)\sigma + sT^*(B^\dagger)\rho T^*(B)] \\ &\leq \text{tr} [T^*(B^\dagger B)\sigma + sT^*(B B^\dagger)\rho], \end{aligned} \quad (\text{B4})$$

since $\rho, \sigma > 0$ and due to the operator inequality $T^*(B^\dagger)T^*(B) \leq T^*(B^\dagger B)$. This inequality holds for any B since T is a channel and by that trace preserving, hence $T^*(\mathbb{1}) = \mathbb{1}$. With $\text{tr} [T^*(B^\dagger B)\sigma] = \text{tr} [B^\dagger B T(\sigma)]$, we can write

$$\begin{aligned} \text{tr} [T^*(B^\dagger)[R_\sigma + sL_\rho]T^*(B)] &\leq \text{tr} [B^\dagger B T(\sigma) + sB^\dagger B T(\rho)] \\ &= \text{tr} [B^\dagger [R_{T(\sigma)} + sL_{T(\rho)}]B] = \text{tr} [B^\dagger T(A)] = \text{tr} \left[T(A^\dagger) \frac{1}{R_{T(\sigma)} + sL_{T(\rho)}} T(A) \right]. \end{aligned} \quad (\text{B5})$$

□

¹P. Diaconis and D. Stroock, "Geometric bounds on eigenvalues of Markov chains," *Ann. Appl. Probab.* **1**, 36 (1991).

²J. A. Fill, "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains with an application to the exclusion process," *Ann. Appl. Probab.* **1**, 62 (1991).

³M. Mihail, Ph.D. dissertation, Department of Computer Science (Harvard University, Cambridge, MA, 1989).

⁴D. A. Levin, Y. Peres and E. L. Wilmer "Markov chains and mixing times," *Am. Math. Soc.*, (2008)

⁵K. Pearson, "On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling," *Philos. Mag.* **50**, 157 (1900).

⁶F. Verstraete, M. M. Wolf and J. I. Cirac, "Quantum computation and quantum-state engineering driven by dissipation," *Nature Phys.* **5**, 633 (2009).

⁷K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, "Quantum metropolis sampling," e-print arXiv:0911.3635v1.

- ⁸M. Fannes, B. Nachtergaele, and R. F. Werner, "Finitely correlated states on quantum spin chains", *Commun. Math. Phys.* **144**, 443 (1992)
- ⁹S. Rommer and S. Östlund, "Class of ansatz wave functions for one-dimensional spin systems and their relation to the density matrix renormalization group," *Phys. Rev.* **B55**, 2164 (1997).
- ¹⁰A. S. Holevo, *Statistical Structure of Quantum Theory*, Springer Lecture Notes in Physics (Springer-Verlag, New York, 2001).
- ¹¹G. Lindbald, *Commun. Math. Phys.* **48**, 119 (1976).
- ¹²M. Sanz, D. Perez-Garcia, M. M. Wolf, and J. I. Cirac, "A quantum version of Wielandt's inequality," e-print arXiv:0909.5347.
- ¹³D. Petz, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.* **23**, 57 (1986).
- ¹⁴D. Petz, "Monotone metrics on matrix spaces," *Linear Algebr. Appl.* **244**, 81 (1996).
- ¹⁵D. Petz and Cs. Sudar, "Geometries of quantum states," *J. Math. Phys.* **37**, 2662 (1996).
- ¹⁶D. Petz and M. B. Ruskai, "Contraction of generalized relative entropy under stochastic mappings on matrices," *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1**, 83 (1998).
- ¹⁷D. Petz and C. Ghinea, "Introduction to quantum Fisher information," e-print arXiv:1008.2417 (2010).
- ¹⁸A. Lesniewski and M. B. Ruskai, "Monotone Riemannian metrics and relative entropy on noncommutative probability spaces," *J. Math. Phys.* **40**, 5702 (1999).
- ¹⁹H. Araki, "Relative entropy of state of von Neumann algebras," *Publ. RIMS Kyoto Univ.* **9**, 809 (1976).
- ²⁰H. Araki, "Recent progress in entropy and relative entropy," *Proceedings of the 8th International Congress on Mathematical Physics*, edited by R. Seneor and M. Mebkhout (World Scientific, Singapore, 1987), pp. 354–365.
- ²¹R. Bhatia, *Matrix Analysis* (Springer-Verlag, New York, 1997).
- ²²D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publ. RIMS. Kyoto Univ.* **21**, 781 (1985), *Infinite Dimensional Anal.* **1**, 83 (1998).
- ²³R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2007).
- ²⁴R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 2006).
- ²⁵H. Araki, "On an inequality of Lieb and Thirring," *Lett. Math. Phys.* **19**, 167 (1990).
- ²⁶M. B. Ruskai, "Beyond strong subadditivity? Improved bounds on the contraction of generalized relative entropy," *Rev. Math. Phys.* **6**, 1147 (1994).
- ²⁷M. B. Ruskai and F. H. Stillinger, "Convexity inequalities for estimating free energy and relative entropy," *J. Phys. A* **23**, 2421 (1990).
- ²⁸E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture" *Adv. Math.* **11**, 267 (1973).
- ²⁹A. Frigerio, "Simulated annealing and quantum detailed balance," *J. Stat. Phys.* **58**, 325 (1990).
- ³⁰M.B. Hastings, "Random unitaries give quantum expanders," *Phys. Rev. A* **76**, 032315 (2007).
- ³¹E. A. Morozova and N. N. Chentsov, "Markov invariant geometry on the state manifolds (Russian)," *Itogi Nauki Tekh.* **36**, 69 (1990).
- ³²D. Perez-Garcia, M. M. Wolf, D. Petz, and M. B. Ruskai, *J. Math. Phys.* **47**, 083506 (2006).
- ³³B. M. Terhal and D. P. DiVincenzo, *Phys. Rev. A* **61**, 2301 (2000).
- ³⁴R. Alicki, *Rep. Math. Phys.* **10**, 249 (1976).
- ³⁵A. Kossakowski, A. Frigerio, V. Gorini, and M. Verri, *Commun. Math. Phys.* **57**, 97 (1977).
- ³⁶W. A. Majewski, "The detailed balance condition in quantum statistical mechanics," *J. Math. Phys.* **25**, 614 (1984).
- ³⁷A. Majewski and R. F. Streater, "Detailed balance and quantum dynamical maps," *J. Phys. A* **31**, 7981 (1998).
- ³⁸A. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs Vol. 191 (American Mathematical Society, Providence and Oxford University Press, New York, 2000).
- ³⁹A. Jenčová and M. B. Ruskai, "A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality," *Rev. Math. Phys.* **22**, 1099 (2010).
- ⁴⁰H. Hasegawa, " α -divergence of the non-commutative information geometry," *Reports Math. Phys.* **33**, 87 (1993).
- ⁴¹F. Hansen, "Metric adjusted skew information," *Proc. Nat. Acad. Sci. USA* **105**, 9909 (2008).