UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Information and thermodynamics

Matteo Scandi

Institut de Ciències Fotòniques

PhD Thesis

Information and thermodynamics

by

Matteo Scandi

Supervisor: **Antonio Acín**

July, 2023

Contents

List of publications

Publications contained in this work

[1] **M. S.**, P. Abiuso, D. De Santis, and J. Surace, "Quantum Fisher information and its dynamical nature," arXiv.2304.14984 (2023).

[2] J. Surace, **M. S.**, "State retrieval beyond Bayes' retrodiction and reverse processes," Quantum 3, 197 (2023).

[3] **M. S.**, D. Barker, S. Lehmann, K. A. Dick, V. F. Maisi, and M. Perarnau-Llobet, "Minimally dissipative information erasure in a quantum dot via thermodynamic length," Physical Review Letters 129, 270601 (2022).

[4] P. Abiuso, **M. S.**, D. De Santis, and J. Surace, "Characterizing (non-)Markovianity through Fisher information," arXiv.2204.04072 (2022).

[5] D. Barker, **M. S.**, S. Lehmann, C. Thelander, K. A. Dick, M. Perarnau-Llobet, and V. F. Maisi, "Experimental verification of the work fluctuation-dissipation relation for information-to-work conversion," Physical Review Letters 128, 040602 (2022).

[6] P. Abiuso, H. J. D. Miller, M. Perarnau-Llobet, and **M. S.**, "Geometric optimisation of quantum thermodynamic processes," Entropy 22, 1076 (2020).

[7] **M. S.**, H. J. D. Miller, J. Anders, and M. Perarnau-Llobet, "Quantum work statistics close to equilibrium," Physical Review Research 2, 023377 (2020).

[8] **M. S.**, H. J. D. Miller, J. Anders, and M. Perarnau-Llobet, "Work fluctuations in slow processes: quantum signatures and optimal control," Physical Review Letters 123, 230603 (2019).

[9] **M. S.** and M. Perarnau-Llobet, "Thermodynamic length in open quantum systems," Quantum 3, 197 (2019).

Publications not contained in this work

[10] **M. S.** and J. Surace, "Undecidability in resource theory: Can you tell resource theories apart?" Physical Review Letters 127, 270501 (2021).

[11] N. Pancotti, **M. S.**, M. T. Mitchison, and M. Perarnau-Llobet, "Speed-ups to isothermality: Enhanced quantum thermal machines through control of the system-bath coupling," Physical Review X 10, 031015 (2020).

Abstract

The modern understanding of physics is deeply linked with the concept of information. The revival of the study of quantum mechanics in the form of quantum information is just an example of a more general trend showing how the incorporation of ideas from information theory into the practice of physics is not simply a fertile opportunity to find new results, but it also offers a radically new understanding of what physics should be describing: emblematic of this paradigm shift is the transition from the infinite dimensional space associated to the wave functions appearing in the Schrödinger equation, to the extreme simplicity of the modern cornerstone of quantum mechanics, the qubit. A special place in this context is taken by thermodynamics: on the one hand, because it was one of the first branches of physics in which the role of information was explicitly recognised; on the other, as the formal correspondence between Shannon and Boltzmann entropy hints at a deep connection between the two. Ultimately, it almost feels like these two theories will end up coinciding, and one will speak about thermodynamics just as information theory with erasure. We are still far away from this claim, but the continuous appearance of information quantifiers in genuinely thermodynamics setting, especially when characterising the dissipation, cannot but corroborate this belief.

The aim of this thesis is to move further in the identification of the two theories, by focusing on some aspects of information geometry and showing how these naturally apply to the study of thermodynamic transformations. In particular, the main object of interest is the family of quantum Fisher information metrics, thoroughly studied in the first part of the thesis. In there we prove a fact that motivates the interest in this quantity: despite the statistical setting in which it was originally formulated, it has such a deep dynamical nature that all physical evolutions can actually be defined just in terms of their behaviour with respect to

the Fisher information. In the second part of the thesis, we connect this discussion to the field of thermodynamics. In this context, we show that the Fisher information metrics naturally emerge in the description of the dissipation in near-isothermal transformations, that is whenever the driving is slow enough for the system to be close to equilibrium during the whole protocol. This shows another example of what seems to be a general rule: the naturality with which structures developed in the context of statistical inference and information theory apply to the study of entropy production.

Resumen

La comprensión moderna de la física está enlazada al concepto de información. El renacimiento en el interés por la mecánica cuántica causado por el nacimiento de la información cuántica es solo un ejemplo de una tendencia más general que muestra cómo la incorporación de ideas de la teoría de la información en la práctica de la física no es simplemente una oportunidad fértil para encontrar nuevos resultados, sino que también ofrece una comprensión radicalmente nueva de lo que la física debería describir: emblemático de este cambio de paradigma es la transición del espacio de dimensión infinita asociado a las funciones de onda que aparecen en la ecuación de Schrödinger, a la extrema simplicidad de la reciente piedra angular de la mecánica cuántica, el qubit. Un espacio especial en este contexto lo ocupa la termodinámica: por un lado, porque fue una de las primeras ramas de la física en la que se reconoció explícitamente el papel de la información; por el otro, por el hecho que la correspondencia formal entre la entropía de Shannon y la de Boltzmann sugiere una conexión profunda entre las dos. En última instancia, casi parece que estas dos teorías terminarán coincidiendo, y se hablará sobre termodinámica como la teoría de la información con pérdidas. Todavía estamos lejos de esta afirmación, pero la continua aparición de cuantificadores de información en contextos genuinamente termodinámico, especialmente al momento de caracterizar la disipación, no puede no corroborar esta creencia.

El objetivo de esta tesis es avanzar en la identificación de las dos teorías, centrándose en algunos aspectos de la geometría de la información y mostrando cómo se aplican naturalmente al estudio de las transformaciones termodinámicas. En particular, el principal objeto de interés es la familia de métricas de información cuántica de Fisher, ampliamente estudiada en la primera parte de la tesis. Allí probamos un hecho que motiva el interés en esta cantidad: a pesar del contexto estadístico en el que se formuló originalmente, tiene una natura dinámica tan profunda

que todas las evoluciones físicas pueden definirse solo en términos de sus comportamientos con respecto a la Información de Fisher. En la segunda parte de la tesis, conectamos esta discusión con el campo de la termodinámica. En este contexto, mostramos que las métricas de información de Fisher emergen naturalmente en la descripción de la disipación en transformaciones casi-isotérmicas, es decir, siempre que el desplazamiento sea lo suficientemente lento como para que el sistema esté cerca del equilibrio durante todo el protocolo. Esto muestra otro ejemplo de lo que parece ser una regla general: la naturalidad con la que estructuras desarrolladas en el contexto de la inferencia estadística y de la teoría de la información se aplican al estudio de la producción de entropía.

Overview

This work is naturally divided in two parts, distinguished both in subject and style. The first part, about the definition and characterisation of the Fisher information metric, is mathematical in nature, rather technical, and aims at giving an organic treatment to a plethora of results that were scattered in the literature. The second part of the thesis takes a more physical stance, which results in a discursive style and less heavy formalism. The first two chapters of this part, about the first and second law of thermodynamics, do not contain original work, but are instead dedicated to what I would have liked to know during my master, when I was first exposed to the subject of modern thermodynamics.

Despite the order of presentation would suggest otherwise, depending on the inclination of the reader, it is possible to start from the second part and to use the first simply as an appendix. Still, given the significance of the results contained in Chapter 1, we preferred to dedicate a full chapter to the Fisher information, rather than to delegate it to an appendix. Finally, we also refer to Chapter 5 for a the summary of the results obtained during the PhD.

Part I: Information

Chap. 1: Fisher information and the geometry of quantum states

The problem of assessing the similarity of different classical statistical distributions does not have a straightforward answer: different methods yield different quantifications, and there is no clear argument that would single out a unique strategy over all the others. As a matter of fact, this difficulty actually reflects the many different behaviours that a probability distribution shows depending on the regime one is focusing on: two distributions could be very similar in the asymptotic regime, but show substantial differences when one restricts their attention to single-shot

experiments. For this reason, rather than trying to reduce the richness of the phenomenology that one can focus on, Csizár introduced a family of possible quantifiers, the contrast functions (see Sec. 1.1), which encompasses the many different aspects that one could want to characterise. Remarkably, despite the plethora of different possibilities, when one considers close-by probability distributions (in the sense that their statistics are very similar) all the contrast functions collapse into a single quantity: the Fisher information metric.

For quantum states, however, this simplification does not happen: the local behaviour of different contrast functions only coincides for diagonal states, i.e., classical statistics, but otherwise to each (symmetrised) contrast function there corresponds a different local metric. This family is what one refers to as the quantum Fisher information metrics, and the non-uniqueness of these quantities shows that the estimation scheme one is considering stays relevant even for close-by states. The resulting rich phenomenology is discussed in Sec. 1.4, which can be considered as a field guide to the many different mathematical expressions that the quantum Fisher information can take.

It should be noticed that there is another important way of characterising the Fisher information: the Chentsov theorem states that this is the unique metric on the space of probability distributions that contracts under arbitrary stochastic maps (Thm. 1). A similar result was later obtained by Petz (Thm. 3), showing that the same family of quantum Fisher information discussed above can also be characterised as all the metrics that contract under arbitrary physical evolutions. There are two important consequences of this theorem: first, it legitimates the identification of the family of quantum Fisher information obtained from the local expansion of contrast functions as the correct extension of the classical Fisher information to quantum systems (if this second characterisation was not present, one could wonder whether it would be possible to single out a unique quantum Fisher information from additional axioms); on the other hand, it reveals a hidden link between the local behaviour of distinguishability measures and physical evolutions.

This second aspect of the Fisher information metric remained long unexplored. Only recently we found that the Chentsov-Petz theorem can be actually reversed: one can define all physical evolutions as exactly the unique family of linear maps that contract the Fisher information (see Thm. 5 and [1, 4]). In this way, whereas the Chentsov-Petz theorem would suggest that the Fisher metric is a derived concept, as it can be

defined in terms of the dynamics, in reality there is no hierarchy between the two, as one can also characterise the physical evolutions starting from the Fisher information.

This connection is further explored in Sec. 1.5.1–1.5.5, where we show that a number of properties of dynamical semigroups can be formulated in terms of the relation between their generators and the Fisher information metrics. In particular, Markovianity is equivalent to a monotonous contraction of the Fisher information on the space of states (together with an ancilla of the same dimension, see Thm. 9), while detailed balance corresponds to a self-adjoint generator with respect to the scalar products induced by the Fisher metrics.

The results just discussed demonstrate the intimate dynamical nature of the Fisher information metric, hinting at a subtle connection between statistics and physical evolutions that was so far overlooked.

Part II: Thermodynamics

Chap. 2: The first law

The first law of thermodynamics has two main messages: first, it encodes the law of conservation of energy; secondly, it naturally splits the energy into an accessible and an inaccessible component, reflecting the fact that an agent only has control over a part of the universe, namely the system. Indeed, any energy inserted on the system through a driving (i.e., in the form of work), will eventually flow into the environment due to a thermalisation process, becoming de facto irretrievable. In order to account for this apparent local violation of the conservation of energy, one is led to introduce the concept of heat, as all that energy disappearing into the thermal bath.

Interestingly, while this division of the total energy into work and heat is unproblematic for classical systems, in the quantum regime one has to be more careful, as the measurement scheme employed becomes relevant to the definition of these two quantities. Indeed, since the Hamiltonian is driven, it could in general not commute with itself at different times, so there are incompatibility issues in the definition of a probability distribution for the work (or, equivalently, for the heat). The most standard approach in this context is to define it through the two-point measurement (TPM) scheme: at the beginning and at the end of the protocol, one performs a global measurement of the energy of system and bath, and defines the change of energy as the difference between the two outcomes.

Since there are no other energy sources, this equals the work, as the only change of energy can come from the driving of the system Hamiltonian. In this way, one can define the probability of extracting a given amount of work w as the sum of the probabilities of all possible energy gaps equal to w (see Eq. (2.16)).

Setting aside the feasibility of performing a global measurement on system and bath (which is usually considered to be macroscopic), this definition has an important drawback: if there is any coherence in the initial state of the system, the average change of energy is actually different from the average computed from the probability defined above. Indeed, the initial energy measurement dephases the state, erasing any memory of the initial coherences. Moreover, in this context one can always find a non-contextual ontological model reproducing the statistics of the work, meaning that any genuinely quantum phenomenon is actually washed away by the initial measurement (Thm. 16).

Given these two important issues, one could wonder why we considered this definition in the first place. It is true that the work defined for classical systems is actually compatible with the measurement scheme just discussed, but one might wonder whether there is a way of modifying it by preserving this compatibility, and recover the equality between the difference of average energy and the average computed from the probability distribution. The negative answer to this endeavour is given by Thm. 15: there is no definition of a quantum work probability distribution satisfying both these conditions.

At this point there are two possible ways forward: either one sticks with the definition of work arising from the TPM scheme, aware of its limitations and scope, which is the approach taken in this thesis; or one drops the assumption that the work is actually connected to a probability distribution. Indeed, if one relaxes the definition to quasiprobabilities, not only one can have the compatibility with both the requirements discussed above, but can also obtain genuinely quantum phenomena (i.e., not reproducible through non-contextual models, see Thm. 17). The only drawback of this approach is that one partly loses the connection between dissipation and information quantifiers (as one can see by comparing Eq. (2.28) with Eq. (2.51)), which is the main reason to consider the two point energy measurement scheme definition of work.

Chap. 3: The second law

The second law of thermodynamics tells us that generic transforma-

tions cannot be undone, as the universe has the tendency to maximise its entropy with time. The tension between this statement and the reversibility of microscopic laws of motion led Maxwell to formulate a thought experiment that would demonstrate the statistical nature of the second law: his famous demon, i.e, a mechanism with complete control over the atomic elements of the system which can extract work from a thermal state without inputting any extra energy in the system.

Despite the advent of statistical mechanics, Maxwell's demon has had the power to stir controversy for over a century, with discussions mostly focussed on finding loopholes in the argument which would protect the second law even in the extreme regime in which the demon was formulated. As recounted in Sec. 3.1, rather than being just an academic skirmish, this debate led to one of the biggest revolution of modern physics: the idea that information is actually physical. Indeed, the pioneering approach of Szilárd showed that there was an impediment to the operation of Maxwell's demon coming from the way in which information was handled in its memory. In its most modern formulation, the arguments hinted at by Szilárd became the basis for the Landauer's principle, the fact that the erasure of a bit of information must dissipate at least $\mathrm{k_B}T\log 2$ of energy. As the memory of the demon is finite, this means that in order to continue operating it has to start erasing it at some point, releasing heat into the environment, and ultimately saving the second law from this attack.

Still, it should be noted that, despite the indisputable importance of the attempts just discussed, the second law is indeed statistical in nature. It is a fact that there are fluctuations in the entropy, signalling a local and sporadic inversion of the arrow of time, without the need to go to astronomical periods as the ones involved in the Poincaré recurrence theorem. Indeed, whereas these inversions are seldom observed at a macroscopic scale, when moving to smaller systems they become more and more relevant. It is quite remarkable, then, that some of the most important new results in modern thermodynamics come from taking these fluctuations seriously.

Indeed, whereas the universality of most of thermodynamics' results requires the system to be always at equilibrium, it was proven by Jarzynski that one could still find universal results by averaging over many realisations of the same transformation. His celebrated fluctuation theorem (see Eq. (3.22)) connects the statistics of work extracted during transformations that bring the system arbitrarily out of equilibrium to a

functional of the difference of free energy, i.e., to a state function. It is important to highlight that this result could only be derived once the existence of fluctuations in any thermodynamic transformation was actually acknowledged.

Along the same lines, Crooks relations (see Eq. (3.44)) connect the dissipation during a thermodynamic transformation with the probability of observing it as compared with its time reversed version. In particular, it shows that a system starting at equilibrium has the tendency to move along the time direction that dissipate more. It should be pointed out though that all the results presented are based on the assumptions that the state at the beginning of the transformation is thermal, so one cannot use Crooks relations to directly infer the directionality in the arrow of time. Still, this result shows that once the thermalisation of generic systems is proven, the asymmetry in time naturally emerges even when bringing them out of equilibrium.

These fluctuation theorems were initially deduced for classical systems. Still, when one uses the TPM definition of work, it can be shown that they also hold in the quantum regime. Interestingly, it has also been proven that this is a defining property of the TPM scheme: Thm. 20 states that any probability distribution satisfying a minimal compatibility with the classical definition of work and for which the Jarzynski equality holds coincides almost everywhere with the TPM definition of work. This result also justifies the interest in the TPM scheme, and allows for a cleaner restatement of the no-go theorem from Chapter 2, in the form of Thm. 21.

Chap. 4: Thermodynamic transformations close to equilibrium

Thermodynamics applies to a huge spectrum of systems, ranging from gases in a box to biological entities. Apart from the clear structural differences between the two examples provided, from a thermodynamic point of view what makes the first system much more tractable than the second, is that the transformations considered in the study of gases are usually ideal, i.e., in which the system is always at thermal equilibrium, while, in order to function, biological systems need to be out of equilibrium. Indeed, even an unstructured medium as can be an ideal gas is difficult to describe when one moves away from equilibrium.

Restricting the scope of the investigation to ideal transformations only, while making the applicability of the results found less general, allows for the definition of universal laws. The most famous example in

this context is the efficiency of a Carnot engine: if the working medium is always at equilibrium during the cycle, the efficiency of the engine only depends on the temperatures of the two baths, and not on any characteristics of the system Hamiltonian. When moving out of equilibrium, equalities become inequalities, and the possibility of finding laws that are independent of the system becomes not only unfeasible, but in principle impossible, given the huge range of different systems that these laws should encompass (it is important to notice that this is exactly the reason that makes fluctuation theorems such remarkable results).

A middle ground between the two extreme regimes described above is the one of slow driving transformations: in this context, despite the system is moved out of equilibrium by the driving, this happens so slowly compared to every thermalisation timescale, that effectively the state is almost thermal during the whole transformation. This approximation allows to examine systems out of equilibrium, while retaining a certain generality in the results obtained. In particular, a general feature emerging in this context is that the space of parameters naturally possesses a Riemannian structure, where the metric describes the way the system dissipates at a local level. In this context, while the metric is still system dependent, the results one can obtain from the existence of this structure are not.

In particular, we find that there are two main metric structures associated to every quantum system: one describing the average dissipation, the other describing its fluctuations. Interestingly, the two metric structures that emerge in this context are directly connected to the quantum Fisher information described in Chapter 1. While the first of these two metrics was already known for classical systems at least from the 70s, the second is a result only recently derived. The reason for this delay is that the two metrics actually coincide for any classical systems, and show a difference in behaviour only when coherences are present. Moreover, one can also prove that the production of coherences during a transformation is witnessed both from the breakdown of a relation connecting the average dissipation to its fluctuations (see Eq. (4.124)), and from the appearance of non-Gaussian features in the work probability distribution. Despite the particular form of this distribution is system dependent, we can give a fairly general characterisation of some of its main features: for example, we find that all the cumulants of the dissipation are positive, signalling a tendency of this quantity to substantially fluctuate over its average value (see the discussion in Sec. 4.5). Moreover, in Sec. 4.6 we

present general features of the way in which systems close to equilibrium dissipate, which subsume the simplifications observed in the characterisation of slowly driven transformations. Finally, in Sec. 4.7 we discuss the optimisation of engines in this regime, providing six general principles one can use to design engines operating at maximum power only by using quite high level information about the system and the baths.

As we discussed in Chapter 2, the definition of a probability distribution for work can be problematic in the quantum regime. While the results just discussed are valid for the average dissipation regardless of the measurement scheme employed, for all higher order statistics these depend on the use of the TPM definition. In particular, this means that no genuinely quantum phenomena can be observed, as there will always be the possibility of actually reproducing the statistics of the work via a non-contextual model. Indeed, since systems in the slow driving regime are always close to equilibrium, one could wonder whether the continuous thermalisation that they experience would result in a similar degradation of the coherence that is at the root of the non-contextuality of the TPM work distribution (as the initial measurement completely erases all the off-diagonal terms). For this reason, we include at the end of the Chapter the discussion of a system in the slow driving regime showing a purely contextual advantage in its power output (see Eq. (4.216) in Sec. 4.8). This proves that the restriction to transformations close to equilibrium does not prevent the observation of genuinely quantum phenomena.

Acknowledgements

It is usually common to refrain from explicitly recognising the value that people bring to our existence, as it is often the case that people behave the way they do without expecting anything in return, but just because of their inner nature. For this reason expressions of gratitude can feel like thanking a tree for its shadow, or the rain for easing the summer heat: something either too egocentric or deeply out of fashion. Still, in this occasion it is just necessary to face the risk of appearing *cursi* (that roughly translates to excessively sentimental, with the additional benefit of sounding exactly like its meaning), and wholeheartedly express my appreciation for the people who supported me during these years.

So, proceeding chronologically, I should start by thanking the friends from my home town, who are now scattered around Europe. Pierpaolo and Davide, your passion for your subjects is unrivalled; thanks for always reminding me why we do this job. Barbero, Herr Doktor, I consider you my *arbiter elegantiae*; I hope not to write about Hannah Arendt ever again, but in that case you will be the first to read it. I also want to thank the friends who were with me in Munich, and who made the time there special. Karan, you made me understand how deep a soul can be; thanks for always pushing me not to lose what is essential in life. I hope you will find what you are looking for. Angelo, I will always admire your stubbornness and patience, thanks for all you wrote. The Italian spirit forced me to leave and in Germany I understood what it means to be a stranger; however painful, this is an experience that is necessary in order to comprehend. In this respect, I am also very grateful to the people from places far away (too many to mention, but who would deserve better) for making me understand deeply that humanity is singular.

This last period of my life was so important because of the people that were part of it: Paolo, you are a moral compass and the best person to party with; thanks for showing me the quiet beauty of solving life little by

little; Arturo, you showed me what self discovery really means, thanks for continuing searching; I would love to be half as brave as you are; Dario, thanks for demonstrating how the support of a friend can make you achieve heights you would have never dreamt of, both in life and in climbing; mi raccomando, sbraga tutto; Jacopo, you have a special way of finding what makes true things really beautiful and to express them as they appear to you; I would hope to have your ability even just once to convey how grateful I am to be your friend: Barcelona is beautiful with you. Before coming here there was always something holding me back from experiencing and expressing what is truly important in life. Thanks to you guys, I understood better. We are not alone to this world, and now I know it.

The time in Barcelona was also special because I found a place I could call a home. Thanks Korbinian, you were the real initiator of this beautiful chapter, probably without your resolution I would be living in the streets. And thanks to the people who were part of the family of Plaça Revolució: Joan, thanks for everything you taught me, I've learned a lot from you; Ilaria, your freedom warms my heart; I am really grateful for all the good times together; Anthéa, from the first moment, it was a pleasure to talk with you; thanks for always suggesting the best places; Morghan, the parties we had were incredible; thanks for all the great memories; Jakub, you arrived last, but you rejoiced the house; thanks for all your energy.

Before finishing, I should also thank the people from the QIT group, in particular Cristian (you are one of the people I've seen growing the most during these years, I really admire you for that), Vicky and Mate (being around you makes me happy), together with Flavio, who welcomed me in the group, and Alexia and Joe, who host the best parties in town. For what regards the science, I would like to thank Martí, who introduced me to actual research and suggested me the best city in the world, and Toni, who always gave me enough freedom to pursue my personal interests.

Finally, I am very grateful to my family, to my parents, who taught me everything I know, and who I have always confided in, and to my sister, whose support has helped me a lot. I also want to dedicate one last word to my late grandfather, who was a model of what a good man should be and the constant reminder of what is important in life.

To conclude, thanks to Antonio Gaudí for Sagrada Familia, and Federico García Lorca por haber entrado en su jardín.

"Caminante no hay camino,
se hace camino al andar..."
Golpe a golpe, verso a verso.

Antonio Machado

Part I

Information

Chapter 1

Fisher information and the geometry of quantum states

Distinguishing classical or quantum probability distributions based on their observed statistics is a delicate matter: different operational criteria lead to different quantifiers, and there is no general consensus on which would be the most appropriate principle to single out one. The canonical family of tools in this context is given by the set of contrast functions, a generalisation of the relative entropy, which share with it many important properties, above all the contractivity under physical evolutions. This property mirrors the intuitive behaviour noisy transformations of loosing information about the initial conditions as the dynamics unfolds, leading to a progressive decrease in the ability of agents to distinguish different initial states. For classical probabilities, despite the plethora of different contrast functions, their local expansion leads to a unique quantity called the Fisher information metric, which endows the space of probability with a Riemannian structure. On the other hand, for quantum states the non-commutativity of different observables precludes the same uniqueness result, but one can anyway give a complete characterisation of the functional expression of general Fisher metrics.

Another way of characterising the same family of metrics is by requiring their contractivity under physical maps, result that goes under the name of Chentsov theorem for classical systems, and that was generalised by Petz to the quantum regime. Interestingly, it was recently shown that this relation goes both ways: all the physical evolutions are the ones and only the ones that contract the Fisher metrics. This result shows their inherently dynamical nature. As this family arises from distinguishabil-

ity measures and not from dynamical objects, this conclusion could in principle be surprising. For this reason, the connection is further corroborated by the fact that other dynamical properties of evolutions can be expressed in terms of Fisher information. On the one hand, this motivates the interest in this topic, on the other, it shows how physical evolutions could be defined just in terms of statistical quantities.

1.1 The classical case: from contrast functions to classical Fisher information

Suppose an agent has to evaluate which one between two models better describes an experiment. A possible strategy is the following: repeat the same observation independently many times, so to generate a list of data $\{x_i\}_{i\in\{1,\dots,N\}}$, and then compare the probability of getting this result given the two distributions. If the first model is associated to the probability distribution ρ and the second to σ, the ratio between the two probabilities can be rewritten as:

$$\frac{p(\{x_i\}_{i\in\{1,\dots,N\}}|\rho)}{p(\{x_i\}_{i\in\{1,\dots,N\}}|\sigma)} = \prod_{i=1}^{N} \frac{p(x_i|\rho)}{p(x_i|\sigma)} = \tag{1.1}$$

$$= \exp N \left(\frac{1}{N} \sum_{i=1}^{N} \left(\log p(x_i|\rho) - \log p(x_i|\sigma)\right) \right) , \tag{1.2}$$

where in the first line we used the independency between different runs to pass to the product. Suppose now that the true distribution is given by ρ. Thanks to the law of large numbers, the sum in the exponent will asymptotically converge to the value:

$$D_{\mathrm{KL}}(\rho||\sigma) := \mathrm{Tr}\left[\rho\left(\log\rho - \log\sigma\right)\right] , \tag{1.3}$$

where ρ and σ are diagonal density matrices encoding the corresponding classical distributions. This is just one of the many ways in which one can introduce the relative entropy in Eq. (1.3), and it directly connects it to the logarithm of observed statistics in independent observations.

The relative entropy has a series of useful properties:

1. *positivity:* that is $D_{KL}(\rho||\sigma) \geq 0$, with equality *iff* $\rho \equiv \sigma$;
2. *homogeneity:* corresponding to the condition $D_{KL}(\lambda\,\rho||\lambda\,\sigma) = \lambda\, D(\rho||\sigma)$, for positive λ;

3. *joint convexity:* which means that for every $0 \leq \lambda \leq 1$ it holds that

$$D_{KL}(\lambda\,\rho_1 + (1-\lambda)\,\rho_2||\lambda\,\sigma_1 + (1-\lambda)\,\sigma_2) \leq \\ \leq \lambda\, D_{KL}(\rho_1||\sigma_1) + (1-\lambda)\, D_{KL}(\rho_2||\sigma_2)\,;$$

4. *monotonicity:* meaning that for any stochastic map ϕ, $D(\phi(\rho)||\phi(\sigma)) \leq D(\rho||\sigma)$.

As it was mentioned in the introduction, different strategies give rise to different quantifiers, among which we mention two in particular, namely:

- the χ^2-divergence: given by $\chi^2(\rho||\sigma) = \mathrm{Tr}\left[\frac{(\rho-\sigma)^2}{\rho}\right]$, and whose operational interpretation will be discussed in the following, and,

- the Hellinger distance: defined as $H_{\mathrm{H}}(\rho||\sigma) := \frac{1}{2}\,\mathrm{Tr}\left[(\sqrt{\rho}-\sqrt{\sigma})^2\right]$.

All of these measures satisfy the four conditions listed above, so one can take them to give an axiomatic construction of possible distinguishability quantifiers. Out of the many functional forms that can be chosen to do so, the most commonly used ansatz is given by the family of g-divergences:

$$H_g(\rho||\sigma) := \mathrm{Tr}\left[\rho\, g(\sigma/\rho)\right]\,, \tag{1.4}$$

also called contrast functions or Csizár divergences (chapter 4 in [12]). All the examples above are contained in this family, as it can be verified by choosing $g(x) = -\log x$ to obtain the usual relative entropy, whereas the χ^2-divergence corresponds to $g(x) = (1-x)^2$ and the Hellinger distance to $g(x) = 1 - \sqrt{x}$. Whereas the homogeneity of Eq. (1.4) follows by construction, one needs to impose extra constraints on the function g for the other conditions to be satisfied: for example, the positivity corresponds to the request that $g(x) \geq 0$ for positive x and $g(1) = 0$. Moreover, $g(x)$ needs to be strictly convex at $x = 1$ for joint convexity and monotonicity to hold [12].

Remarkably, despite the generality of Eq. (1.4), all the g-divergences locally behave in the same way. This can be easily verified by considering an arbitrary diagonal state ρ and a small perturbation $\delta\rho$. Then, by

Taylor expanding the contrast functions one obtains:

$$H_g(\rho+\delta\rho\,||\rho) = \mathrm{Tr}\left[\rho\, g\left(\mathbb{1}+\frac{\delta\rho}{\rho}\right)\right] = \tag{1.5}$$

$$= \cancel{g(1)\mathrm{Tr}\,[\rho]} + \cancel{g'(1)\mathrm{Tr}\,[\delta\rho]} + \frac{g''(1)}{2}\mathrm{Tr}\left[\frac{\delta\rho^2}{\rho}\right] + \mathcal{O}\left(|\delta\rho|^3\right) = \tag{1.6}$$

$$= \frac{g''(1)}{2}\chi^2(\rho+\delta\rho\,||\rho) + \mathcal{O}\left(|\delta\rho|^3\right)\,, \tag{1.7}$$

where one uses the fact that, thanks to strict convexity, g has a minimum in $x = 1$, which implies that $g'(1) = 0$ and $g''(1) > 0$. This shows that locally all contrast functions coincide up to a constant factor with the χ^2-divergence. Equivalently, one can regard the χ^2-divergence as an extension of the local parabolic behaviour of contrast functions to the whole space of probability distributions.

We can also give an operational meaning to the χ^2-divergence, which naturally emerges in the context of estimating probability distributions from frequencies [13]. Assume the same setting as for Eq. (1.2), but now one has only access to the vector of frequencies $\{f_i\}_{i\in\{1,\ldots,n\}}$, where we indicate by n the number of different outputs and by N the number of repetitions of the experiment. If the true distribution is $\{p_i\}_{i\in\{1,\ldots,n\}}$, the probability of observing a given frequency vector is given by the multinomial distribution:

$$P(\{f_i\}_{i\in\{1,\ldots,n\}}) = \frac{N!}{(f_1N)!\ldots(f_nN)!}\, p_1^{f_1N}\ldots p_n^{f_nN}\,. \tag{1.8}$$

In the limit of many repetitions, one can use Stirling's approximation $N! \sim \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$, to obtain:

$$P(\{f_i\}_{i\in\{1,\ldots,n\}}) \overset{N\gg 1}{=} \frac{1}{\left(\sqrt{2\pi N}\right)^{n-1}\sqrt{f_1\ldots f_n}}\, e^{-ND_{\mathrm{KL}}(f||p)}\,. \tag{1.9}$$

Moreover, thanks to the law of large numbers, $\{f_i\}_{i\in\{1,\ldots,n\}}$ is exponentially close to the true distribution, so the Kullback-Leibler divergence reduces to $\chi^2(f||p)$. Suppose now that we want to characterise the probability distributions compatible with a given observed frequency vector $\{f_i\}_{i\in\{1,\ldots,n\}}$. Then, it follows from Eq. (1.9) that any probability distribution $p := f + \mathrm{d}f$ satisfying:

$$\chi^2(f||p) = \chi^2(f||f+\mathrm{d}f) = \sum_{i=1}^{n}\frac{\mathrm{d}f^i\mathrm{d}f^i}{f_i} \leq \frac{1}{N} \tag{1.10}$$

will be a good guess. This shows how the characterisation of the local geometry induced by the χ^2-divergence can be of interest for estimation theory.

Hence, to this end, we consider the metric defined at the point p by the tensor $g_{i,j} = \frac{\delta_{i,j}}{p_i}$. Clearly, $\chi^2(p||p + \mathrm{d}p) = \mathrm{d}p^i \mathrm{d}p^j \, g_{i,j}$. Interestingly, it is possible to make this metric Euclidean by an appropriate change of coordinates. Consider, indeed, the substitution given by $X_i := \sqrt{p_i}$ (called the purification of p); then the metric is given by:

$$g_{i,j} \, \mathrm{d}p^i \mathrm{d}p^j = g_{i,j} (4 \, X_i X_j) \, \mathrm{d}X^i \mathrm{d}X^j = \tag{1.11}$$

$$= \frac{\delta_{i,j}}{X_i^2} (4 \, X_i X_j) \, \mathrm{d}X^i \mathrm{d}X^j = 4 \delta_{i,j} \, \mathrm{d}X^i \mathrm{d}X^j = \tilde{g}_{i,j} \, \mathrm{d}X^i \mathrm{d}X^j \, . \tag{1.12}$$

The change of variables made the metric proportional to the Euclidean one. This mapping also provides a geometrical interpretation for the Hellinger distance. Given two distributions ρ and σ, corresponding to the two purifications $X := \sqrt{\rho}$ and $Y := \sqrt{\sigma}$, one can see that the Euclidean distance in the latter space is proportional to the Hellinger distance:

$$D_{\mathrm{Euc}}(X, Y)^2 = 4 \sum_{i=1}^{n} \, (X_i - Y_i)^2 = \tag{1.13}$$

$$= 4 \sum_{i=1}^{n} \, (\sqrt{\rho_i} - \sqrt{\sigma_i})^2 = 4 \, H_{\mathrm{H}}(\rho||\sigma) \, . \tag{1.14}$$

Despite this appealing connection, simply using the Euclidean distance in the space of purification is not sufficient to define Riemannian neighbourhoods on the space of probability distributions. In fact, in order for the normalisation constraint to hold, one has to require that:

$$\sum_{i=1}^{n} \, p_i = \sum_{i=1}^{n} \, X_i^2 = 1 \, . \tag{1.15}$$

This restrict the purifications associated to an actual probability distribution to an octant of the sphere (as the additional constraint $X_i \geq 0$ comes from the choice of the root for $\sqrt{p}$). Since spheres do not contain any straight line, one needs to use the metric induced by embedding them in a Euclidean space. In this case, geodesics are great circles, and the corresponding distance is measured by the angle in radians between the endpoints of the path (since for a radius 1 circle, the angle in radians

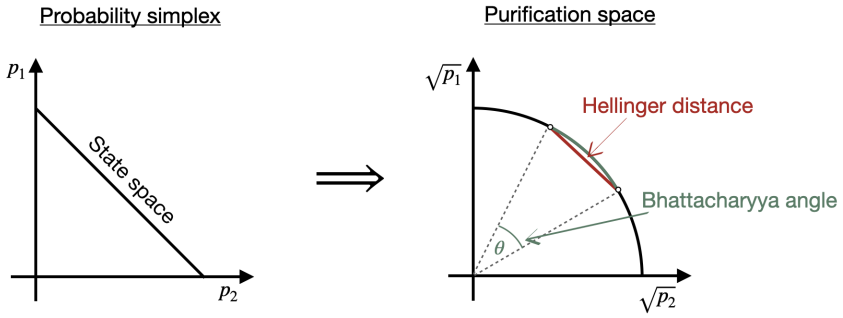


Figure 1.1: The space of classical probability distributions is the simplex defined by the conditions $\sum\ p_i = 1$ and $p_i \geq 0$. The square root maps the simplex into an octant of the sphere, the space of purifications. In the latter space, there is a straightforward geometrical interpretation for the Hellinger distance: it corresponds to the Euclidean distance measured in the purification space. This is not a geodesic distance, as the points crossed by the connecting chord lay outside the allowed states. The Bhattacharyya angle, or distance, on the other hand corresponds to the geodesic distance for the Fisher information metric.

corresponds to the length of the subtended arc). This construction takes the name of Bhattacharyya distance and can be expressed in formulae as:

$$D_{\mathrm{BH}}(X,Y) = 2\arccos\left(X\cdot Y\right) = 2\arccos\left(\sum_{i=1}^{n}\sqrt{\rho_i\sigma_i}\right). \qquad (1.16)$$

Notice that the additional factor 2 comes from the fact that the Fisher metric is just proportional to the Euclidean one.

The discussion above shows the importance of the χ^2-divergence: it generalises the quadratic local behaviour of all g-divergences to all states, which means that all contrast functions locally reduce to it. For this reason, the Riemannian metric locally induced by the χ^2-divergence coincides with the one derived from any other g-divergence. This metric takes the name of Fisher information metric, and its importance is further strengthened by the following result:

Theorem 1 (Chentsov [14])**.** *The Fisher information is the unique Riemannian metric on the space of smooth classical probability distributions which is contractive under arbitrary stochastic maps.*

A proof of this theorem can be found in [15], and its significance is further explored in Sec. 1.5.

In order to get some insights about the expression of the Fisher metric, it is useful to look at at a paradigmatic example. Consider the family of Gaussian probability distributions parametrised by μ and σ^2:

$$\mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,. \tag{1.17}$$

The Fisher information can be computed from the expansion:

$$\chi^2(\mathcal{N}_{\mu+\mathrm{d}\mu,\sigma+\mathrm{d}\sigma}||\mathcal{N}_{\mu,\sigma}) = \int_{-\infty}^{\infty} \mathrm{d}x\ \frac{(\mathcal{N}_{\mu+\mathrm{d}\mu,\sigma+\mathrm{d}\sigma}(x) - \mathcal{N}_{\mu,\sigma}(x))^2}{\mathcal{N}_{\mu,\sigma}(x)} = \tag{1.18}$$

$$\simeq \int_{-\infty}^{\infty} \mathrm{d}x\ \frac{(\partial_\mu \mathcal{N}_{\mu,\sigma}(x)\mathrm{d}\mu + \partial_\sigma \mathcal{N}_{\mu,\sigma}(x)\mathrm{d}\sigma)^2}{\mathcal{N}_{\mu,\sigma}(x)} = \tag{1.19}$$

$$= \int_{-\infty}^{\infty} \mathrm{d}x\ \left(\partial_\mu \log \mathcal{N}_{\mu,\sigma}(x)\, \mathrm{d}\mu + \partial_\sigma \log \mathcal{N}_{\mu,\sigma}(x)\, \mathrm{d}\sigma\right)^2 \mathcal{N}_{\mu,\sigma}(x)\,. \tag{1.20}$$

The last step, i.e., passing from a derivative on the distribution to a derivative on its logarithm (called log-likelihood), is quite standard, and it is particularly useful for exponential families. There are three integrals that need to be computed:

$$\int_{-\infty}^{\infty} \mathrm{d}x\ \left(\partial_\mu \log \mathcal{N}_{\mu,\sigma}\right)^2 \mathcal{N}_{\mu,\sigma}(x) = \int_{-\infty}^{\infty} \mathrm{d}x\ \left(\frac{x-\mu}{\sigma^2}\right)^2 \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sigma^2}\,; \tag{1.21}$$

$$\int_{-\infty}^{\infty} \mathrm{d}x\ \left(\partial_\sigma \log \mathcal{N}_{\mu,\sigma}(x)\right)^2 \mathcal{N}_{\mu,\sigma}(x) = \\ = \int_{-\infty}^{\infty} \mathrm{d}x\ \left(\frac{(x-\mu)^2 - \sigma^2}{\sigma^3}\right)^2 \mathcal{N}_{\mu,\sigma}(x) = \frac{2}{\sigma^2}\,; \tag{1.22}$$

$$\int_{-\infty}^{\infty} \mathrm{d}x\ \left(\partial_\mu \log \mathcal{N}_{\mu,\sigma}(x)\, \partial_\sigma \log \mathcal{N}_{\mu,\sigma}(x)\right) \mathcal{N}_{\mu,\sigma}(x) = 0\,. \tag{1.23}$$

Hence the Fisher information metric is given by:

$$g_{i,j}\mathrm{d}p^i \mathrm{d}p^j = \frac{1}{\sigma^2}(\mathrm{d}\mu^2 + 2\mathrm{d}\sigma^2)\,. \tag{1.24}$$

This is the metric of the Poincaré half-plane, a well studied model of hyperbolic geometry. In this case the geodesics are either vertical lines in the (μ, σ) plane, or half-circles whose origin lays on the μ axis (see Fig. 1.2). The interpretation of the metric is quite clear: in the case $\sigma \to 0$ there is

a deterministic output corresponding to μ, hence distributions with different expectation values are infinitely far away from each other, because they can be perfectly distinguished; in the opposite limit, as $\sigma \to \infty$, it is impossible to distinguish different distributions, as the average value is completely hidden by the fluctuations, and, for this reason, the metric trivialises to zero. Moreover, the half-circle geodesics can be interpreted as follows: consider the task of changing a probability distribution from $\mathcal{N}_{\mu_1,0}$ to $\mathcal{N}_{\mu_2,0}$ in such a way that locally (i.e., for finite statistics) the difference between subsequent samples is minimal. At the beginning, the only possibility is to increase the fluctuations, because any deviations from the average value μ_1 would witness a change in the probability distribution; once the fluctuations are strong enough, they start hiding the change in the expectation value, which can be modified more and more as the fluctuations increase; finally, the same process is reversed to reach $\mathcal{N}_{\mu_2,0}$. This example also shows a possible interpretation of the Fisher geodesics: they provide the trajectory in the probability space that adiabatically changes one distribution into another, where adiabatically is intended in the sense of minimal local change throughout the trajectory.

1.2 Quantum contrast functions

Any sensible generalisation of the concept of contrast functions to quantum states should satisfy the same properties that we found in the classical case. We rewrite them here for convenience:

1. *positivity:* $H(\rho||\sigma) \geq 0$, with equality *iff* $\rho \equiv \sigma$;
2. *homogeneity:* $H(\lambda\rho||\lambda\sigma) = \lambda\, H(\rho||\sigma)$, for $\lambda > 0$;
3. *joint convexity:* namely the condition that

$$H(\lambda\rho_1 + (1-\lambda)\rho_2||\lambda\sigma_1 + (1-\lambda)\sigma_2) \leq \\ \leq \lambda H(\rho_1||\sigma_1) + (1-\lambda)H(\rho_2||\sigma_2) \qquad (1.25)$$

   for $0 \leq \lambda \leq 1$;
4. *monotonicity:* for any Completely Positive Trace Preserving (CPTP) map Φ, it should hold that $H(\Phi(\rho)||\Phi(\sigma)) \leq H(\rho||\sigma)$;
5. *differentiability:* the function $h_{\rho,\sigma}(x,y) := H(\rho + xA||\sigma + yB)$ for A and B Hermitian operators is C^∞.

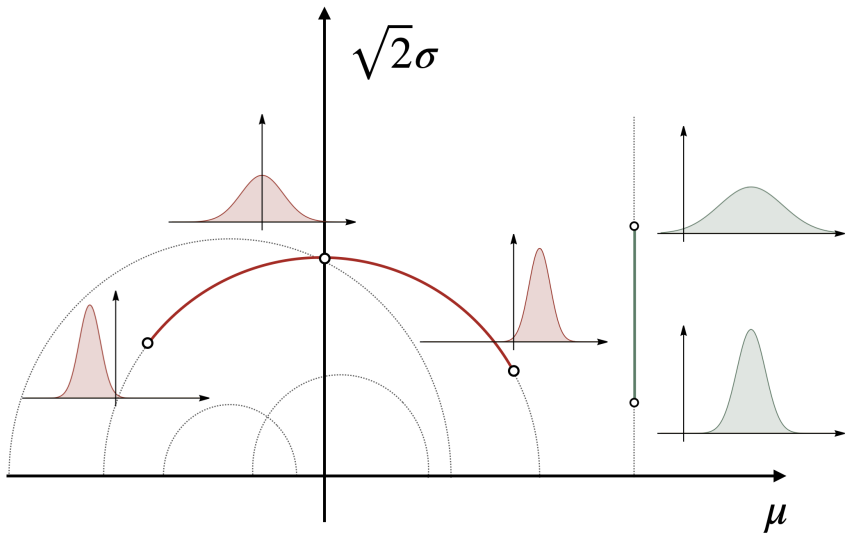


Figure 1.2: The Fisher information geometry associated to Gaussian distributions is the Poincaré half-plane, a well known model of hyperbolic geometry. Geodesics are given by half-circles with their origin on the x-axis (with vertical lines in the limit of infinite radius circles). Each point in the half-plane corresponds to a different Gaussian, and geodesics should be interpreted as a trajectory in which the statistics of the model are minimally varied. In the figure above we plot the Gaussians corresponding to the highlighted points, which are chosen to exemplify this behaviour.

Notice the addition of the last requirement, as we already expect that the expansion of the contrast functions will give rise to quantum Fisher informations, so it is sensible to already require some smoothness properties.

The non-commutativity of general quantum observables gives a certain freedom in the choice of contrast functions, corresponding to different ways in which one could define the division in the argument of $g(x)$ in Eq. (1.4). The most canonical choice is the following, proposed in [16]:

$$H_g(\rho||\sigma) := \mathrm{Tr}\left[\rho^{1/2}\, g\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big)\left[\rho^{1/2}\right]\right], \tag{1.26}$$

where $\mathbb{L}_\rho$ and $\mathbb{R}_\rho$ are the left and right multiplication superoperators, acting as $\mathbb{L}_\rho[\sigma] = \rho\,\sigma$ and $\mathbb{R}_\rho[\sigma] = \sigma\,\rho$, while g is an operator convex function defined on $(0,\infty)$, satisfying $g(1) = 0$. This choice is mainly justified by the content of Thm. 2, and by the use of the superoperator $\mathbb{L}_\sigma\mathbb{R}_\rho^{-1}$ in the context of W*-algebras, where it takes the name of relative modular operator [16]. It should be noticed that if one chooses $g(x) = -\log x$, then Eq. (1.26) gives the usual relative entropy.

We can now explore which constraints have to be imposed on $g(x)$ to satisfy the requirements above. The positivity condition 1 corresponds to imposing that $g(1) = 0$ is the only zero of the function. Moreover, by noticing the fact that the contrast function associated to $g(x) = (x-1)$ is identically zero (as $H_g(\rho||\sigma) = \mathrm{Tr}\,[\sigma - \rho] \equiv 0$) in order to ensure the positivity of $H_g(\rho||\sigma)$ it is sufficient to require that $g(x) + a(x-1) \geq 0$, for some arbitrary constant a. On the other hand, condition 2 is automatically satisfied, while one needs to require that $g(x)$ is matrix convex at $x = 1$ for the joint convexity to hold (condition 3). This directly implies the monotonicity of the contrast functions, condition 4, fact that can be proved as follows: first, it should be noticed that $H_g(\rho||\sigma)$ are unitary invariant. Indeed, one has:

$$H_g(U\,\rho\,U^\dagger||U\,\sigma\,U^\dagger) := \tag{1.27}$$

$$= \mathrm{Tr}\left[U\,\rho^{1/2}\,U^\dagger\, g\big(\mathbb{L}_{U\,\sigma\,U^\dagger}\mathbb{R}_{U\,\rho^{-1}\,U^\dagger}\big)\left[U\,\rho^{1/2}\,U^\dagger\right]\right] = \tag{1.28}$$

$$= \mathrm{Tr}\left[U\,\rho^{1/2}\,U^\dagger\,U\, g\big(\mathbb{L}_\sigma\mathbb{R}^{-1}_{\rho^{-1}}\big)\left[\rho^{1/2}\right]\,U^\dagger\right] = H_g(\rho||\sigma)\,, \tag{1.29}$$

where the step from Eq. (1.28) to the Eq. (1.29) can be verified either in coordinates (see Eq. (1.40)), or by expanding $g(x)$ in Laurent series, while in Eq. (1.29) we exploited the unitarity of U and the cyclicity of

the trace. Second, also notice that for generic contrast functions it holds that:

$$H_g(\rho\otimes\tau||\sigma\otimes\tau) = \tag{1.30}$$
$$= \mathrm{Tr}\left[g\big(\mathbb{L}_{\sigma\otimes\tau}\mathbb{R}^{-1}_{\rho\otimes\tau}\big)\left[\rho\otimes\tau\right]\right] = \mathrm{Tr}\left[g\big(\mathbb{L}_{\sigma}\mathbb{R}^{-1}_{\rho}\otimes\mathbb{I}\big)\left[\rho\otimes\tau\right]\right] = \tag{1.31}$$
$$= \mathrm{Tr}\left[g\big(\mathbb{L}_{\sigma}\mathbb{R}^{-1}_{\rho}\big)\left[\rho\right]\right]\mathrm{Tr}\left[\tau\right] = H_g(\rho||\sigma)\,, \tag{1.32}$$

where in Eq. (1.31) we used the fact that $\mathbb{L}_\tau\mathbb{R}^{-1}_\tau$ coincides with the identity operator on the commutant of τ. This identity can also be verified by expanding $g(x)$ in Laurent series or from the coordinate expression in Eq. (1.40). These two facts allow to deduce the monotonicity of the contrast functions in Eq. (1.26) from their joint convexity. Indeed, given a CPTP map Φ, one can express it in terms of its Stinespring's dilation:

$$\Phi(\rho) = \mathrm{Tr}_E\left[U\left(\rho\otimes|\psi\rangle\langle\psi|\right)U^\dagger\right]\,, \tag{1.33}$$

where U is a unitary operator and $|\psi\rangle\langle\psi|$ is an environmental pure state of dimension d_E. Take a unitary basis $\{V_i\}$ for the space of bounded operators of dimension d_E^2. It is a well-known result that $\sum_i(V_i(\rho_E)V_i^\dagger)/d_E^2 = \mathbb{1}_E/d_E$ for any ρ_E [17]. We denote this operator by $\Delta_1(\rho)$. This identity, together with Eq. (1.33) allows us to rewrite the action of the channel as:

$$\Phi(\rho)\otimes\frac{\mathbb{1}_{d_E}}{d_E} = \frac{1}{d_E^2}\sum_{i=1}^{d_E^2}(\mathbb{I}\otimes V_i)U\left(\rho\otimes|\psi\rangle\langle\psi|\right)U^\dagger(\mathbb{I}\otimes V_i^\dagger)\,, \tag{1.34}$$

where we used $\mathbb{1}$ to indicate the identity matrix, and $\mathbb{I}$ for the identity superoperator. Putting together this expression with Eq. (1.32) we can finally prove monotonicity. Indeed, one has:

$$H_g(\Phi(\rho)||\Phi(\sigma)) = H_g\left(\Phi(\rho)\otimes\frac{\mathbb{1}_{d_E}}{d_E}\middle|\middle|\Phi(\sigma)\otimes\frac{\mathbb{1}_{d_E}}{d_E}\right) = \tag{1.35}$$
$$= H_g\left((\mathbb{1}\otimes\Delta_1)\left(U\left(\rho\otimes|\psi\rangle\langle\psi|\right)U^\dagger\right)||(\mathbb{1}\otimes\Delta_1)\left(U\left(\sigma\otimes|\psi\rangle\langle\psi|\right)U^\dagger\right)\right) \leq \tag{1.36}$$
$$\leq \frac{1}{d_E^2}\sum_{i=1}^{d_E^2}H_g(\rho\otimes|\psi\rangle\langle\psi|\,||\sigma\otimes|\psi\rangle\langle\psi|) = H_g(\rho||\sigma)\,, \tag{1.37}$$

where in the first line and in the third line we used Eq. (1.32), while in the step from the second to the third line we used the decomposition in

Eq. (1.34) together with the unitary invariance and joint convexity of H_g. This proves condition 4.

In order to highlight the similarities and the differences with the classical contrast functions in Eq. (1.4), we provide here the coordinate expression of Eq. (1.26). First, define the eigensystem of ρ and σ as:

$$\rho = \sum_i \rho_i \, |\rho_i\rangle\langle\rho_i| \, , \qquad\qquad \sigma = \sum_j \sigma_j \, |\sigma_j\rangle\langle\sigma_j| \, . \tag{1.38}$$

Since $\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}[|\sigma_j\rangle\langle\rho_i|] = \frac{\sigma_j}{\rho_i} \, |\sigma_j\rangle\langle\rho_i|$, the relative modular operator is diagonal in the basis given by $\{|\sigma_j\rangle\langle\rho_i|\}$. Hence, Eq. (1.26) can be expanded as:

$$H_g(\rho||\sigma) := \sum_{i,j} \, \mathrm{Tr}\left[g\left(\frac{\sigma_j}{\rho_i}\right) |\sigma_j\rangle\langle\sigma_j| \, \rho \, |\rho_i\rangle\langle\rho_i| \right] = \tag{1.39}$$

$$= \sum_{i,j} \rho_i \, g\left(\frac{\sigma_j}{\rho_i}\right) |\langle\sigma_j|\rho_i\rangle|^2 \, . \tag{1.40}$$

In particular, it is clear that if ρ and σ are diagonal in the same basis, then Eq. (1.26) reduces to the classical expression in Eq. (1.4).

The operator convex functions with $g(1) = 0$ have the following integral expression

$$g(x) = a_g \, (x-1) + b_g \, (x-1)^2 + c_g \, \frac{(x-1)^2}{x} + \int_0^\infty \mathrm{d}\nu_g(s) \, \frac{(x-1)^2}{x+s} \, , \tag{1.41}$$

where $b_g, c_g > 0$, and ν_g is a positive measure with finite mass [16]. Moreover, using the two identities:

$$(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1} - \mathbb{1}) \left[\rho^{1/2}\right] = (\sigma - \rho)\rho^{-1/2} = \mathbb{R}_\rho^{-1/2}(\sigma - \rho) \, , \tag{1.42}$$

$$(\mathbb{R}_\sigma \mathbb{L}_\rho^{-1} - \mathbb{1}) \left[\rho^{1/2}\right] = \rho^{-1/2}(\sigma - \rho) = \mathbb{L}_\rho^{-1/2}(\sigma - \rho) \, , \tag{1.43}$$

one can rewrite the contrast functions in Eq. (1.26) as:

$$H_g(\rho||\sigma) = \mathrm{Tr}\left[\rho^{1/2} g\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big) \left[\rho^{1/2}\right]\right] = \tag{1.44}$$

$$= \mathrm{Tr}\left[\mathbb{L}_\rho^{-1/2}(\rho - \sigma) \; g\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big)(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1} - \mathbb{1})^{-2}[\mathbb{R}_\rho^{-1/2}(\rho - \sigma)]\right] = \tag{1.45}$$

$$= \mathrm{Tr}\left[(\rho - \sigma) \, \mathbb{R}_\rho^{-1} h\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big)[(\rho - \sigma)]\right] , \tag{1.46}$$

where in Eq. (1.45) we rewrote $\rho^{1/2}$ using the identities above, and we used the fact that $\mathbb{L}_\sigma^\dagger = \mathbb{R}_\sigma$ to group together the superoperators $(\mathbb{R}_\sigma \mathbb{L}_\rho^{-1} - \mathbb{1})^{-1}$ and $(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1} - \mathbb{1})^{-1}$. Finally, in Eq. (1.46) we defined the function $h(x) := g(x)/(x-1)^2$. Putting together the expression in Eq. (1.41) for $g(x)$ with what we just obtained in Eq. (1.46), it is a matter of simple algebra to give the general integral form of quantum contrast functions:

$$\begin{aligned} H_g(\rho||\sigma) =& \mathrm{Tr}\left[(\rho-\sigma)(b_g\,\rho^{-1} + c_g\,\sigma^{-1})(\rho-\sigma)\right] + \\ &+ \int_0^\infty \mathrm{d}\nu_g(s)\ \mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma + s\mathbb{R}_\rho)^{-1}[(\rho-\sigma)]\right] . \end{aligned} \tag{1.47}$$

As it was discussed above, the linear contribution disappears from the expression, meaning that, for some arbitrary a_g, just $g(x) + a_g(x-1)$ needs to be positive for the contrast function to satisfy the condition 1.

In general, the divergence in Eq. (1.26) is not symmetric. One might then wonder what is the connection between $H_g(\rho||\sigma)$ and $H_g(\sigma||\rho)$. This is given by the transformation $g(x) \to \hat{g}(x) := x\,g(x^{-1})$. In fact, it is straightforward to verify that:

$$H_{\hat{g}}(\rho||\sigma) = \mathrm{Tr}\left[\rho^{1/2}\,\hat{g}\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big)[\rho^{1/2}]\right] = \mathrm{Tr}\left[\hat{g}\big(\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}\big)[\rho]\right] = \tag{1.48}$$

$$= \mathrm{Tr}\left[g\big((\mathbb{L}_\sigma \mathbb{R}_\rho^{-1})^{-1}\big)[\mathbb{L}_\sigma \mathbb{R}_\rho^{-1}[\rho]]\right] = \mathrm{Tr}\left[g\big(\mathbb{L}_\sigma^{-1}\mathbb{R}_\rho\big)[\sigma]\right] = \tag{1.49}$$

$$= \mathrm{Tr}\left[g\big(\mathbb{L}_\rho \mathbb{R}_\sigma^{-1}\big)[\sigma]\right] = H_g(\sigma||\rho)\,, \tag{1.50}$$

where in the first line we used the cyclicity of the trace to group together the two $\rho^{1/2}$, while in the last line we used the commutation between $\mathbb{L}_\sigma$ and $\mathbb{R}_\rho$, together with the fact that $\mathbb{L}_\sigma[\sigma] = \mathbb{R}_\sigma[\sigma]$. Interestingly, the transformation above is also involutive (meaning that $\hat{g}(x) \to g(x)$), so it really corresponds to an exchange in the arguments of $H_g(\rho||\sigma)$. Hence, a divergence is symmetric if and only if its defining function satisfies $g(x) = \hat{g}(x) = xg(x^{-1})$.

We can then study which constraints symmetry imposes on the functions in Eq. (1.41). First, notice that under the transformation $g(x) \to \hat{g}(x)$ the function $(x-1)^2$ is mapped to $(x-1)^2/x$, so for symmetry to hold one has to require $b_g \equiv c_g$. If we now focus on the measure ν alone, we see that:

$$\hat{g}(x) = \int_0^\infty \mathrm{d}\nu_g(s)\ x\,\frac{(x^{-1}-1)^2}{x^{-1}+s} = \int_0^\infty \mathrm{d}\nu_g(s)\ \frac{(x-1)^2}{1+x\,s} = \tag{1.51}$$

$$= \int_0^\infty \mathrm{d}\nu_g(t^{-1})\ \frac{1}{t^2}\frac{(x-1)^2}{1+x\,t^{-1}} = \int_0^\infty \mathrm{d}\nu_g(t^{-1})\ \frac{1}{t}\frac{(x-1)^2}{t+x}\,, \tag{1.52}$$

where in the second line we used the substitution $t = s^{-1}$. In order for the last integral to coincide with $g(x)$ one then has to require that $\mathrm{d}\nu_g(s) = \mathrm{d}\nu_g(s^{-1})/s$. Hence, any $g(x)$ corresponding to a symmetric divergence can be written as:

$$g(x) = b_g \frac{(x+1)(x-1)^2}{x} + \int_0^1 \mathrm{d}\nu_g(s)\ \frac{(x-1)^2}{x+s} + \int_1^\infty \mathrm{d}\nu_g(s)\ \frac{(x-1)^2}{x+s} = \tag{1.53}$$

$$= b_g \frac{(x+1)(x-1)^2}{x} + \int_0^1 \mathrm{d}\nu_g(s)\ \left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right) = \tag{1.54}$$

$$= \int_0^1 \mathrm{d}N_g(s)\ \left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right), \tag{1.55}$$

where we implicitly transformed the second integral in the first line through the change of variables $s \to s^{-1}$, and used the properties of $\mathrm{d}\nu$. Finally, in the last line we grouped all the terms by defining the measure $\mathrm{d}N_g(s) := b_g\delta(s)\,\mathrm{d}s + \mathrm{d}\nu_g(s)$. Correspondingly, a generic symmetric contrast function takes the form:

$$H_g(\rho||\sigma) = \\ = \int_0^1 \mathrm{d}N_g(s)\ \mathrm{Tr}\left[(\rho-\sigma)\left((\mathbb{L}_\sigma + s\mathbb{R}_\rho)^{-1} + (\mathbb{L}_\rho + s\mathbb{R}_\sigma)^{-1}\right)[(\rho-\sigma)]\right]. \tag{1.56}$$

This expression will be particularly useful in the next section.

1.2.1 From contrast functions to quantum Fisher information

There are two main routes for extending the Fisher information metric to the quantum regime: one can either explore the local behaviour of contrast functions, or one try to extend Chentsov theorem to CPTP maps. In this section, we start from the first of the two approaches, because it shows the analogy and the differences with the classical case. Still, given the apparent arbitrariness in the definition of the contrast functions in Eq. (1.26), one could think that pursuing the second route could lead to a quite different characterisation: remarkably, though, one can single out the same family of metrics by imposing the property of being contractive

under arbitrary evolutions (Thm. 3). This is a first hint at the dynamical nature of the Fisher information, which will be further explored in Sec. 1.5.

In order to study the local behaviour of Eq. (1.26), consider two close-by states ρ and σ, that is $\rho := \pi + \varepsilon A$ and $\sigma := \pi + \varepsilon B$, where A and B are self-adjoint, traceless perturbations. By using the expression for contrast functions in Eq. (1.46) and by retaining only terms up to order $\mathcal{O}\left(\varepsilon^2\right)$, we can expand the divergence as:

$$H_g(\pi + \varepsilon A||\pi + \varepsilon B) = \tag{1.57}$$

$$= \varepsilon^2 \operatorname{Tr}\left[(A-B)\,\mathbb{R}^{-1}_{\pi+\varepsilon A} h\big(\mathbb{L}_{\pi+\varepsilon B}\mathbb{R}^{-1}_{\pi+\varepsilon A}\big)[(A-B)]\right] = \tag{1.58}$$

$$= \frac{\varepsilon^2}{2}\left(\operatorname{Tr}\left[(A-B)\,\left(\mathbb{R}^{-1}_{\pi+\varepsilon A} h\big(\mathbb{L}_{\pi+\varepsilon B}\mathbb{R}^{-1}_{\pi+\varepsilon A}\big)\right)[(A-B)]\right] + \operatorname{Tr}\left[A \leftrightarrow B\right]\right) = \tag{1.59}$$

$$= \frac{\varepsilon^2}{2}\operatorname{Tr}\left[(A-B)\,\mathbb{R}^{-1}_{\pi}\,\frac{1}{f}\big(\mathbb{L}_{\pi}\mathbb{R}^{-1}_{\pi}\big)\,[(A-B)]\right] + \mathcal{O}\left(\varepsilon^3\right)\,, \tag{1.60}$$

where Eq. (1.58) is exact, while in Eq. (1.59) we denote by $\operatorname{Tr}\left[A \leftrightarrow B\right]$ the first trace of the line but with A and B exchanged, and we used the fact that exchanging the arguments of $H_g(\rho||\sigma)$ does not affect the result at this order of approximation. The function defined in the last line is given by:

$$f(x) = \frac{(x-1)^2}{g(x) + x g(x^{-1})}\,. \tag{1.61}$$

We remind the reader that exchanging the arguments in the contrast function corresponds to the transformation $g(x) \to x g(x^{-1})$. In this way, we see that every $f(x)$ is in one to one relation with a unique symmetric contrast function. Interestingly, the properties of $f(x)$ are quite well characterised. First, since $g(x)$ has to be matrix convex, it follows that $f(x)$ is matrix monotone [18]. Moreover, it should be noticed that $f(x)$ satisfies the symmetry $f(x) = x f(x^{-1})$. Finally, without loss of generality, we can require the normalisation condition $f(1) = 1$. The functions satisfying these three conditions are called standard monotone functions. Interestingly, they can be completely classified using the following:

Lemma 1 (Theorem 4.43 from [18])**.** *Consider a function $f : (0,\infty) \to (0,\infty)$. The following conditions are equivalent:*

1. *$f(x)$ is matrix monotone;*

2. $[Tf](x) := x/f(x)$ *is matrix monotone;*

3. $f(x)$ *is matrix concave.*

This theorem directly implies that the standard monotones are all contained in the interval:

$$\frac{2x}{x+1} \leq f(x) \leq \frac{x+1}{2}\,. \tag{1.62}$$

This can be shown as follows: thanks to the condition $f(x) = xf(x^{-1})$ one only needs to characterise the properties of $f(x)$ in the interval $[0, 1]$. Moreover, the same condition also implies that $f'(1) = \frac{1}{2}$. In fact, this can be easily verified from the equation:

$$f'(x) = f(x^{-1}) - \frac{1}{x} f'(x^{-1})\,, \tag{1.63}$$

by setting x to 1, and noticing the normalisation $f(1) = 1$. Then, from concavity it follows that $f(x) \leq f(1)+f'(1)(x-1) = (x+1)/2$. The upper bound satisfies all the necessary constraints, so it can be identified as the largest standard monotone function, which we indicate by f_B. Finally, notice that the transformation $f \to Tf$ inverts the inequality and maps standard monotones into standard monotones.

Hence, in analogy to what happens in the classical case, when one considers infinitesimally close distributions, contrast functions can be expanded according to the theorem:

Theorem 2 (Lesniewski, Ruskai [16])**.** *For each g satisfying the required properties to define a contrast function, one can locally approximate H_g up to corrections of order $\mathcal{O}\left(\varepsilon^3\right)$ as:*

$$H_g(\pi + \varepsilon A||\pi + \varepsilon B) = \frac{\varepsilon^2}{2} \operatorname{Tr}\left[(A-B)\,\mathbb{J}_f^{-1}\big|_\pi[(A-B)]\right]\,, \tag{1.64}$$

where the superoperator $\mathbb{J}_f$ is given by:

$$\mathbb{J}_f\big|_\pi := \mathbb{R}_\pi\, f\big(\mathbb{L}_\pi \mathbb{R}_\pi^{-1}\big)\,, \tag{1.65}$$

and g and f are connected by the equation:

$$f(x) = \frac{(x-1)^2}{g(x) + x\, g(x^{-1})}\,. \tag{1.66}$$

In contrast with what happened in classical statistics, quantum divergences do not collapse on a single quantity when considering their local behaviour. Still, it is possible to claim that the class of operators $\mathbb{J}_f^{-1}$ generalises the Fisher information to the quantum regime by using the concept of monotone metrics. These are scalar products $K_\pi(A, B)$ that satisfy the two properties:

1. *smoothness:* $K_\pi(A, B)$ depends smoothly on π;
2. *monotonicity:* for every CPTP Φ the metric is contractive:

$$K_{\Phi(\pi)}(\Phi(A), \Phi(A)) \leq K_\pi(A, A)\,.$$

In the classical setting, Chentsov theorem identifies the Fisher information as the unique monotone metric on the space of probability distributions. For quantum states, on the other hand, we have:

Theorem 3 (Petz [19])**.** *The monotone metrics on self-adjoint operators are all given in the form:*

$$K_{f,\pi}(A, B) := \mathrm{Tr}\left[A\ \mathbb{J}_f^{-1}\big|_\pi[B]\right], \tag{1.67}$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function. Moreover, requiring that $K_{f,\pi}(A, B)$ is real and that it reduces to the classical Fisher information for commuting variables constrains f to be a standard monotone function.

This last result corroborates the interpretation of $\mathbb{J}_f^{-1}$ as the natural extension of Fisher information scalar product to quantum mechanical systems. Moreover, it also shows that the definition of contrast functions in Eq. (1.26) is well justified, as their local behaviour correctly reduces to the quantum Fisher information.

Due to the importance of the quantum Fisher information and the richness in the structure of the $\mathbb{J}_f$, in the next subsection we study in depth the properties of this class of operators.

1.3 Properties of quantum Fisher information

In the previous section it was shown that the expansion of the quantum contrast functions constrains f to be a standard monotone, that is to satisfy the symmetry $f(x) = xf(x^{-1})$, and additionally we imposed the

normalisation $f(x) = 1$. The same characterisation is obtained in the context of monotone metrics when one requires $K_{f,\pi}(A, B)$ to be real and to reduce to the classical Fisher information for commuting states. Metrics satisfying this last requirement are said to be Fisher adjusted, and this condition follows directly from requiring $f(1) = 1$. In fact, first notice that $[B, \pi] = 0$ implies that $\mathbb{L}_\pi \mathbb{R}_\pi^{-1} B = B$. Then, a simple calculation gives:

$$\mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi[B]\right] = \mathrm{Tr}\left[A\, \mathbb{R}_\pi^{-1} \frac{1}{f}(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})[B]\right] = \mathrm{Tr}\left[A\, \mathbb{R}_\pi^{-1} \frac{1}{f}(\mathbb{I})[B]\right] = \tag{1.68}$$

$$= \mathrm{Tr}\left[AB\pi^{-1}\right] , \tag{1.69}$$

which is exactly the same result one would obtain if A and B were classical vectors. Moreover, for the same reason, the action of $\mathbb{J}_f$ on operators that commute with π is simply the multiplication by the state.

On the other hand, the condition $f(x) = xf(x^{-1})$ enforces that $\mathbb{J}_f$ maps self-adjoint operators into self-adjoint operators. Indeed, one has that, if $A = A^\dagger$, it follows that:

$$(\mathbb{J}_f\big|_\pi[A])^\dagger = (\mathbb{R}_\pi f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})[A])^\dagger = \mathbb{L}_\pi f(\mathbb{R}_\pi \mathbb{L}_\pi^{-1})[A^\dagger] = \tag{1.70}$$

$$= \cancel{\mathbb{L}_\pi \mathbb{L}_\pi^{-1}} \mathbb{R}_\pi f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})[A] = \mathbb{J}_f\big|_\pi[A] , \tag{1.71}$$

where we used the fact that $(\mathbb{L}_\pi[A])^\dagger = \mathbb{R}_\pi[A^\dagger]$ (and similarly for $\mathbb{R}_\pi$), the commutation $[\mathbb{L}_\pi, \mathbb{R}_\pi] = 0$ and the property $f(x) = xf(x^{-1})$.

The monotonicity of $f(x)$ is expressed in formulae as:

$$\mathrm{Tr}\left[\Phi(A)\, \mathbb{J}_f^{-1}\big|_{\Phi(\pi)}[\Phi(A)]\right] \leq \mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi[A]\right] , \tag{1.72}$$

for every self-adjoint operator A and CPTP map Φ. By polarisation, one can transfer the condition on the superoperators themselves obtaining:

$$\Phi^\dagger \circ \left(\mathbb{J}_f^{-1}\big|_{\Phi(\pi)}\right) \circ \Phi \;\leq\; \left(\mathbb{J}_f^{-1}\big|_\pi\right) ; \tag{1.73}$$

$$\Phi \circ \left(\mathbb{J}_f\big|_\pi\right) \circ \Phi^\dagger \;\leq\; \left(\mathbb{J}_f\big|_{\Phi(\pi)}\right) , \tag{1.74}$$

where the first inequality directly follows from Eq. (1.72), while the second can be obtained with a two lines proof (Lemma 2 in [19]). Notice that the equality is obtained if and only if Φ is a unitary.

Interestingly, the class of standard monotones also comes with a partial ordering. We say that $f_1 \leq f_2$ if $f_1(x) \leq f_2(x)$ for every x. This partial order is reflected in the operators $\mathbb{J}_f$. In fact, it holds that:

$$f_1 \leq f_2 \quad \Longrightarrow \quad (\mathbb{J}_{f_1} \leq \mathbb{J}_{f_2}) \quad \wedge \quad \left(\mathbb{J}_{f_1}^{-1} \geq \mathbb{J}_{f_2}^{-1}\right) . \tag{1.75}$$

Since $\mathbb{J}_f\big|_\pi$ is diagonal in a basis independent of f (see Eq. (1.81)), this fact can be verified directly in coordinates. Hence, there exists a minimum and a maximum quantum Fisher information, respectively corresponding to $f(x) = (x+1)/2$ and $f = 2x/(x+1)$. Interestingly, the same ordering is reflected also on the symmetrised version of the contrast functions. In particular, one has that:

$$H_{\frac{(x-1)^2}{x+1}}(\rho||\sigma) \leq H_g^{\text{symm}}(\rho||\sigma) \leq H_{(x-1)^2}^{\text{symm}}(\rho||\sigma) , \tag{1.76}$$

where it should be noticed that the lower bound is already symmetrised thanks to the symmetry of the corresponding g.

We can now pass to give some explicit expressions of the quantum Fisher information. First, one can replicate the computations that led to Eq. (1.56) to obtain:

$$H_g(\pi + \varepsilon A||\pi + \varepsilon B) = \frac{\varepsilon^2}{2} \operatorname{Tr}\left[(A-B)\, \mathbb{J}_f^{-1}\big|_\pi[(A-B)]\right] = \tag{1.77}$$

$$= \frac{\varepsilon^2}{2} \int_0^1 \mathrm{d}N_g(s) \operatorname{Tr}\left[(A-B)\left((\mathbb{L}_\pi + s\mathbb{R}_\pi)^{-1} + (\mathbb{R}_\pi + s\mathbb{L}_\pi)^{-1}\right)[(A-B)]\right] , \tag{1.78}$$

where we extended the notation $\mathrm{d}N(s)$ to asymmetric $g(x)$ as $\mathrm{d}N(s) := (b_g + b_{\hat{g}} + c_g + c_{\hat{g}})\delta(s)\mathrm{d}s/2 + (\mathrm{d}\nu_g(s) + \mathrm{d}\nu_g(s^{-1})/s)$. From this expression, together with the requirement that $\mathbb{J}_f^{-1}$ should be Fisher adjusted, we obtain the normalisation condition on the measure $\mathrm{d}N(s)$:

$$\int_0^1 \mathrm{d}N_g(s)\, \frac{2}{1+s} = 1 , \tag{1.79}$$

as it can be verified by noticing that $(\mathbb{L}_\pi + s\mathbb{R}_\pi)^{-1}[A] = \frac{A\pi^{-1}}{(1+s)}$ for commuting operators.

We can also use the functional expression of $\mathbb{J}_f$ to give its coordinate in the basis of π, denoted by:

$$\pi = \sum \pi_i \, |i\rangle\langle i| \, . \tag{1.80}$$

It is straightforward to compute the action of $\mathbb{J}_f$ on $|i\rangle\langle j|$:

$$\mathbb{J}_f\big|_\pi [|i\rangle\langle j|] = f\left(\frac{\pi_i}{\pi_j}\right)\pi_j \;\; |i\rangle\langle j| \; ; \tag{1.81}$$

$$\mathbb{J}_f^{-1}\big|_\pi [|i\rangle\langle j|] = \left(f\left(\frac{\pi_i}{\pi_j}\right)\pi_j\right)^{-1} \;\; |i\rangle\langle j| \; . \tag{1.82}$$

Consider now an operator A which can be written in coordinates as $A := \sum A_{i,j}\, |i\rangle\langle j|$. Then, $\mathbb{J}_f$ acts as:

$$\mathbb{J}_f\big|_\pi[A] = \sum f\left(\frac{\pi_i}{\pi_j}\right)\pi_j \; A_{i,j} \; |i\rangle\langle j| = \sum (\mathcal{J}_f \circ A)_{i,j} \; |i\rangle\langle j| \; , \tag{1.83}$$

where we introduced the matrix $\mathcal{J}_f := \sum \mathbb{J}_f[|i\rangle\langle j|]$, and we use the circle to denote the Hadamard product. Analogously, the same computation can be carried out for $\mathbb{J}_f^{-1}$.

The expression just obtained is particularly useful to discuss the positivity properties of $\mathbb{J}_f$. In fact, a result that goes under the name of Schur product theorem states that the Hadamard product of two operators is positive if they are both positive. This means that in order for $\mathbb{J}_f$ to be positivity preserving, the matrix $\mathcal{J}_f$ has to be positive semidefinite for any probability vector $\{\pi_i\}_{i\in\{1,\dots,n\}}$ and any $n \in \mathbb{N}$ (and similarly for $\mathbb{J}_f^{-1}$). Since every principal sub-matrix should also be positive semidefinite, from the positivity of the determinant of the 2×2 matrix containing only the (i,j)-components, it follows that $\pi_i\pi_j - f(\pi_i/\pi_j)^2\pi_j^2 \geq 0$. Hence, one can deduce that a necessary condition for $\mathbb{J}_f$ to be positive preserving is that $f(x) \leq \sqrt{x}$, while for $\mathbb{J}_f^{-1}$ the condition becomes $f(x) \geq \sqrt{x}$.

Consider now the question about the complete positivity of $\mathbb{J}_f$. It is straightforward to see that P $\implies$ CP. In fact, complete positivity corresponds to the request that $\mathbb{J}_f\big|_\pi \otimes \mathbb{I}_n$ is positivity preserving for any n. But this can be written as $\mathbb{J}_f\big|_{\pi\otimes\mathbb{1}_n}$, which proves the claim.

Finally, it is interesting to explore the properties of the transformation $f(x) \to [Tf](x) := x/f(x)$. From Lemma 1 it follows that this map sends standard monotones into standard monotones. Using the property $f(x) = xf(x^{-1})$ we can rewrite this transformation as:

$$[Tf](x) = \frac{x}{f(x)} = \frac{1}{f(x^{-1})} \, . \tag{1.84}$$

From this expression it is evident that T is involutive, meaning that $TTf = f$. Moreover, it has a unique fixed point, given by $f(x) = \sqrt{x}$.

Since it flips inequalities, this also means that it maps the region $f(x) \geq \sqrt{x}$ to the region $f(x) \leq \sqrt{x}$, that is it maps the region in which $\mathbb{J}_f^{-1}$ can be CP to the region in which $\mathbb{J}_f$ can be CP. In particular, one can rewrite $\mathbb{J}_{Tf}$ as:

$$\mathbb{J}_{Tf}\big|_\pi = \mathbb{R}_\pi \, Tf(\mathbb{L}_\pi \mathbb{R}_\pi^{-1}) = \mathbb{R}_\pi \frac{1}{f}(\mathbb{L}_\pi^{-1}\mathbb{R}_\pi) = \tag{1.85}$$

$$= \mathbb{R}_{\pi^{-1}}^{-1} \frac{1}{f}(\mathbb{L}_{\pi^{-1}}\mathbb{R}_{\pi^{-1}}^{-1}) = \mathbb{J}_f^{-1}\big|_{\pi^{-1}} \, . \tag{1.86}$$

From this relation we can deduce that T maps $\mathbb{J}_f$ that are CP into $\mathbb{J}_{Tf}^{-1}$ that are CP (and vice versa). In fact, suppose that $\mathbb{J}_f$ is given in the general Kraus form:

$$\mathbb{J}_f\big|_\pi [\rho] = \sum_i K_i(\pi) \, \rho \, K_i(\pi)^\dagger \, , \tag{1.87}$$

where i ranges on a possibly uncountable set, and $K_i(\pi)$ are π-dependent Kraus operators. Then, $\mathbb{J}_{Tf}^{-1}$ can be expressed as:

$$\mathbb{J}_{Tf}^{-1}\big|_\pi [\rho] = \mathbb{J}_f\big|_{\pi^{-1}} [\rho] = \sum_i K_i(\pi^{-1}) \, \rho \, K_i(\pi^{-1})^\dagger \, , \tag{1.88}$$

so it also admits an expression in terms of Kraus operators, i.e., it is also completely positive.

This transformation is also useful to study the convexity properties of the set of standard monotones. Indeed, it was shown in [20] that the inverse of a standard monotone can be rewritten as:

$$\frac{1}{f(x)} = \int_0^1 \mathrm{d}\mu_f(\lambda) \, \left(\frac{\lambda+1}{2}\right)\left(\frac{1}{x+\lambda} + \frac{1}{1+\lambda\, x}\right) \, , \tag{1.89}$$

where $\mathrm{d}\mu_f(\lambda)$ is a probability measure. Then, following Lemma 1, we know that $[Tf](x) = x/f(x)$ is also a matrix monotone, and, since the transformation T is invertible and involutive, for every standard monotone f there corresponds a unique Tf and all the standard monotones can be written as the image of some other function, namely Tf, under T. Hence, we can manipulate Eq. (1.89) to give the following integral expression:

$$f(x) = \int_0^1 \mathrm{d}\mu_{Tf}(\lambda) \, \left(\frac{1+\lambda}{2}\right)\left(\frac{x}{x+\lambda} + \frac{x}{1+\lambda\, x}\right) \, . \tag{1.90}$$

This shows that the set of standard monotone functions has the property of being a Bauer simplex, and that its extreme points are exactly the functions in the integrand of Eq. (1.90).

Finally, it should be noticed that we can lift the transformation T to act on symmetric contrast functions. In particular, since there is a one-to-one relation between the standard monotone functions and the convex functions $g^{\text{symm}}(x)$ corresponding to symmetric divergences, the range of the latter is given by:

$$\frac{(x-1)^2}{(x+1)} \leq g^{\text{symm}}(x) \leq \frac{(x+1)(x-1)^2}{4x}\,. \tag{1.91}$$

We then define the transformation $\tilde{T}$ acting on $g^{\text{symm}}(x)$ as:

$$[\tilde{T}g^{\text{symm}}](x) := \frac{(x-1)^4}{4\,x\,g^{\text{symm}}(x)}\,. \tag{1.92}$$

The definition is chosen so that the following diagram commutes:

$$\begin{array}{ccc} g^{\text{symm}} & \xrightarrow{\tilde{T}} & \tilde{T}g^{\text{symm}} \\ {\scriptstyle\text{Locally}}\downarrow & & \downarrow{\scriptstyle\text{Locally}} \\ f & \xrightarrow{T} & Tf\,. \end{array}$$

Now that we have discussed the main properties of the quantum Fisher information operators, in the next section we present a compendium of notable examples.

1.4 A garden of quantum Fisher information

Before starting to consider the rich range of specific examples that quantum Fisher informations offer us, we study the general properties of the set of standard monotones, which we denote by $\mathcal{F}$. As we discussed above, this is a convex set, compact under pointwise convergence, and whose extreme points are given in Eq. (1.90), that is, functions of the form:

$$f_\lambda(x) = \left(\frac{1+\lambda}{2}\right)\left(\frac{x}{x+\lambda} + \frac{x}{1+\lambda\,x}\right)\,, \tag{1.93}$$

where $\lambda \in [0,1]$. Starting from the definition in Eq. (1.65), we can give the explicit expression:

$$\mathbb{J}_{f_\lambda}\big|_\pi[A] = \mathbb{R}_\pi\, f_\lambda\big(\mathbb{L}_\pi \mathbb{R}_\pi^{-1}\big)[A] = \tag{1.94}$$

$$= \left(\frac{1+\lambda}{2}\right)\left((\mathbb{L}_\pi\mathbb{R}_\pi^{-1}+\lambda)^{-1}[\pi A] + (1+\lambda\mathbb{L}_\pi\mathbb{R}_\pi^{-1})^{-1}[\pi A]\right) = \tag{1.95}$$

$$= \left(\frac{1+\lambda}{2}\right)\left((\mathbb{L}_\pi+\lambda\mathbb{R}_\pi)^{-1} + (\mathbb{R}_\pi+\lambda\mathbb{L}_\pi)^{-1}\right)[\pi A\pi]\,, \tag{1.96}$$

where in the step from the first to the second line we used the commutation between $\mathbb{L}_\pi$ and $\mathbb{R}_\pi$. This implies that generic $\mathbb{J}_f$ can be written as:

$$\mathbb{J}_f\big|_\pi[A] = \tag{1.97}$$

$$= \int_0^1 \mathrm{d}\mu_{Tf}(\lambda)\, \left(\frac{1+\lambda}{2}\right)\left((\mathbb{L}_\pi+\lambda\mathbb{R}_\pi)^{-1} + (\mathbb{R}_\pi+\lambda\mathbb{L}_\pi)^{-1}\right)[\pi A\pi]\,. \tag{1.98}$$

Thanks to Eq. (1.86), we can also express any Fisher information metric as:

$$\mathbb{J}_f^{-1}\big|_\pi[A] = \mathbb{J}_{Tf}\big|_{\pi^{-1}}[A] = \tag{1.99}$$

$$= \int_0^1 \mathrm{d}\mu_f(\lambda)\, \left(\frac{1+\lambda}{2}\right)\left((\mathbb{L}_\pi^{-1}+\lambda\mathbb{R}_\pi^{-1})^{-1} + (\mathbb{R}_\pi^{-1}+\lambda\mathbb{L}_\pi^{-1})^{-1}\right)[\pi^{-1}A\pi^{-1}] = \tag{1.100}$$

$$= \int_0^1 \mathrm{d}\mu_f(\lambda)\, \left(\frac{1+\lambda}{2}\right)\left((\mathbb{L}_\pi+\lambda\mathbb{R}_\pi)^{-1} + (\mathbb{R}_\pi+\lambda\mathbb{L}_\pi)^{-1}\right)[A]\,, \tag{1.101}$$

as it could also be directly inferred from Eq. (1.89). Interestingly, we are back to the expression in Eq. (1.78), so that we can identify the two defining measures through the formula:

$$\mathrm{d}N_g(\lambda) := \mathrm{d}\mu_f(\lambda)\, \left(\frac{1+\lambda}{2}\right)\,. \tag{1.102}$$

This shows that the expansion in Eq. (1.78) can be interpreted as a decomposition into extreme points of the space of standard monotones.

It is interesting to notice that for $\lambda \geq 3-2\sqrt{2}$, the extreme points satisfy $f_\lambda(x) \leq \sqrt{x}$, whereas for any $\lambda \in (0, 3-2\sqrt{2})$ the corresponding monotone has two additional crossing with the graph of the square root

(other than $x = 0$ and $x = 1$) implying that $f_\lambda(x)$ neither lays all beneath nor all above $\sqrt{x}$. This shows that the set of monotone functions is not totally ordered. Moreover, it also shows that there are fs for which both $\mathbb{J}_f$ and $\mathbb{J}_f^{-1}$ are not CP.

Indeed, regarding the question of which operators are completely positive, we have seen that a necessary condition is that the corresponding defining function is ordered with respect to the square root. In the next subsection we will provide a more in-depth analysis of sufficient and necessary conditions for complete positivity to hold.

1.4.1 Completely positive Fisher information functionals

Before providing more precise criteria for complete positivity, it is useful to introduce the set of inverse standard monotones, namely:

$$\mathcal{K} := \left\{ \frac{1}{f(x)} \,\middle|\, f \in \mathcal{F} \right\}, \tag{1.103}$$

or, equivalently, the set of all operator convex functions satisfying $k(x) = k(x^{-1})x^{-1}$ and such that $k(1) = 1$. Indeed, one can readily verify the equivalence between the two definitions as the inverse of a matrix monotone is matrix convex (and vice versa), and the two extra conditions are given by mapping $k(x) \to \frac{1}{k(x)}$. This transformation is a bijection between $\mathcal{F}$ and $\mathcal{K}$, and will be denoted by I in the following (i.e., $[If](x) := 1/f(x)$). Similarly to T, this map is also involutive, but whereas T sends $\mathcal{F}$ into itself, I maps it to $\mathcal{K}$.

We call the members of $\mathcal{K}$ standard convex functions. This family is also a convex, compact set, and one can deduce from Eq. (1.89) that the corresponding extreme points are given by the functions:

$$k_\lambda(x) := \left(\frac{\lambda+1}{2}\right)\left(\frac{1}{x+\lambda} + \frac{1}{1+\lambda\, x}\right). \tag{1.104}$$

These functions are decreasing in λ, so every standard convex satisfies:

$$k_1(x) = \frac{2}{x+1} \leq k(x) \leq \frac{x+1}{2x} = k_0(x)\,. \tag{1.105}$$

It should be also noticed that $k_0 = If_1$ and $k_1 = If_0$.

The isomorphism between $\mathcal{F}$ and $\mathcal{K}$ makes the choice of which one to use in order to define the quantum Fisher information largely a matter of

$g(x)$	$H_g(\rho\,\|\,\|\,\sigma)$	$\mathrm{d}N(s)$	$f(x)$	$\mathbb{J}_f[A]$	$\mathbb{J}_f^{-1}[A]$
$\frac{(x-1)^2}{x+1}$	$\mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma+\mathbb{R}_\rho)^{-1}(\rho-\sigma)\right]$	$\frac{\delta(1-s)}{2}$	$\frac{x+1}{2}$	$\frac{1}{2}\{\pi, A\}$	$\int_0^\infty \mathrm{d}t\, e^{-t\pi/2}\, A\, e^{-t\pi/2}$
$\frac{(x-1)^2}{x^\gamma+x^{1-\gamma}}$	$\mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma^\gamma\mathbb{R}_\rho^{1-\gamma}+\mathbb{L}_\sigma^{1-\gamma}\mathbb{R}_\rho^{\gamma})^{-1}[(\rho-\sigma)]\right]$	???	$\frac{x^\gamma+x^{1-\gamma}}{2}$	$\frac{1}{2}\{\pi^{1-2\gamma}, \pi^\gamma A\,\pi^\gamma\}$	$\int_0^\infty \mathrm{d}t\, \mathrm{e}^{-(t\pi^{1-2\gamma})/2}\,\pi^{-\gamma}\mathrm{A}\,\pi^{-\gamma}\,\mathrm{e}^{-(t\pi^{1-2\gamma})/2}$
$\frac{x^\alpha-1}{\alpha(\alpha-1)}$	$\frac{1}{\alpha(\alpha-1)}(\mathrm{Tr}\left[\sigma^\alpha\rho^{1-\alpha}\right]-1)$	$\frac{c_\alpha(s^\alpha+s^{1-\alpha})}{2(1+s)^2}$	$\frac{\alpha(\alpha-1)(x-1)^2}{(x-x^\alpha)(x^{-\alpha}-1)}$	???	$\int_0^\alpha \frac{\mathrm{d}\beta}{\alpha(1-\alpha)}\int_\beta^{1-\beta}\mathrm{d}\gamma\,\mathbb{J}_L^{-2}[\pi^{1-\gamma}\,A\,\pi^\gamma]$
$4(1-\sqrt{x})$	$4(1-\mathrm{Tr}[\sqrt{\rho}\sqrt{\sigma}])$	$\frac{4\sqrt{s}}{\pi(1+s)^2}$	$\frac{1}{4}\left(1+\sqrt{x}\right)^2$	$\frac{1}{4}\{\sqrt{\pi}, \{\sqrt{\pi}, A\}\}$	$\int_0^\infty \mathrm{d}t\int_0^\infty \mathrm{d}s\, e^{-(t+s)\sqrt{\pi}/2}\, A\, e^{-(t+s)\sqrt{\pi}/2}$
$-\log x$	$\mathrm{Tr}\left[(\rho\,(\log\rho-\log\sigma)\right]$	$\frac{1}{2(1+s)}$	$\frac{x-1}{\log x}$	$\mathbb{J}_L=\int_0^1 \mathrm{d}t\,\pi^{1-t}\,A\,\pi^t$	$\mathbb{J}_L^{-1}=\int_0^\infty \mathrm{d}t\,(\pi+t)^{-1}\,A\,(\pi+t)^{-1}$
$\frac{1}{2}(\log x)^2$	$\mathrm{Tr}\left[(\rho\,(\log\rho-\log\sigma)^2\right]$	$\frac{-\log s}{(1+s)^2}$	$\frac{2\,(x-1)^2}{(x+1)(\log x)^2}$	$\int_0^\infty \mathrm{d}t\, e^{-t\pi/2}\,\mathbb{J}_L^2[A]\, e^{-t\pi/2}$	$\frac{1}{2}\,\mathbb{J}_L^{-2}\left[\{\pi, A\}\right]$
$\sqrt{x^{-1}}-\sqrt{x}$	$\mathrm{Tr}\left[\sqrt{\rho}(\rho-\sigma)\sqrt{\sigma^{-1}}\right]$	$\frac{1}{2\pi\sqrt{s}}$	$\sqrt{x}$	$\sqrt{\pi}\,A\,\sqrt{\pi}$	$\sqrt{\pi^{-1}}\,A\,\sqrt{\pi^{-1}}$
$\frac{(x-1)^2}{4x(x^\gamma+x^{1-\gamma})^{-1}}$	$\mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma^{\gamma-1}\mathbb{R}_\rho^{-\gamma}+\mathbb{L}_\sigma^{-\gamma}\mathbb{R}_\rho^{\gamma-1})[(\rho-\sigma)]\right]$	$C_\gamma\frac{\lambda^{\gamma-1}+\lambda^{-\gamma}}{2}$	$\frac{2x}{x^\gamma+x^{1-\gamma}}$	$\int_0^\infty \mathrm{d}t\mathrm{e}^{-(t\pi^{2\gamma-1})/2}\,\pi^\gamma\mathrm{A}\,\pi^\gamma\,\mathrm{e}^{-(t\pi^{2\gamma-1})/2}$	$\frac{1}{2}\{\pi^{2\gamma-1}, \pi^{-\gamma}\,A\,\pi^{-\gamma}\}$
$\frac{(x-1)^2}{2}$	$\frac{1}{2}(\mathrm{Tr}\left[\sigma^2\rho^{-1}\right]-1)$	$\frac{\delta(s)}{2}$	$\frac{2x}{x+1}$	$\int_0^\infty \mathrm{d}t\, e^{-t\pi^{-1}/2}\, A\, e^{-t\pi^{-1}/2}$	$\frac{1}{2}\{\pi^{-1}, A\}$

Figure 1.3: In the table above we summarise the standard monotone functions $f(x)$ analysed in the text. The table is divided in two parts: in the first half we present the properties of the associated contrast function, in the second the corresponding standard monotone. In particular, notice that the measure $\mathrm{d}N(s)$ is the one defined in Eq. (1.78) for asymmetric contrast functions. The question marks correspond to the entries for which no explicit form has been found. For reason of space, we have also introduced the constants $c_\alpha := \frac{\sin(\pi\alpha)}{\pi\alpha(1-\alpha)}$ and $C_\gamma = \frac{\sin(\pi\gamma)}{\pi}$.

taste. Indeed, one could have started by introducing the superoperator associated to the function $k \in \mathcal{K}$:

$$\Omega_k := k(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})\, \mathbb{R}_\pi^{-1} = \mathbb{J}_{Ik}^{-1} . \tag{1.106}$$

The main reason to use Ω_k is that it allows to lift the convex structure of $\mathcal{K}$ to quantum Fisher informations, thanks to the identity $\Omega_{\alpha k_1 + (1-\alpha)k_2} = \alpha\, \Omega_{k_1} + (1-\alpha)\Omega_{k_2}$. The same happened for $\mathbb{J}_f$ and functions of $\mathcal{F}$, as $\mathbb{J}_{\alpha f_1 + (1-\alpha) f_2} = \alpha\, \mathbb{J}_{f_1} + (1-\alpha)\mathbb{J}_{f_2}$, whereas $\mathbb{J}^{-1}_{\alpha f_1 + (1-\alpha) f_2}$ has no simple relations with $\mathbb{J}^{-1}_{f_1}$ and $\mathbb{J}^{-1}_{f_2}$.

We are now ready to start exploring the complete positivity of the quantum Fisher information superoperators. To this end, let us introduce the two sets:

$$\mathcal{F}^+ := \{ f \,|\, \mathbb{J}_f \text{ is CP} \} ; \tag{1.107}$$

$$\mathcal{F}^- := \left\{ f \,|\, \mathbb{J}_f^{-1} \text{ is CP} \right\} . \tag{1.108}$$

Thanks to Eq. (1.86) $\mathcal{F}^+ = T\mathcal{F}^-$. Moreover, it is also useful to introduce the dual sets:

$$\mathcal{K}^+ := \{ k \,|\, \Omega_k \text{ is CP} \} ; \tag{1.109}$$

$$\mathcal{K}^- := \left\{ k \,|\, \Omega_k^{-1} \text{ is CP} \right\} . \tag{1.110}$$

As it is clear from Eq. (1.106), one has the mapping $\mathcal{K}^\pm = I\mathcal{F}^\mp$. The reason why we need so many sets is that $\mathcal{F}^+$ and $\mathcal{K}^+$ are convex, while it was shown in [21] that $\mathcal{F}^-$ and $\mathcal{K}^-$ are not. As convexity greatly simplifies the characterisation of sets, this is a desirable property.

Interestingly, one can completely characterise $\mathcal{F}^+/\mathcal{K}^+$ in terms of a partial order. In order to do so, one needs to introduce the concept of positive definite continuous functions. These are functions $h : \mathbb{R} \to \mathbb{C}$ such that for any vector of reals $\{t_i\}_{i \in \{1,\dots,n\}}$ of arbitrary size n, the matrix A defined in coordinates as $A_{i,j} = h(t_i - t_j)$ is positive semidefinite. This class of function is closed under multiplication, and the corresponding functions are uniformly bounded by their value in zero. Finally, Bochner's theorem says that a function is positive definite if and only if it is the Fourier transform of a finite positive measure on $\mathbb{R}$ [22]. This last property will be particularly useful later.

Then, we define the partial ordering $\preceq$ on both $\mathcal{F}$ and $\mathcal{K}$ by saying that $h_1 \preceq h_2$ if $h_1(e^t)/h_2(e^t)$ is positive definite or, equivalently, if the

matrix defined in coordinates as:

$$A_{i,j} := \frac{h_1(\pi_i/\pi_j)}{h_2(\pi_i/\pi_j)} \tag{1.111}$$

is positive semidefinite for any probability vector $\{\pi_i\}_{i\in\{1,\dots,n\}}$ of any fixed size n. It was shown in [23] that $h_1 \preceq h_2$ implies $h_1(x) \leq h_2(x)$ for every x. Then, we can prove that $\preceq$ is indeed an order relation. First, notice that it is reflexive, i.e., $h \preceq h$, as a matrix with all entries equal to 1 is positive semidefinite. Secondly, it is antisymmetric, meaning that $h_1 \preceq h_2$ and $h_2 \preceq h_1$ if and only if $h_1 = h_2$, since $\preceq$ implies the pointwise order, and no different functions can lay at the same time above and below each other. Finally, we can deduce transitivity from the fact that the product of positive definite functions is positive definite: indeed, from $h_1 \preceq h_2$ and $h_2 \preceq h_3$ it follows that $h_1 \preceq h_3$, as $h_1(e^t)/h_3(e^t)$ can be rewritten as $(h_1(e^t)/h_2(e^t))(h_2(e^t)/h_3(e^t))$, and the functions inside the two parenthesis are positive definite by assumption. These are the defining properties of a partial order.

Then, we can verify that $\mathcal{F}^+$ is the set of all functions satisfying $f(x) \preceq \sqrt{x}$. Indeed, this means that the matrix:

$$A_{i,j} := f\left(\frac{\pi_i}{\pi_j}\right)\sqrt{\frac{\pi_j}{\pi_i}} = \left(f\left(\frac{\pi_i}{\pi_j}\right)\pi_j\right)\frac{1}{\sqrt{\pi_i\pi_j}} = \tag{1.112}$$

$$= \frac{1}{\sqrt{\pi_i}}\,(\mathcal{J}_f)_{i,j}\,\frac{1}{\sqrt{\pi_j}}\,, \tag{1.113}$$

is positive semidefinite. The last step shows that A is similar to the matrix $\mathcal{J}_f$ defined in Eq. (1.83), so it has the same positivity properties. Hence, the condition that $f(x) \preceq \sqrt{x}$ corresponds to the requirement that $\mathcal{J}_f$ is positive semidefinite, which, as we saw in the Sec. 1.3, makes $\mathbb{J}_f$ completely positive thanks to Schur product theorem (see the discussion below Eq. (1.83)).

Moreover, in the same way we can prove that $\mathcal{F}^-$ is given by the complementary set $\sqrt{x} \preceq f(x)$, as this corresponds to the positivity of the matrix:

$$A_{i,j} := \left(f\left(\frac{\pi_i}{\pi_j}\right)\right)^{-1}\sqrt{\frac{\pi_i}{\pi_j}} = \left(f\left(\frac{\pi_i}{\pi_j}\right)\pi_j\right)^{-1}\sqrt{\pi_i\pi_j}\,, \tag{1.114}$$

which thanks to Eq. (1.82) and Eq. (1.83) implies the complete positivity of $\mathbb{J}_f^{-1}$. Using the transformation I, we then have that $\mathcal{K}^+$ corresponds

to functions $k(x) \preceq \sqrt{x^{-1}}$ and $\mathcal{K}^-$ to $\sqrt{x^{-1}} \preceq k(x)$, since $\mathcal{K}^\pm = I\mathcal{F}^\mp$ and $f_1 \preceq f_2 \implies If_2 \preceq If_1$.

Hence, we have shown that all the functions below and above $\sqrt{x}$with respect to $\preceq$ are exactly all the f for which either $\mathbb{J}_f$ or $\mathbb{J}_f^{-1}$ are CP. In Sec. 1.3 it was shown that a similar necessary condition could be given in terms of the pointwise order, so it is natural to wonder whether this is in fact enough. The refutation of this conjecture was presented in [21], where it was proven that for any $f_\lambda(x)$ with $\lambda \neq 1$, $f_\lambda \notin \mathcal{F}^+$, even if for any $\lambda \geq 3 - 2\sqrt{2}$ the corresponding function lays below the square root. Hence, even if for $\lambda \in [3 - 2\sqrt{2}, 1)$ one has $f_\lambda(x) \leq \sqrt{x}$, it still holds that $f_\lambda(x) \not\preceq \sqrt{x}$.

Interestingly, using the characterisation in terms of $\preceq$ one can find the following analytical expressions [21]:

Theorem 4. *For any $f \in \mathcal{F}^+$ there exists a symmetric probability measure $\mathrm{d}\nu(s)$ on $\mathbb{R}$ such that:*

$$\mathbb{J}_f|_\pi[A] = \int_{-\infty}^{\infty} \mathrm{d}\nu(s)\, \pi^{is+\frac{1}{2}}\, A\, \pi^{-is+\frac{1}{2}}\,. \tag{1.115}$$

Similarly, for any $f \in \mathcal{F}^-$ there exists a symmetric probability measure $\mathrm{d}\mu(s)$ on $\mathbb{R}$ such that:

$$\mathbb{J}_f^{-1}|_\pi[A] = \int_{-\infty}^{\infty} \mathrm{d}\mu(s)\, \pi^{is-\frac{1}{2}}\, A\, \pi^{-is-\frac{1}{2}}\,. \tag{1.116}$$

Proof. This result is a straightforward application of Bochner's theorem. Let us first analyse the case in which $f \in \mathcal{F}^+$. As we discussed above, this implies that $f(x) \preceq \sqrt{x}$, meaning that $e^{-t/2} f(e^t)$ is positive definite. Hence, from Bochner's theorem there exists a unique probability measure $\mathrm{d}\nu(s)$ on $\mathbb{R}$ such that:

$$e^{-t/2} f(e^t) = \int_{-\infty}^{\infty} \mathrm{d}\nu(s)\, e^{its}\,. \tag{1.117}$$

Since f is standard it follows that $e^{-t/2} f(e^t) = e^{-t/2} e^t f(e^{-t}) = e^{t/2} f(e^{-t})$, showing that $\mathrm{d}\nu(s)$ has to be symmetric with respect to zero. Then, applying the definition of $\mathbb{J}_f$ in Eq. (1.65) we have:

$$\mathbb{J}_f\big|_\pi[A] = \mathbb{R}_\pi\, f\big(\mathbb{L}_\pi \mathbb{R}_\pi^{-1}\big)[A] = \tag{1.118}$$

$$= \mathbb{R}_\pi \int_{-\infty}^{\infty} \mathrm{d}\nu(s)\, (\mathbb{L}_\pi \mathbb{R}_\pi^{-1})^{is+1/2}[A] = \tag{1.119}$$

$$= \int_{-\infty}^{\infty} \mathrm{d}\nu(s)\, \pi^{is+1/2}\, A\, \pi^{-is+1/2}\,, \tag{1.120}$$

proving Eq. (1.115). Similarly, for $f \in \mathcal{F}^-$, Bochner's theorem implies the existence of a unique probability measure $\mathrm{d}\mu(s)$ on $\mathbb{R}$ such that:

$$\frac{e^{t/2}}{f(e^t)} = \int_{-\infty}^{\infty} \mathrm{d}\mu(s)\, e^{its}\,, \tag{1.121}$$

and from the fact that f is standard we can verify that $\mathrm{d}\mu(s)$ has to be symmetric with respect to zero. Then, the same kind of computations as above shows that:

$$\mathbb{J}_f^{-1}\big|_\pi[A] = \mathbb{R}_\pi^{-1}\,\frac{1}{f}(\mathbb{L}_\pi\mathbb{R}_\pi^{-1})[A] = \tag{1.122}$$

$$= \mathbb{R}_\pi^{-1}\int_{-\infty}^{\infty} \mathrm{d}\mu(s)\,(\mathbb{L}_\pi\mathbb{R}_\pi^{-1})^{is-1/2}[A] = \tag{1.123}$$

$$= \int_{-\infty}^{\infty} \mathrm{d}\nu(s)\,\pi^{is-1/2}\,A\,\pi^{-is-1/2}\,, \tag{1.124}$$

proving the claim. □

Thm. 4 shows that if $f \in \mathcal{F}^{\pm}$, one can find an explicit Kraus form for $\mathbb{J}_f^{\pm 1}$ by using Eq. (1.115) or Eq. (1.116), where the probability measure can be obtained by inverting Eq. (1.117) or Eq. (1.121) respectively.

After having discussed these general families of standard monotones, in the next subsections we study some specific examples. Since the set of monotones is partially ordered, this will be mirrored in the ordering of the sections, where we start from the largest standard monotone and we go down from there.

1.4.2 The Bures metric

Among the standard monotone functions, the maximum is given by:

$$f_B(x) = \frac{x+1}{2}\,. \tag{1.125}$$

This function can be obtained from Eq. (1.66) if one chooses $g(x)$ to be:

$$g_B(x) = \frac{(x-1)^2}{x+1} = \int_0^1 \mathrm{d}N_B(s)\,\left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right)\,, \tag{1.126}$$

with $\mathrm{d}N_B(s) := \delta(1-s)/2$. The corresponding contrast function is given by:

$$H_{\mathrm{B}}(\rho||\sigma) = \mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma + \mathbb{R}_\rho)^{-1}(\rho-\sigma)\right]\,. \tag{1.127}$$

Since the quantity above is real, taking its complex conjugate does not affect it, while it exchanges $\mathbb{L}$ with $\mathbb{R}$ and vice versa. This shows that $H_B(\rho||\sigma)$ is symmetric in its arguments. The same conclusion could also be derived by noticing that $g_B(x) = x g_B(x^{-1})$. Hence, $H_{\rm B}(\rho||\sigma) = H_{\rm B}(\sigma||\rho)$.

The associated quantum Fisher information operators take the form:

$$\mathbb{J}_B\big|_\pi[A] = \frac{1}{2}\{\pi, A\}\,; \qquad \mathbb{J}_B^{-1}\big|_\pi[A] = \int_0^\infty \mathrm{d}t\ e^{-t\pi/2}\, A\, e^{-t\pi/2}\,. \quad (1.128)$$

The metric generated by $\mathbb{J}_B^{-1}$ is called Bures metric. This operator gives the explicit expression of what is called the symmetric logarithmic derivative, see Eq. (1.138). One can verify this expression of $\mathbb{J}_B^{-1}$ by noticing that:

$$\mathbb{J}_B^{-1}[\mathbb{J}_B[A]] = \int_0^\infty \mathrm{d}t\ \left(\left(e^{-t\pi/2}\frac{\pi}{2}\right) A\, e^{-t\pi/2} + e^{-t\pi/2} A \left(e^{-t\pi/2}\frac{\pi}{2}\right)\right) = \quad (1.129)$$

$$= -\int_0^\infty \mathrm{d}t\ \left(\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-t\pi/2} A\, e^{-t\pi/2}\right)\right) = (e^{-t\pi/2} A\, e^{-t\pi/2})\Big|_{t=\infty}^{t=0} = A\,. \quad (1.130)$$

Interestingly, there is a close form for the geodesic distance associated to this metric [24]. In particular, it takes the form:

$$d_B(\rho,\,\sigma) = 2\arccos\left(\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]\right) = \quad (1.131)$$

$$= 2\arccos\left(\sqrt{F(\rho,\,\sigma)}\right)\,, \quad (1.132)$$

where in the last step we have implicitly defined the fidelity $F(\rho,\sigma)$. The quantity just obtained should be compared with the Bhattacharyya distance for classical probability vector in Eq. (1.16), to which it reduces for commuting states. Eq. (1.132) should also be compared to the Bures length, defined as [25]:

$$D_B(\rho,\,\sigma)^2 = \inf\left\{\mathrm{Tr}\left[(W-X)(W-X)^\dagger\right]\,\middle|\, WW^\dagger = \rho,\ XX^\dagger = \sigma\right\} = \quad (1.133)$$

$$= 2\left(1 - \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]\right) = 2\left(1 - \cos\left(\frac{d_B(\rho,\,\sigma)}{2}\right)\right)\,. \quad (1.134)$$

This further shows the analogy with the classical case. Indeed, one can think of the Bures length as the analogous of the Hellinger distance for quantum states: in fact, Eq. (1.133) simply tells us that $D_B(\rho, \sigma)^2$ coincides with the minimum Hilbert-Schmidt distance on purifications of ρ and σ, which is the non-commutative analogous of the Euclidean metric. Indeed, when one restricts to diagonal states, this exactly reduces to Eq. (1.13).

In the literature the Bures metric corresponds to the most canonical choice of quantum Fisher information. In fact, it appears in the context of quantum metrology in the famous Cramer-Rao bound, which is a limit on the quality with which one can estimate parameters encoded in a state. In order to see this, consider a family of density matrices $\rho(\theta)$ depending on some parameters θ. Without loss of generality, suppose that the true value is $\theta = 0$. We define a locally unbiased estimator to be an observable satisfying [26]:

$$\frac{\partial}{\partial\theta} \operatorname{Tr}\left[\rho(\theta)A\right]\Big|_{\theta=0} = 1\,. \tag{1.135}$$

This equation tells us that by measuring A in a neighbourhood of $\theta = 0$, one obtains the correct value of θ on average:

$$\exists\,\varepsilon \;\big|\quad \operatorname{Tr}\left[\rho(\theta)A\right] = \theta \quad \forall\theta \in (-\varepsilon, \varepsilon)\,. \tag{1.136}$$

Define now the symmetric logarithmic derivative (SLD) of $\rho(\theta)$ to be the operator L such that, for any B, we have:

$$\frac{\partial}{\partial\theta} \operatorname{Tr}\left[\rho(\theta)B\right]\Big|_{\theta=0} = \operatorname{Tr}\left[B\,L\rho(0)\right] = \operatorname{Tr}\left[B\,\mathbb{J}_B\big|_{\rho(0)}[L]\right]\,. \tag{1.137}$$

Since this relation holds for any B, one can impose the equality even without the trace, giving:

$$\frac{\partial\rho(\theta)}{\partial\theta}\Big|_{\theta=0} = \frac{1}{2}\left(L\rho(0) + \rho(0)L\right)\,. \tag{1.138}$$

This expression makes it clear where the name for SLD comes from, as L can be thought as a symmetric generalisation of the differential operator to non-commutative variables.

We are now ready to prove the Cramer-Rao bound: this is a fundamental limit on how small the variance of locally unbiased operators can

be. Writing it down explicitly, we have:

$$\mathrm{Tr}\left[\rho(0)A^2\right] = \mathrm{Tr}\left[A\,\mathbb{J}_B\big|_{\rho(0)}[A]\right] \geq \tag{1.139}$$

$$\geq \frac{\mathrm{Tr}\left[A\,\mathbb{J}_B\big|_{\rho(0)}[L]\right]}{\mathrm{Tr}\left[L\,\mathbb{J}_B\big|_{\rho(0)}[L]\right]} = \frac{1}{\mathrm{Tr}\left[\partial_\theta\rho(\theta)\,\mathbb{J}_B\big|^{-1}_{\rho(\theta)}[\partial_\theta\rho(\theta)]\right]\big|_{\theta=0}}\,, \tag{1.140}$$

where in the step between the first and the second line we used the Cauchy-Schwartz inequality for $\mathbb{J}_B$, and in the second line we used the definition of local unbiased operators (Eq. (1.135)) and we inverted Eq. (1.138). This shows that the ability to estimate the parameter θ is intrinsically connected to the statistical difference between $\rho(0)$ and $\rho(0+\partial\theta)$.

It should be pointed out that the steps presented above can in principle be replicated for the other quantum Fisher information. In fact, it is sufficient to define the generalised derivative L_f as:

$$\left.\frac{\partial\rho(\theta)}{\partial\theta}\right|_{\theta=0} = \mathbb{J}_f\big|_{\rho(0)}[L_f]\,, \tag{1.141}$$

to obtain:

$$\mathrm{Tr}\left[A\,\mathbb{J}_f\big|_{\rho(0)}[A]\right] \geq \frac{1}{\mathrm{Tr}\left[L_f\,\mathbb{J}_f\big|_{\rho(0)}[L_f]\right]} = \frac{1}{\mathrm{Tr}\left[\partial_\theta\rho(\theta)\,\mathbb{J}_f^{-1}\big|_{\rho(0)}[\partial_\theta\rho(\theta)]\right]\big|_{\theta=0}}\,. \tag{1.142}$$

This procedure gives a whole family of bounds. Whereas for the Bures case the variance of an observable is a quite standard object, the problem here is to find an operational interpretation to the generalised variance on the left in the above equation. Still, this can be done in some contexts, as for example for the relative entropy, Sec. 1.4.6.

1.4.3 The Heinz family

Let us introduce the one-parameter family of functions $g_{\gamma_>}(x)$ defined as:

$$g_{\gamma_>}(x) = \frac{(x-1)^2}{x^\gamma + x^{1-\gamma}}\,, \tag{1.143}$$

for $\gamma \in [0,1]$, corresponding to the contrast functions:

$$H_{\gamma_>}(\rho||\sigma) = \mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma^\gamma\mathbb{R}_\rho^{1-\gamma} + \mathbb{L}_\sigma^{1-\gamma}\mathbb{R}_\rho^\gamma)^{-1}[(\rho-\sigma)]\right] \tag{1.144}$$

where in the first line we used the expression in Eq. (1.46). It is useful for what follows to provide an integral equation to this contrast function. To do so, we need to invert the operator:

$$(\mathbb{L}_\sigma^\gamma \mathbb{R}_\rho^{1-\gamma} + \mathbb{L}_\sigma^{1-\gamma}\mathbb{R}_\rho^\gamma)[A] = \sigma^\gamma A \rho^{1-\gamma} + \sigma^{1-\gamma} A \rho^\gamma = \tag{1.145}$$

$$= (\sigma^\gamma A \rho^\gamma)\rho^{1-2\gamma} + \sigma^{1-2\gamma}(\sigma^\gamma A \rho^\gamma) = \tag{1.146}$$

$$= (\mathbb{L}_\sigma^{1-2\gamma} + \mathbb{R}_\rho^{1-2\gamma})\mathbb{L}_\sigma^\gamma \mathbb{R}_\rho^\gamma[A]\,. \tag{1.147}$$

The inverse of $(\mathbb{L}_\sigma \mathbb{R}_\rho)^\gamma$ is simply given by $(\mathbb{L}_\sigma \mathbb{R}_\rho)^{-\gamma}$. On the other hand, we can use the formula:

$$\sigma A + A\rho = L \implies L = \int_0^\infty \mathrm{d}t\ e^{-t\sigma} A\, e^{-t\rho}\,, \tag{1.148}$$

to invert the operator $(\mathbb{L}_\sigma^{1-2\gamma} + \mathbb{R}_\rho^{1-2\gamma})$. The proof of Eq. (1.148) is completely analogous to the procedure presented in Eq. (1.130), so we do not repeat it here. Then, one can express the contrast function in Eq. (1.144) as:

$$H_{\gamma_>}(\rho||\sigma) = \int_0^\infty \mathrm{d}t\ \mathrm{Tr}\left[(\rho-\sigma)e^{-t\sigma^{1-2\gamma}}\sigma^{-\gamma}(\rho-\sigma)\rho^{-\gamma}e^{-t\rho^{1-2\gamma}}\right]. \tag{1.149}$$

The corresponding family of standard monotone functions is given by:

$$f_{\gamma_>}(x) := \frac{x^\gamma + x^{1-\gamma}}{2}\,. \tag{1.150}$$

Thanks to the Löwner-Heinz inequality we know that x^γ is matrix monotone for all $\gamma \in [0,1]$ [18], which means that $f_{\gamma_>}(x)$ is matrix monotone as well. Using the integral expression for powers $\gamma \in (0,1)$:

$$x^\gamma = \frac{\sin\pi\gamma}{\pi}\int_0^\infty \mathrm{d}\lambda\ \lambda^{\gamma-1}\frac{x}{x+\lambda} = \tag{1.151}$$

$$= \frac{\sin\pi\gamma}{\pi}\int_0^1 \mathrm{d}\lambda\ \left(\lambda^{\gamma-1}\frac{x}{x+\lambda} + \lambda^{-\gamma}\frac{x}{1+\lambda\,x}\right), \tag{1.152}$$

we can rewrite the functions $f_{\gamma_>}(x)$ in terms of the extreme points as:

$$f_{\gamma_>}(x) = \int_0^1 \mathrm{d}\lambda \left(\frac{\sin\pi\gamma}{\pi}\left(\frac{\lambda^{\gamma-1}+\lambda^{-\gamma}}{1+\lambda}\right)\right) f_\lambda(x)\,, \tag{1.153}$$

where we isolated the terms which defines the measure $\mathrm{d}\mu_{Tf_{\gamma_>}}(\lambda)$. Moreover, it is easy to show that all $f_{\gamma_>}(x)$ lay above $\sqrt{x}$, as it easily follows from the fact that x^γ has a unique minimum at $\gamma = 1/2$. It should be noticed that for $\gamma = 0$ or $\gamma = 1$ one has $f_{0_>} = f_{1_>} = f_B$.

We can then use Eq. (1.149) to obtain the expression of the quantum Fisher information. Using Thm. 2 we can indeed identify the superoperator:

$$\mathbb{J}^{-1}_{f_{\gamma_>}}\big|_\pi[A] = \int_0^\infty \mathrm{d}t\ e^{-(t\pi^{1-2\gamma})/2}\,\pi^{-\gamma}A\,\pi^{-\gamma}\,e^{-(t\pi^{1-2\gamma})/2}\,. \tag{1.154}$$

Since $\mathbb{J}^{-1}_{f_{\gamma_>}}$ is given in Kraus form, it is immediately clear that it is completely positive. This shows that $f_{\gamma_>} \in \mathcal{F}^-$ for all $\gamma \in [0,1]$ or, equivalently, $\sqrt{x} \preceq f_{\gamma_>}(x)$. Moreover, the complex hull of standard convex $k_{\gamma_>} := 2/(x^\gamma + x^{1-\gamma})$ is also in $\mathcal{K}^+$. Moreover, using Eq. (1.128) we can also provide an expression for $\mathbb{J}_{f_{\gamma_>}}$, namely:

$$\mathbb{J}_{f_{\gamma_>}}\big|_\pi[A] = \frac{1}{2}\{\pi^{1-2\gamma}, \pi^\gamma\, A\,\pi^\gamma\}\,. \tag{1.155}$$

Thanks to the properties of the map T, the image $f_{\gamma_<} := Tf_{\gamma_>}$ is contained in $\mathcal{F}^+$. This family of monotones are given by:

$$f_{\gamma_<}(x) := [Tf_{\gamma_>}](x) = \frac{2x}{x^\gamma + x^{1-\gamma}}\,. \tag{1.156}$$

From the definition $f_{\gamma_<}(x) \preceq \sqrt{x}$, which in particularly implies that $f_{\gamma_<}(x) \leq \sqrt{x}$ for each x. Indeed, one can use Eq. (1.86) to derive the expression of $\mathbb{J}^{\pm}_{f_{\gamma_<}}$ from Eq. (1.154) and Eq. (1.155), which reads:

$$\mathbb{J}_{f_{\gamma_<}}\big|_\pi[A] = \mathbb{J}_{Tf_{\gamma_>}}\big|_\pi[A] = \int_0^\infty \mathrm{d}t\ e^{-(t\pi^{2\gamma-1})/2}\,\pi^\gamma A\pi^\gamma\,e^{-(t\pi^{2\gamma-1})/2}\,; \tag{1.157}$$

$$\mathbb{J}^{-1}_{f_{\gamma_<}}\big|_\pi[A] = \mathbb{J}^{-1}_{Tf_{\gamma_>}}\big|_\pi[A] = \frac{1}{2}\,\{\pi^{2\gamma-1}, \pi^{-\gamma}A\pi^{-\gamma}\}\,, \tag{1.158}$$

Again, one can confirm that all $\mathbb{J}_{f_{\gamma_<}}$ are CP, as they are presented in Kraus form. In this case, it should be noticed that the convex hull of $f_{\gamma_<}$ is still in $\mathcal{F}^+$, without the need to pass to the inverse. Moreover, it is interesting to point out that the only function both in $f_{\gamma_>}$ and in the $f_{\gamma_<}$ is the square root, which follows from the fact that this is the only function both in $\mathcal{F}^+$ and $\mathcal{F}^-$.

The contrast functions associated to Eq. (1.156) come from the family of convex functions:

$$g_{\gamma<} = \frac{(x^\gamma + x^{1-\gamma})(x-1)^2}{4x} \tag{1.159}$$

and take the form:

$$H_{\gamma>}(\rho||\sigma) = \mathrm{Tr}\left[(\rho-\sigma)(\mathbb{L}_\sigma^{\gamma-1}\mathbb{R}_\rho^{-\gamma} + \mathbb{L}_\sigma^{-\gamma}\mathbb{R}_\rho^{\gamma-1})[(\rho-\sigma)]\right] \tag{1.160}$$

Moreover, thanks to Eq. (1.153) and Eq. (1.102) one also has:

$$g_{\gamma<}(x) = \int_0^1 \mathrm{d}\lambda \left(\frac{\sin\pi\gamma}{\pi}\left(\frac{\lambda^{\gamma-1}+\lambda^{-\gamma}}{2}\right)\right)\left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right) \tag{1.161}$$

so that we can identify $\mathrm{d}N_g(s) := \frac{\sin\pi\gamma}{\pi}\left(\frac{\lambda^{\gamma-1}+\lambda^{-\gamma}}{2}\right)\mathrm{d}s$.

1.4.4 The family of α-divergences

Consider the family of convex functions:

$$g_\alpha(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)} = \frac{\sin\pi\alpha}{\pi\alpha(1-\alpha)}\int_0^\infty \mathrm{d}s\, \frac{s^\alpha}{(1+s)^2}\left(\frac{(x-1)^2}{x+s}\right). \tag{1.162}$$

It should be noticed that in the integral above one has an extra term of the form $(1-x)/(1-\alpha)$, which we ignore as it does not contribute to the expression of the contrast function (see the beginning of Sec. 1.2). The corresponding symmetrised version is given by:

$$g_\alpha^{\mathrm{symm}}(x) = \frac{(1-x^\alpha)(1-x^{1-\alpha})}{2\alpha(1-\alpha)}, \tag{1.163}$$

which is matrix convex for $\alpha \in [-1, 2]$. Moreover, we can use Eq. (1.162) to provide an expression for $\mathrm{d}N_g(s)$, which can be inferred from the following integral:

$$g_\alpha^{\mathrm{symm}}(x) = \frac{\sin\pi\alpha}{\pi\alpha(1-\alpha)}\int_0^1 \mathrm{d}s\, \frac{(s^\alpha + s^{1-\alpha})}{2(1+s)^2}\left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right), \tag{1.164}$$

so that we can identify $\mathrm{d}N_g(s) := \frac{\sin\pi\alpha}{\pi\alpha(1-\alpha)}\frac{(s^\alpha+s^{1-\alpha})}{2(1+s)^2}\mathrm{d}s$. This expression corrects the one provided in [16].

Another interesting integral expression for $g_\alpha(x)$ is given by:

$$g_\alpha(x) = \frac{1}{\alpha(\alpha-1)} \int_0^\alpha \mathrm{d}\beta \; (x^\beta \; \log x) \,. \tag{1.165}$$

This can be used to compute the associated contrast functions:

$$H_\alpha(\rho||\sigma) = \frac{1}{\alpha(\alpha-1)} \left(\mathrm{Tr}\left[\sigma^\alpha \rho^{1-\alpha}\right] - 1 \right) = \tag{1.166}$$

$$= \frac{1}{\alpha(\alpha-1)} \int_0^\alpha \mathrm{d}\beta \; \mathrm{Tr}\left[\rho^{1-\beta}(\log\sigma - \log\rho)\sigma^\beta\right] \,. \tag{1.167}$$

We call these quantities α-divergences. A similar family is the one of Rényi divergences, given by:

$$S_\alpha(\rho||\sigma) := \frac{1}{\alpha-1} \log \; \mathrm{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right] = \tag{1.168}$$

$$= \frac{1}{\alpha-1} \log \left(1 + \int_0^\alpha \mathrm{d}\beta \; \mathrm{Tr}\left[\rho^\beta (\log\rho - \log\sigma)\sigma^{1-\beta}\right] \right) \,. \tag{1.169}$$

It is easy to see that the two are related by the equation:

$$H_\alpha(\rho||\sigma) = \frac{e^{-\alpha S_{1-\alpha}(\rho||\sigma)} - 1}{\alpha(\alpha-1)} \,. \tag{1.170}$$

Interestingly, the α-contrast functions and the corresponding Rényi divergences locally give rise to the same metric structure, which is related to the statistics of work close to equilibrium (see Eq. (2.28)).

An expansion of these quantities was provided in [7], but we present here a more compact derivation. First, it should be noticed that locally $H_\alpha(\rho||\sigma)$ is indistinguishable from $H_\alpha^{\mathrm{symm}}(\rho||\sigma)$, so we can do the expansion directly on the latter. In particular, one can use the following integral expression:

$$2\alpha(\alpha-1)\, H_\alpha^{\mathrm{symm}}(\rho||\sigma) = \left(\mathrm{Tr}\left[\sigma^\alpha \rho^{1-\alpha}\right] + \mathrm{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right] - 2 \right) = \tag{1.171}$$

$$= \int_0^\alpha \mathrm{d}\beta \; \left(\mathrm{Tr}\left[\rho^{1-\beta}(\log\sigma - \log\rho)\sigma^\beta\right] + \mathrm{Tr}\left[\rho^\beta(\log\rho - \log\sigma)\sigma^{1-\beta}\right] \right) = \tag{1.172}$$

$$= \int_0^\alpha \mathrm{d}\beta \left(\int_0^\beta \mathrm{d}\gamma \; \mathrm{Tr}\left[\rho^{1-\gamma}(\log\sigma - \log\rho)\sigma^\gamma(\log\sigma - \log\rho)\right] + \right.$$

$$\left. - \int_\beta^1 \mathrm{d}\gamma \; \mathrm{Tr}\left[\rho^\gamma(\log\rho - \log\sigma)\sigma^{1-\gamma}(\log\rho - \log\sigma)\right] \right) \,, \tag{1.173}$$

where the last equality can be verified by carrying out the integrals and see that it correctly retrieves Eq. (1.172). Substituting $\gamma \to 1-\gamma$ in the second integral of Eq. (1.173), one finally obtains:

$$2\alpha(\alpha-1)\, H_\alpha^{\rm symm}(\rho||\sigma) = \tag{1.174}$$

$$= \int_0^\alpha \mathrm{d}\beta \left(\int_0^\beta \mathrm{d}\gamma - \int_0^{1-\beta} \mathrm{d}\gamma \right) \left(\mathrm{Tr}\left[\rho^{1-\gamma} (\log\sigma - \log\rho) \sigma^\gamma (\log\sigma - \log\rho) \right] \right) = \tag{1.175}$$

$$= -\int_0^\alpha \mathrm{d}\beta \int_\beta^{1-\beta} \mathrm{d}\gamma\ \mathrm{Tr}\left[\rho^{1-\gamma} (\log\sigma - \log\rho) \sigma^\gamma (\log\sigma - \log\rho) \right] . \tag{1.176}$$

It should be noticed that the symmetry in the arguments of $H_\alpha^{\rm symm}(\rho||\sigma)$ is reflected in the symmetry under the transformation $\alpha \to 1-\alpha$. Moreover, thanks to the appearance of the two differences of logarithms in Eq. (1.176), it is straightforward to give the local expansion of the α-divergences. In fact, denote by $\mathbb{J}_L^{-1}$ the Fréchet derivative of the logarithm, which reads in formulae [18]:

$$\mathbb{J}_L^{-1}[A] := \lim_{\varepsilon\to 0} \frac{\log(\pi+\varepsilon A) - \log\pi}{\varepsilon} = \int_0^\infty \mathrm{d}t\ \ (\pi+t)^{-1} A\, (\pi+t)^{-1} . \tag{1.177}$$

Hence, one can approximate the operator $\log(\pi+\varepsilon A)$ by $\log(\pi)+\varepsilon\mathbb{J}_L^{-1}[A]$. Then, a simple substitution shows that the expansion of the α-divergences is given by:

$$H_\alpha(\pi||\pi+\varepsilon\delta\rho) = \frac{S_{1-\alpha}(\pi||\pi+\varepsilon\delta\rho)}{1-\alpha} = \tag{1.178}$$

$$= \frac{\varepsilon^2}{2\,\alpha(1-\alpha)} \int_0^\alpha \mathrm{d}\beta \int_\beta^{1-\beta} \mathrm{d}\gamma\ \mathrm{Tr}\left[\pi^{1-\gamma}\, \mathbb{J}_L^{-1}[\delta\rho]\, \pi^\gamma \mathbb{J}_L^{-1}[\delta\rho] \right] = \tag{1.179}$$

$$= \frac{\varepsilon^2}{2\,\alpha(1-\alpha)} \int_0^\alpha \mathrm{d}\beta \int_\beta^{1-\beta} \mathrm{d}\gamma\ \mathrm{cov}_\pi^\gamma(\mathbb{J}_L^{-1}[\delta\rho],\, \mathbb{J}_L^{-1}[\delta\rho]) , \tag{1.180}$$

where in the last line we implicitly defined the γ-covariance. This quantity can be connected to the Wigner-Yanase-Dyson skew information, defined as:

$$I^x(\pi, L) = -\frac{1}{2}\mathrm{Tr}\left[[\pi^x, L][\pi^{1-x}, L] \right] = \tag{1.181}$$

$$= \mathrm{Tr}\left[L^2\, \pi \right] - \mathrm{Tr}\left[\pi^{1-x}\, L\, \pi^x\, L \right] , \tag{1.182}$$

which can be interpreted as a quantifier of the quantum uncertainty of the observable L as measured in the state π [27]. In fact, by adding and subtracting the variance of $\mathbb{J}_L^{-1}[\delta\rho]$ to Eq. (1.180), one obtains:

$$H_\alpha(\pi||\pi+\varepsilon\delta\rho) = \tag{1.183}$$

$$= \frac{\varepsilon^2}{2}\mathrm{Tr}\left[\left(\mathbb{J}_L^{-1}[\delta\rho]\right)^2\,\pi\right] + \frac{\varepsilon^2}{2\,\alpha(\alpha-1)}\int_0^\alpha \mathrm{d}\beta\int_\beta^{1-\beta}\mathrm{d}\gamma\; I^\gamma(\pi,\mathbb{J}_L^{-1}[\delta\rho])\,. \tag{1.184}$$

This expression is particularly useful when one wants to isolate the effects of the coherences in the basis of π. Notice in fact that for full rank states π, the Wigner-Yanase-Dyson skew information $I^\gamma(\pi,\mathbb{J}_L^{-1}[\delta\rho]) = 0$ if and only if $[\pi,\delta\rho]=0$.

The standard monotones associated to $H_\alpha(\rho||\sigma)$ are given by:

$$f_\alpha(x) = \frac{\alpha(1-\alpha)(x-1)^2}{(1-x^\alpha)(1-x^{1-\alpha})}\,, \tag{1.185}$$

which are clearly symmetric under the exchange $\alpha\to 1-\alpha$. Applying Thm. 2 to Eq. (1.180) one can deduce that the corresponding family of quantum Fisher information takes the form:

$$\mathbb{J}_\alpha^{-1}\big|_\pi[A] = \frac{1}{\alpha(1-\alpha)}\int_0^\alpha \mathrm{d}\beta\int_\beta^{1-\beta}\mathrm{d}\gamma\;\mathbb{J}_L^{-1}\left[\,\pi^{1-\gamma}\,\mathbb{J}_L^{-1}[A]\,\pi^\gamma\right]\,. \tag{1.186}$$

Moreover, in [28] another integral expression, only valid for $\alpha\in(0,1)$, was provided, namely:

$$\mathbb{J}_\alpha^{-1}\big|_\pi[A] = \frac{(\sin\pi\alpha)^2}{\pi^2\alpha(1-\alpha)}\int_0^\infty \mathrm{d}s\int_0^\infty \mathrm{d}t\;\frac{s^\alpha}{(\pi+s)(\pi+t)}A\frac{t^{1-\alpha}}{(\pi+s)(\pi+t)}\,, \tag{1.187}$$

which clearly shows that in this range $\mathbb{J}_\alpha^{-1}$ is completely positive, which means that the corresponding monotone function satisfies $f_\alpha\in\mathcal{F}^-$. Moreover, in [21] it was shown that $f_\alpha\in\mathcal{F}^+$ for $\alpha\in[-1,-\frac{1}{2}]\cup[\frac{3}{2},2]$, meaning that for this set of parameters $\mathbb{J}_\alpha$ is CP, whereas, for $\alpha\in[-\frac{1}{2},0)\cup(1,\frac{3}{2},2]$ neither $\mathbb{J}_\alpha$ or $\mathbb{J}_\alpha^{-1}$ are CP. Unfortunately, we weren't able to find a general expression for $\mathbb{J}_\alpha$ not even in the parameter range for which it is completely positive.

In the next sections, different examples of quantum Fisher information as α varies. First, it should be noticed that in the range $\alpha\in[0,1/2]$

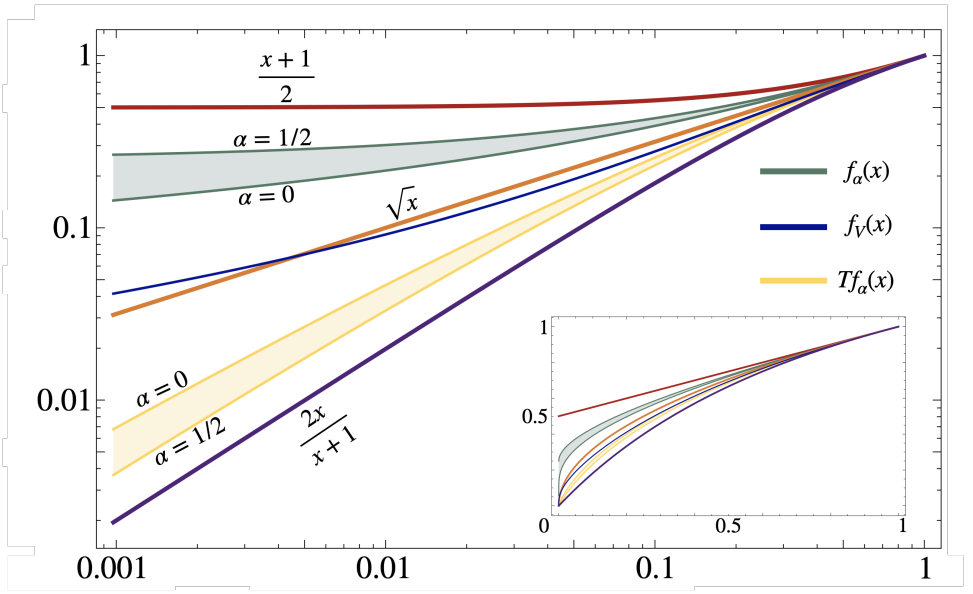


Figure 1.4: In the figure some of the most notable standard monotones are presented in a log-log scale (in the inset we show their behaviour in linear coordinates). In particular, we show the two extrema (f_B and f_H), the square root, the family f_α of α-divergences in the range $\alpha \in [0,1]$ and its transform Tf_α, together with the standard monotone associated with the quantum information variance $f_V(x)$ defined in Eq. (1.229). The shading in the two curves associated to f_α indicates that this family interpolates between f_0, the standard monotone for the relative entropy, and the maximum value $f_{1/2} = \frac{1}{4}(1+\sqrt{x})^2$, corresponding to the Wigner-Yanase skew information. It is interesting to notice that the monotone associated to the quantum information variance does not satisfy $f_V(x) \geq \sqrt{x}$ nor $f_V(x) \leq \sqrt{x}$. This shows that there are monotones for which both $\mathbb{J}_f$ and $\mathbb{J}_f^{-1}$ are not CP.

$f_\alpha(x)$ are monotonically increasing in α, whereas for $\alpha \in [1/2, 2]$ they are monotonically decreasing. There are three limits that are notable enough to deserve a name: the Fisher information associated with the relative entropy is the one given by the limit $\alpha \to 0$, called Kubo-Mori-Bogoliubov (KMB) inner product; in the limit $\alpha \to 1/2$ one obtains the largest among the family, the Wigner-Yanase metric; finally, for $\alpha \to 2$ one gets the minimal function $f_H(x) = 2x/(x+1)$, called the harmonic mean.

1.4.5 The Wigner-Yanase skew information ($\alpha = 1/2$)

The first case we consider is given by the Wigner-Yanase metric. This is associated to the case $\alpha = \frac{1}{2}$, giving rise to the following contrast function:

$$g_{WY}(x) = 4(1 - \sqrt{x}), \qquad H_{WY}(\rho||\sigma) = 4(1 - \mathrm{Tr}\left[\sqrt{\rho}\,\sqrt{\sigma}\right]). \quad (1.188)$$

The corresponding standard monotone function takes the particularly simple form:

$$f_{WY}(x) = \left(\frac{1+\sqrt{x}}{2}\right)^2. \quad (1.189)$$

This expression should be compared with $f_B(x)$ in Eq. (1.125). In fact, it can be easily verified that $f_{WY}(x) \equiv (f_B(\sqrt{x}))^2$. Hence, the quantum Fisher information for the Wigner-Yanase metric can be directly computed from the one for the Bures as:

$$\mathbb{J}_{WY}\big|_\pi[A] = \frac{1}{4}\{\sqrt{\pi}, \{\sqrt{\pi}, A\}\}; \quad (1.190)$$

$$\mathbb{J}^{-1}_{WY}\big|_\pi[A] = \int_0^\infty \mathrm{d}t \int_0^\infty \mathrm{d}s\; e^{-(t+s)\sqrt{\pi}/2}\, A\, e^{-(t+s)\sqrt{\pi}/2}. \quad (1.191)$$

It is a remarkable fact that the space of d-dimensional density matrices endowed with the metric generated by $\mathbb{J}^{-1}_{WY}$ is locally isometric to an n-sphere of radius 2 in $d^2 - 1$ dimensions [29]. Hence, one can give a closed form for the geodesic distance, i.e.:

$$d_{WY}(\rho, \sigma) = 2\,\arccos \mathrm{Tr}\left[\sqrt{\rho}\sqrt{\sigma}\right], \quad (1.192)$$

where the similarity with the Bures geodesic distance in Eq. (1.132) is evident. Indeed, both reduce to the same quantity for commuting

states, i.e., to the Bhattacharyya distance in Eq. (1.16). Moreover, since $\mathrm{Tr}\left[\sqrt{\rho}\sqrt{\sigma}\right] \leq \sqrt{F(\rho,\sigma)}$ [30], it also holds that $d_B(\rho,\sigma) \leq d_{WY}(\rho,\sigma)$, as the arccosine is monotonically decreasing. To the best of the author's knowledge, these are the only two cases for which one has an analytical expression for the geodesic distance. In this case, one even has a simple expression for the geodesic path connecting any two density matrices ρ and σ, which can be parametrised as:

$$\gamma_{WY}^{\rho\to\sigma}(t) = 2\,\frac{\left((1-t)\sqrt{\rho}+t\sqrt{\sigma}\right)^2}{\mathrm{Tr}\left[\left((1-t)\sqrt{\rho}+t\sqrt{\sigma}\right)^2\right]}\,. \tag{1.193}$$

Another important property of this metric is that it can be used to express the quantum Chernoff bound [31]. This arise in the following setting: consider the task of distinguishing two different states ρ_0 and ρ_1, knowing that each is prepared with a probability p_0 and p_1. In this context, the symmetric distinguishability problem consists in finding a POVM E_0, E_1 such that the probability of error $P_e := p_0\mathrm{Tr}\left[E_1\rho_0\right] + p_1\mathrm{Tr}\left[E_0\rho_1\right]$ is minimal. By defining the positive and negative part of a hermitian operator as $A_\pm := (|A| \pm A)/2$, one can prove that the optimal measurement is obtained by setting E_1 to be the projector on the range of $(p_1\rho_1 - p_0\rho_0)_+$, which yields the minimum error probability [31]:

$$P_{e,min} = \frac{1}{2}(1 - \mathrm{Tr}\left[|p_1\rho_1 - p_0\rho_0|\right])\,. \tag{1.194}$$

This discussion was done in the single copy scenario. If one allows more copies of $\rho_{0/1}$ to be prepared at the same time with the probability $p_{0/1}$, one can again infer that the optimal error probability is given by:

$$P_{e,min,n} = \frac{1}{2}(1 - \mathrm{Tr}\left[|p_1\rho_1^{\otimes n} - p_0\rho_0^{\otimes n}|\right])\,. \tag{1.195}$$

Differently from what happened for the single copy scenario, though, this probability scales with n, and in particular it asymptotically decrease as $P_{e,min,n} \simeq e^{-\xi_{QCB}n}$ for $n \gg 1$. In [31] it was proven that the exponent takes the form:

$$\xi_{QCB} := -\lim_{n\to\infty}\frac{\log P_{e,min,n}}{n} = \max_{0\leq s\leq 1}\left(-\log\mathrm{Tr}\left[\rho_0^s\rho_1^{1-s}\right]\right). \tag{1.196}$$

This result goes under the name of Chernoff bound.

The position at which the maximum is found usually depends on the particular form of ρ_0 and ρ_1. Still, if one restricts to the case in which

$\rho_1 = \rho_0 + \delta\rho$, with $\delta\rho \ll 1$, then one can apply the methods from Sec. 1.4.4 to express ξ_{QCB} in terms of $\mathbb{J}_\alpha^{-1}$. Then, the maximum in the family is attained for $s = \frac{1}{2}$, meaning that:

$$\xi_{QCB} = \frac{1}{8}\mathrm{Tr}\left[\delta\rho\, \mathbb{J}_{WY}^{-1}\big|_{\rho_0}[\delta\rho]\right] . \tag{1.197}$$

This further motivates the interest in the Wigner-Yanase metric.

1.4.6 The relative entropy ($\alpha = 0$)

The most renowned among the α-divergences, and among the contrast functions in general, is the one obtained in the limit $\alpha \to 0$, namely the relative entropy. In fact, carrying out the limit of Eq. (1.180) one gets:

$$g_L(x) := \lim_{\alpha\to 0} \frac{1}{\alpha(\alpha-1)} \int_0^\alpha \mathrm{d}\beta\, (x^\beta \log x) = -\log x\,. \tag{1.198}$$

The corresponding contrast function takes the familiar form:

$$S(\rho||\sigma) := H_L(\rho||\sigma) = \mathrm{Tr}\left[\rho\,(\log\rho - \log\sigma)\right] . \tag{1.199}$$

Moreover, its symmetrised version has the following integral expression:

$$\frac{g_L(x) + x\, g_L(x^{-1})}{2} = \int_0^1 \mathrm{d}N_L(s)\, \left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right) , \tag{1.200}$$

where $\mathrm{d}N_L(s) := \frac{1}{2(1+s)}\mathrm{d}s$. This divergence not only is very relevant in the context of quantum information, but it is also of key importance in the context of resource theory and thermodynamics. It is related to the von Neumann entropy by the equation:

$$S(\rho||\mathbb{1}) = -S(\rho)\,, \tag{1.201}$$

and it has the special property that it is additive on tensor products:

$$S(\rho_A \otimes \rho_B||\sigma_A \otimes \sigma_B) = S(\rho_A||\sigma_A) + S(\rho_B||\sigma_B)\,. \tag{1.202}$$

Thanks to this, one can express the mutual information in a state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ as:

$$I(A:B) := S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = \tag{1.203}$$
$$= \mathrm{Tr}\left[\rho_{AB}\log\rho_{AB}\right] - \mathrm{Tr}\left[\rho_A\log\rho_A\right] - \mathrm{Tr}\left[\rho_B\log\rho_B\right] = \tag{1.204}$$
$$= \mathrm{Tr}\left[\rho_{AB}\log\rho_{AB}\right] - \mathrm{Tr}\left[\rho_{AB}\log\rho_A\otimes\rho_B\right] = S(\rho_{AB}||\rho_A\otimes\rho_B)\,, \tag{1.205}$$

where $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$ (and similarly for B) and in the last line we used the additivity of the tensor products when taking the logarithm of positive states. Eq. (1.205) shows that the mutual information is directly connected to the statistical distance of the state ρ_{AB} from a product state locally indistinguishable from it (i.e., $\rho_A \otimes \rho_B$).

As it was mentioned above, the relative entropy also plays a special role in quantum thermodynamics, as it is used in the definition of the relative entropy out of equilibrium:

$$F(\rho, H) := \mathrm{Tr}[\rho H] - \beta^{-1} S(\rho)\,. \tag{1.206}$$

Indeed, this can be transformed as follows:

$$\begin{aligned} F(\rho, H) &= \beta^{-1}(-\mathrm{Tr}\left[\rho \log e^{-\beta H}\right] + \mathrm{Tr}[\rho \log \rho]) = && (1.207)\\ &= -\beta^{-1} \log \mathcal{Z}_H + \beta^{-1}\,\mathrm{Tr}\left[\rho\left(\log\rho - \log\frac{e^{-\beta H}}{\mathcal{Z}_H}\right)\right] = && (1.208)\\ &= F_{\mathrm{eq}}(\pi_\beta(H)) + \beta^{-1}\, S(\rho||\pi_\beta(H))\,, && (1.209) \end{aligned}$$

where we denoted the thermal state by $\pi_\beta(H) := \frac{e^{-\beta H}}{\mathcal{Z}_H}$, and we defined the partition function $\mathcal{Z}_H$ to be $\mathcal{Z}_H := \mathrm{Tr}\left[e^{-\beta H}\right]$. Finally, in the last line, we defined the free energy of equilibrium in the usual way, i.e., $F_{\mathrm{eq}}(\pi_\beta(H)) := -\beta^{-1}\log \mathcal{Z}_H$. Again we see that the excess free energy with respect to the equilibrium one is given by the statistical distance from the Gibbs state, as measured by the relative entropy.

Going back to its mathematical characterisation, it should be noticed that the corresponding standard monotone is given by:

$$f_L(x) = \frac{x-1}{\log x} = \int_0^1 \mathrm{d}\alpha\; x^\alpha\,, \tag{1.210}$$

i.e., it corresponds to $f_>(x)$ in Eq. (1.150) for a uniform measure. This integral expression is particularly useful because it allows the immediate calculation of $\mathbb{J}_L$ as:

$$\mathbb{J}_L\big|_\pi[A] = \int_0^1 \mathrm{d}\alpha\; \pi^{1-\alpha} A\, \pi^\alpha\,. \tag{1.211}$$

Interestingly, this superoperator is the same one gets from the Dyson series of the exponential, i.e., $e^{\log(\pi)+\varepsilon A} \simeq \pi + \varepsilon\, \mathbb{J}_L[A]$. For this reason, one can deduce that $\mathbb{J}_L^{-1}[A]$ will be given by the first term in the expansion

of $\log(\pi + \varepsilon A)$. As we saw in the expansion of the α-contrast functions, this reads:

$$\mathbb{J}_L^{-1}\big|_\pi[A] = \int_0^\infty \mathrm{d}t \ (\pi + t)^{-1} A \, (\pi + t)^{-1} \, . \tag{1.212}$$

Notice that the same conclusion could be reached by directly expanding Eq. (1.199). Since $\mathbb{J}_L^{-1}$ is in Kraus form, it follows that $f_L \in \mathcal{F}^-$. This could be surprising as f_L is obtained from the convex combination of elements of $\mathcal{F}^-$, which is not a convex set [21].

Another useful property of the operator $\mathbb{J}_L^{-1}$ is that it allows us to express the symmetrised relative entropy in the form of an integral. This reads [32]:

$$H_L(\rho||\sigma) + H_L(\sigma||\rho) = \mathrm{Tr}\left[(\rho - \sigma)\left(\log\rho - \log\sigma\right)\right] = \tag{1.213}$$

$$= \int_0^1 \mathrm{d}t \ \mathrm{Tr}\left[(\rho - \sigma)\, \mathbb{J}_L^{-1}\big|_{t\rho + (1-t)\sigma}[(\rho - \sigma)]\right] \, . \tag{1.214}$$

This expression is particularly useful when one needs to compare the global and local behaviour of the relative entropy: for example, in [32] it was used to prove the equality of the contraction coefficients for the symmetrised relative entropy with the one for the corresponding Fisher information (defined in Eq. (1.304) and Eq. (1.306)).

The scalar product defined by $\mathbb{J}_L$ on the space of observables is called Kubo-Mori-Bogoliubov (KMB) inner product. It is particularly relevant in the context of linear response theory due to the following identity:

$$\frac{\partial^2}{\partial x \partial y} \mathrm{Tr}\left[e^{H + xA + yB}\right]\big|_0 = \mathrm{Tr}\left[A \ \left(\mathbb{J}_L\big|_{e^H}\right)[B]\right] \tag{1.215}$$

which connects the KMB inner product to the expansion of the partition function associated with the Hamiltonian H. This quantity naturally appears when one wants to estimate the fluctuations around thermal equilibrium in the linear response regime [33]. Starting from this interpretation we can provide a Cramer-Rao bound of the type in Eq. (1.142). Indeed, suppose one has a Hamiltonian depending on some parameters θ. The corresponding thermal state is given by $\pi_\beta(H(\theta)) := \frac{e^{-\beta H(\theta)}}{\mathcal{Z}_{H(\theta)}}$, and we can use Eq. (1.215) to derive that:

$$\frac{1}{\mathcal{Z}_{H(0)}} \frac{\partial^2}{\partial x^2} \mathrm{Tr}\left[e^{-\beta H(0) + xA}\right]\Big|_{x=0} = \mathrm{Tr}\left[A \ \left(\mathbb{J}_L\big|_{\pi_\beta(H(0))}\right)[A]\right] \, , \tag{1.216}$$

which provides an operational procedure of accessing the generalised variance in Eq. (1.142), namely through the linear response of the thermal state at equilibrium when one perturbs the Hamiltonian by $\beta^{-1}A$. Suppose now that A is a locally unbiased estimator, i.e., that it satisfies:

$$\frac{\partial}{\partial\theta}\,\mathrm{Tr}\left[\pi_\beta(H(\theta))\,A\right]\Big|_{\theta=0} = 1\,. \tag{1.217}$$

Then, from Eq. (1.142) we obtain that:

$$\frac{1}{\mathcal{Z}_{H(0)}}\frac{\partial^2}{\partial x^2}\mathrm{Tr}\left[e^{-\beta H(0)+xA}\right]\Big|_{x=0} \geq \tag{1.218}$$

$$\geq \frac{1}{\mathrm{Tr}\left[\partial_\theta\pi_\beta(H(\theta))\,\mathbb{J}_L^{-1}\big|_{\pi_\beta(H(0))}[\partial_\theta\pi_\beta(H(\theta))]\right]\Big|_{\theta=0}} = \tag{1.219}$$

$$= \frac{1}{\beta^2\,\mathrm{Tr}\left[\Delta\partial_\theta H(\theta)\,\mathbb{J}_L\big|_{\pi_\beta(H(0))}[\Delta\partial_\theta H(\theta)]\right]\Big|_{\theta=0}}\,, \tag{1.220}$$

where in the last line we used the notation $\Delta A := A - \langle A\rangle_{\pi_\beta(H(0))}$ arising from the expansion of the thermal state $\pi_\beta(H(\varepsilon)) = \pi_\beta(H(0)) - \varepsilon\,\beta\,\mathbb{J}_L[\Delta\partial_\theta H(\theta)]$ (see Eq. (4.26) for the details of the calculation). In order to show the relevance of this last formula, suppose one wants to estimate the temperature of a thermal state through the measurement of the observable A. In order for the estimator to be locally unbiased, we require that:

$$\frac{\partial}{\partial\beta}\,\mathrm{Tr}\left[\pi_\beta(H)\,A\right]\Big|_{\beta=\beta_0} = 1 \tag{1.221}$$

where β_0 is the temperature of the system around which we want our thermometer to work. Then, it should be noticed that the denominator of Eq. (1.220) can be connected to the heat capacity as follows:

$$C_H(\beta_0) := \frac{\partial}{\partial T}\mathrm{Tr}\left[H\,\pi_\beta(H)\right]\Big|_{\beta=\beta_0} = \frac{\partial\beta}{\partial T}\,\frac{1}{\partial\beta}\mathrm{Tr}\left[H\,\pi_\beta(H)\right]\Big|_{\beta=\beta_0} = \tag{1.222}$$

$$= \beta^2\,\mathrm{Tr}\left[\Delta H\,\mathbb{J}_L\big|_{\pi_\beta(H)}[\Delta H]\right]\Big|_{\beta=\beta_0} \tag{1.223}$$

where T is the temperature and we used the definition $\beta := 1/T$. Hence, wrapping everything together we see that the fluctuations in the measurement of the temperature are bounded as:

$$\frac{1}{\mathcal{Z}_H}\frac{\partial^2}{\partial x^2}\mathrm{Tr}\left[e^{-\beta_0 H + xA}\right]\Big|_{x=0} \geq \frac{1}{C_H(\beta_0)}\,, \tag{1.224}$$

result that can be interpreted as saying that if one wants to get a good estimation of the temperature the best strategy is to maximise the heat capacity of the thermometer. The discussion above shows how the Cramer-Rao bound can be interesting even when applied to quantum Fisher informations other than the Bures one.

1.4.7 The quantum information variance

The function:

$$g_V(x) = \frac{1}{2}(\log x)^2\,. \tag{1.225}$$

defines what is called the quantum information variance, given by [34]:

$$H_V(\rho||\sigma) = \frac{1}{2}\,\mathrm{Tr}\left[\rho\,(\log\rho - \log\sigma)^2\right]\,. \tag{1.226}$$

This quantity is not part of the family of α-divergences, but it is related to them by the identity:

$$H_V(\rho||\sigma) = \frac{1}{2}\frac{\partial^2}{\partial\alpha^2}\left(\alpha(\alpha-1)H_\alpha(\rho||\sigma)\right)\Big|_{\alpha=0}\,. \tag{1.227}$$

Hence, it is straightforward to verify, for example, that:

$$g_V(x) = \int_0^\infty \mathrm{d}s\;\frac{(-\log s)}{(1+s)^2}\left(\frac{(x-1)^2}{x+s}\right)\,. \tag{1.228}$$

The corresponding monotone function is given by:

$$f_V(x) = \frac{2\,(x-1)^2}{(x+1)(\log x)^2} \tag{1.229}$$

Thanks to the relation in Eq. (1.227), we can also deduce from Eq. (1.186) that the Fisher information takes the form:

$$\mathbb{J}_V^{-1}\big|_\pi[A] = \frac{1}{2}\,\mathbb{J}_L^{-2}\left[\{\pi, A\}\right]\,. \tag{1.230}$$

Interestingly, one could arrive to the same result from the relation $f_V(x) = f_L(x)^2/f_B(x)$ (defined in Eq. (1.210) and Eq. (1.125) respectively). This identification also allows us to express $\mathbb{J}_V$ in terms of $\mathbb{J}_L$ and $\mathbb{J}_B^{-1}$, namely:

$$\mathbb{J}_V\big|_\pi[A] = \int_0^\infty \mathrm{d}t\ e^{-t\pi/2}\,\mathbb{J}_L^2[A]\,e^{-t\pi/2}\,. \tag{1.231}$$

The quantum information variance can be interpreted as a quantifier of the fluctuations in the distinguishability of ρ with respect to σ [34]. For this reason, it also appears in the context of the statistics of work as the natural quantifier of fluctuations (see Sec. 4.4).

1.4.8 The geometric mean

Consider the function:

$$g_{\sqrt{x}}(x) = \sqrt{x^{-1}} - \sqrt{x}\,, \tag{1.232}$$

which corresponds to the contrast function:

$$H_{\sqrt{x}}(\rho||\sigma) = \mathrm{Tr}\left[\sqrt{\rho}\,(\rho-\sigma)\sqrt{\sigma^{-1}}\right]\,. \tag{1.233}$$

Its symmetrised version has the following integral expression:

$$g^{\mathrm{symm}}_{\sqrt{x}}(x) = \frac{(x-1)^2}{2\sqrt{x}} = \int_0^1 \mathrm{d}s\ \frac{1}{2\pi\sqrt{s}}\left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right)\,, \tag{1.234}$$

so that we can identify $\mathrm{d}N_g(s) := \frac{1}{2\pi\sqrt{s}}$. Despite the fact the expression of $H_{\sqrt{x}}(\rho||\sigma)$ might look unfamiliar, the associated standard monotone takes a particularly simple form. In fact, this corresponds to:

$$f_{\sqrt{x}}(x) = \sqrt{x}\,. \tag{1.235}$$

In this case it is quite easy to compute the quantum Fisher information operators, which are given by:

$$\mathbb{J}_{\sqrt{x}}\big|_\pi[A] = \sqrt{\pi}\,A\,\sqrt{\pi}\,; \qquad\qquad \mathbb{J}^{-1}_{\sqrt{x}}\big|_\pi[A] = \sqrt{\pi^{-1}}\,A\,\sqrt{\pi^{-1}}\,. \tag{1.236}$$

As it can be verified by inspection, both $\mathbb{J}_{\sqrt{x}}$ and $\mathbb{J}^{-1}_{\sqrt{x}}$ are CP, as they are presented in Kraus form. Indeed, $f_{\sqrt{x}}$ is the only element in the intersection of $\mathcal{F}^+$ and $\mathcal{F}^-$. Moreover, it can actually be used to define these sets thanks to the order relation $\preceq$ defined above. In this sense, the

square root can be regarded as the middle point in the space of standard monotone functions. It should also be noted that $f_{\sqrt{x}}$ is also the only fixed point of the transformation T. The peculiar property of having both $\mathbb{J}_{\sqrt{x}}$ and $\mathbb{J}_{\sqrt{x}}^{-1}$ CP will be used in the definition of the Petz' recovery map (see Sec. 3.3).

1.4.9 The harmonic mean ($\alpha = 2$)

The last function we consider is again in the family of α-divergences, and it is given by the limit $\alpha \to 2$, for which we obtain:

$$g_2(x) = \frac{x^2 - 1}{2}\,. \tag{1.237}$$

A more standard definition of the same divergence is obtained by using the function:

$$g_H(x) = \frac{(x-1)^2}{2}\,. \tag{1.238}$$

In fact, it should be noticed that the two only differ by a linear term, namely $g_2(x) - g_H(x) = -2(x-1)$, so we indeed have that $H_{g_2}(\rho||\sigma) \equiv H_{g_H}(\rho||\sigma)$. The corresponding contrast function is then given by:

$$H_H(\rho||\sigma) = \frac{1}{2}\,\mathrm{Tr}\left[(\rho-\sigma)\rho^{-1}(\rho-\sigma)\right] = \frac{1}{2}\left(\mathrm{Tr}\left[\sigma^2\rho^{-1}\right] - 1\right)\,, \tag{1.239}$$

while the measure associated to its symmetrised version reads:

$$g_H(x) + x g_H(x^{-1}) = \frac{(x-1)^2}{2} + \frac{(x-1)^2}{2x} = \tag{1.240}$$

$$= \int_0^1 \mathrm{d}N_H(s)\,\left(\frac{(x-1)^2}{x+s} + \frac{(x-1)^2}{1+sx}\right)\,, \tag{1.241}$$

with $\mathrm{d}N_H(s) = \delta(s)/2$. Finally, the standard monotone function associated to it is given by:

$$f_H(x) = \frac{2x}{x+1}\,. \tag{1.242}$$

This is the minimal standard monotone function, also corresponding to the image of the Bures function through T, i.e., $f_H = T f_B$. Thanks to this relation we can use Eq. (1.86) to directly obtain:

$$\mathbb{J}_H\big|_\pi[A] = \int_0^\infty \mathrm{d}t\;\, e^{-t\pi^{-1}/2}\,A\,e^{-t\pi^{-1}/2}\,; \qquad \mathbb{J}_H^{-1}\big|_\pi[A] = \frac{1}{2}\{\pi^{-1}, A\}\,. \tag{1.243}$$

This could also be deduced from the fact that $f_H = f_{0_<}$ (in the notation of Eq. (1.156)), so the formula above could be directly inferred from Eq. (1.157) and Eq. (1.158).

This concludes the survey of the quantum Fisher information metrics.

1.5 The dynamical properties of Fisher information

In the treatment of Fisher information most of the focus goes into its significance as a distinguishability quantifier, and in the many different results linking it to estimation theory, as in the case of the Cramer-Rao bound and of the Chernoff bound (see Eq. (1.140) and Eq. (1.196)). At the same time, Fisher information metrics satisfy the important property of being the only class that monotonically decreases under the action of any CPTP map. Hence, there is a dual manner of defining Fisher information, one statistical and one dynamical, and it is not clear how the two should be connected. In fact, it is not even obvious that these two properties should define the same class of metrics. For this reason, even if the statistical aspects of the Fisher information are quite well studied, this is not very informative about its dynamical nature.

Indeed, not only the Fisher information can be defined in terms of its contractivity under physical evolutions, but also many properties of the dynamical maps can be given solely in terms of their behaviour with respect to the Fisher information. The first result in this direction is given by the following [1]:

Theorem 5. *Consider a trace and adjoint preserving linear map* $\Phi : \mathcal{M}_d(\mathbb{C}) \to \mathcal{M}_d(\mathbb{C})$*, where* $\mathcal{M}_d(\mathbb{C})$ *is the space of* $d \times d$ *complex matrices. Define* $\mathcal{S} \subset \mathcal{M}_d(\mathbb{C})$ *to be the set of positive semidefinite, trace one matrices, and define* $\Phi_{\text{anc}} := \mathbb{I}_d \otimes \Phi$*. If* Φ_{anc} *satisfies the following three properties:*

1. Φ *is full-rank;*
2. Φ_{anc} *maps at least one point from the interior of* $\mathcal{S} \otimes \mathcal{S}$ *into* $\mathcal{S} \otimes \mathcal{S}$*;*
3. *for any two states* ρ *and* $\rho + \delta\rho$ *in* $\Phi_{\text{anc}}^{-1}(\mathcal{S} \otimes \mathcal{S} \cap \Phi_{\text{anc}}(\mathcal{S} \otimes \mathcal{S}))$ *with* $|\delta\rho| \ll 1$*, for any contrast function* H_g*, one has:*

$$H_g(\rho||\rho + \delta\rho) \geq H_g(\Phi_{\text{anc}}(\rho)||\Phi_{\text{anc}}(\rho + \delta\rho))\,; \tag{1.244}$$

then, the original map Φ *is CP.*

This theorem shows that the property of being completely positive can be defined exclusively in statistical terms. Indeed, this characterisation could be regarded as the dual of the Chentsov/Petz theorem: not only the Fisher information is the unique family of metrics that contracts under arbitrary CP maps, but also CP maps are the only set of maps that contract the Fisher information. Without the existence of Thm. 5, one could refer to the Fisher information is a derived concept, as the Chentsov/Petz theorem tells us that it can be derived from the much more fundamental notion of physical evolutions. Instead, the main contribution of Thm. 5 is to show that it is at least as fundamental as the concept of complete positivity, since the latter can be defined in terms of the first. Indeed, one could think of an information theoretic reconstruction of quantum mechanics that defines the allowed evolutions in terms of the Fisher information alone. This is still to hypothetical to be made precise, but it shows the importance in principle of the result just stated.

Proof. We prove the theorem by contradiction: suppose there exists a map Φ that is not CP, but that satisfies Eq. (1.244). Thanks to condition 2 there exists at least one point π in the interior of $\mathcal{S}\otimes\mathcal{S}$ such that $\Phi_{\text{anc}}(\pi)$ is a state. Moreover, from the assumption that Φ is not CP there is also at least one state σ such that $\Phi_{\text{anc}}(\sigma)\notin\mathcal{S}\otimes\mathcal{S}$. Without loss of generality one can choose σ to be in the interior of $\mathcal{S}\otimes\mathcal{S}$: if this is not the case, one can take an ε-ball around σ whose image still lays outside of the state space, and by inspecting the intersection between its preimage and $\mathcal{S}\otimes\mathcal{S}$ one can find a point satisfying the assumption. Hence, the line $\rho_\lambda := (1-\lambda)\,\pi+\lambda\,\sigma$ also lays in the interior of the state space. From this, it follows that the following superior is finite:

$$\sup_{\lambda,\mathrm{Tr}[\delta\rho^2]=1} H_g(\rho_\lambda||\rho_\lambda+\varepsilon\,\delta\rho)\simeq \tag{1.245}$$

$$\simeq \sup_{\lambda,\mathrm{Tr}[\delta\rho^2]=1}\frac{\varepsilon^2}{2}\left(\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_f^{-1}\big|_{\rho_\lambda}[\delta\rho]\right]\right)<\infty\,, \tag{1.246}$$

where $\varepsilon\ll 1$, and we used the fact that the Fisher information is a bounded operator in the interior of the state space.

Since by varying λ, $\Phi_{\text{anc}}(\rho_\lambda)$ interpolates between the positive definite matrix $\Phi_{\text{anc}}(\pi)$ and one with at least one negative eigenvalue, namely $\Phi_{\text{anc}}(\sigma)$, there exists a λ^* such that $\Phi_{\text{anc}}(\rho_{\lambda^*})$ is a state with at least one zero-eigenvalue. We are now ready to prove the claim. Set the state ρ_η

to be given by $\rho_\eta := \rho_{(\lambda^* - \eta)}$, where $\eta \ll 1$ is chosen so that the smallest eigenvalue of $\Phi(\rho_\eta)$ is of order η. Moreover, consider a perturbation $\delta\rho_\eta$ such that $[\Phi_{\rm anc}(\rho_\eta), \Phi_{\rm anc}(\delta\rho_\eta)] = 0$, and having a positive finite contribution along the eigenvectors corresponding to η-eigenvalues (the positivity condition ensures that in the limit $\eta \to 0$ the perturbed state is still in the interior of $\mathcal{S}$ for any finite η). Since Φ is full rank one can always find such an element. Then, the evolved contrast function reads:

$$H_g(\Phi_{\rm anc}(\rho_\eta)||\Phi_{\rm anc}(\rho_\eta + \varepsilon\delta\rho_\eta)) \simeq \frac{\varepsilon^2}{2}\sum_{i=1}^{d^2}\frac{(\delta\tilde{\rho}_\eta)_i^2}{(\tilde{\rho}_\eta)_i}\,, \tag{1.247}$$

which scales as $1/\eta$ as $\eta \to 0$. Hence, we can always find a η small enough such that:

$$H_g(\Phi_{\rm anc}(\rho_\eta)||\Phi_{\rm anc}(\rho_\eta + \varepsilon\delta\rho_\eta)) > \sup_{\lambda, {\rm Tr}[\delta\rho^2]=1} H_g(\rho_\lambda||\rho_\lambda + \delta\rho) \geq \tag{1.248}$$

$$\geq H_g(\rho_\eta||\rho_\eta + \varepsilon\delta\rho_\eta)\,, \tag{1.249}$$

contradicting the assumption that $\Phi_{\rm anc}$ is contracting for any two points in the interior of the space of states (condition 3). This proves the claim. □

The connection between the Fisher information and the dynamics of quantum systems is actually even deeper: not only one can express the CP-ness of a map in terms of its contractivity property with respect to the Fisher information, but also, with the same principle, one can define Markovianity of a semigroup of maps in the same way. Moreover, we show in the following that self-adjointness of an operator with respect to Fisher information metric corresponds to a detailed balanced dynamics. These topics are the argument of the next subsections, and corroborate the interpretation of the Fisher information as an intimately dynamical quantity.

1.5.1 Characterisation of Markovian evolutions

Before starting to investigate the dynamical character of Fisher information, we need to introduce the concept of dynamical semigroup. To this end, consider a family of CP-maps Φ_t depending smoothly on t, such that for any two times t and s ($t \geq s$) one can define an intermediate map $\Phi_{t,s}$ satisfying the property $\Phi_t = \Phi_{t,s} \circ \Phi_s$. This kind of maps are called divisible.

Divisible dynamics Φ_t are naturally endowed with the structure of semigroup (with identity), hence the name of dynamical semigroup. Indeed, define $\{\Phi_{t,s}\}_{t\geq s}$ to be the set of all intermediate maps. Since $\Phi_t = \Phi_{t,t}\circ\Phi_t = \mathbb{1}\circ\Phi_t$, the identity is contained in the set; moreover, the composition rule $\Phi_{t,s}\circ\Phi_{s,r} = \Phi_{t,r}$ satisfies associativity. For this reason, the study of divisible evolutions coincides with the one of semigroups of maps. In particular, if all the intermediate maps are CP, the corresponding semigroup is called CP-divisible. This is the most canonical notion of Markovianity [35], and the one that will be used in this text. It should be noticed that for CP-divisible maps all the contrast functions monotonically decrease, as it can be verified from their derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}H_g(\Phi_t(\rho)||\Phi_t(\sigma)) = \tag{1.250}$$

$$= \lim_{\varepsilon\to 0}\frac{H_g(\Phi_{t+\varepsilon}(\rho)||\Phi_{t+\varepsilon}(\sigma)) - H_g(\Phi_t(\rho)||\Phi_t(\sigma))}{\varepsilon} = \tag{1.251}$$

$$= \lim_{\varepsilon\to 0}\frac{H_g(\Phi_{t+\varepsilon,t}\Phi_t(\rho)||\Phi_{t+\varepsilon,t}\Phi_t(\sigma)) - H_g(\Phi_t(\rho)||\Phi_t(\sigma))}{\varepsilon} \leq \tag{1.252}$$

$$\leq \lim_{\varepsilon\to 0}\frac{H_g(\Phi_t(\rho)||\Phi_t(\sigma)) - H_g(\Phi_t(\rho)||\Phi_t(\sigma))}{\varepsilon} = 0\,, \tag{1.253}$$

where in Eq. (1.252) one uses the contractivity of the contrast functions under the action of quantum channels. Since ρ and σ are arbitrary, the same result also holds for quantum Fisher metrics, and this behaviour is referred to as monotonic decrease of information, which also justifies the identification of CP-divisibility with Markovian dynamics.

As we said above, a semigroup of CP-maps is Markovian without having to put extra constraint on the global structure of the semigroup, as the composition of infinitesimal CP-maps automatically gives a CP-map. On the other hand, a semigroup containing non-CP elements is non-Markovian only if the extra constraint that $\Phi_{t,0}$ is CP for all times is satisfied. This condition is needed to ensure that the global evolution is actually physically realisable.

Since $\Phi_{t,s}$ is defined for any $t \geq s$, one can translate the study of the semigroup to the one of their generators, defined as:

$$\mathcal{L}_t := \lim_{\varepsilon\to 0}\frac{\Phi_{t+\varepsilon,t} - \mathbb{I}}{\varepsilon}\,. \tag{1.254}$$

In the following we present a derivation of the Lindbladian expression of the generator, which will shed light on its connection to CP-divisibility.

First, it is useful to study what kind of constraints complete positivity imposes on the expression of general maps. To this end, consider the maximally entangled state in d dimensions, $|\Omega\rangle := \frac{1}{\sqrt{d}} \sum |i\rangle_A \otimes |i\rangle_B$. The Choi state associated to a map Φ is defined as:

$$\mathcal{C}^{\Phi} := (\mathbb{I}_A \otimes \Phi)[|\Omega\rangle\langle\Omega|] = \frac{1}{d} \sum_{i,j} \Phi_{l|j}^{k|i} \; |i\,k\rangle\langle j\,l| \;, \tag{1.255}$$

where we use the notation $\Phi_{l|j}^{k|i} := \langle k| \Phi(|i\rangle\langle j|) |l\rangle$. A fundamental theorem by Choi states that the map Φ is CP if and only if the corresponding Choi state $\mathcal{C}^{\Phi}$ is positive semidefinite [36]. This characterisation can be used to express the action of CP-maps on states. Given an arbitrary state ρ, one can rewrite $\Phi(\rho)$ in terms of the Choi matrix as:

$$\Phi(\rho) = \sum_{i,j} \rho_{i,j} \, \Phi_{l|j}^{k|i} \, |k\rangle\langle l| = \sum_{i,j} \mathcal{C}_{ik,jl}^{\Phi} \, E_{i,k}^{\dagger} \, \rho \, E_{j,l} \,, \tag{1.256}$$

where we defined $E_{i,j} := |i\rangle\langle j|$. If we now choose another orthonormal basis for the space of bounded operators $F_\alpha := \sum_{i,j} E_{i,j}^{\dagger} \bar{U}_{i,j,\alpha}$ (where α is an index ranging from 1 to d^2 and $U_{\alpha,i,j}$ a unitary matrix), it is easy to see that:

$$\Phi(\rho) = \sum_{\alpha,\beta} c_{\alpha,\beta} \, F_\alpha \, \rho \, F_\beta^{\dagger} \,, \tag{1.257}$$

where $c_{\alpha,\beta} = \sum_{i,j,k,l} U_{\alpha,i,k} \, \mathcal{C}_{ik,jl}^{\Phi} \bar{U}_{j,l,\beta}$ is a positive matrix. Hence, any CP-map admits an expression as in Eq. (1.257) where $c_{\alpha,\beta}$ is required to be positive semidefinite.

Fix now the first element of the basis F_α to be the identity operator. By ignoring the first column and the first row of the corresponding $c_{\alpha,\beta}$ one obtains what is called the reduced GKS (Gorini-Kossakowski-Sudarshan) matrix for Φ, which we denote by $\tilde{c}_{\alpha,\beta}$ [37]. This quantity is particularly useful when talking about generators thanks to the following:

Theorem 6. *Let $\Phi_t := e^{t\mathcal{L}}$, and denote by $\tilde{c}_{\alpha,\beta}(\mathcal{L})$ the reduced GKS matrix associated to $\mathcal{L}$. Φ_t is completely positive if and only if $\tilde{c}_{\alpha,\beta}(\mathcal{L})$ is positive semidefinite.*

Proof. If Φ_t is completely positive, then the corresponding reduced GKS matrix $\tilde{c}_{\alpha,\beta}(\Phi_t)$ is positive semidefinite. Then, one can obtain $\tilde{c}_{\alpha,\beta}(\mathcal{L})$ from the limit:

$$\tilde{c}_{\alpha,\beta}(\mathcal{L}) = \lim_{\varepsilon \to 0} \frac{\tilde{c}_{\alpha,\beta}(\Phi_\varepsilon) - \tilde{c}_{\alpha,\beta}(\mathbb{I})}{\varepsilon} \,, \tag{1.258}$$

where we implicitly defined the reduced GKS matrix of the identity superoperator. By construction this is identically zero, which implies that $\tilde{c}_{\alpha,\beta}(\mathcal{L})$ is positive semidefinite.

In order to prove the opposite direction, notice that for t sufficiently small one has:

$$\tilde{c}_{\alpha,\beta}(\Phi_t) = \tilde{c}_{\alpha,\beta}(\mathbb{I}) + t\,\tilde{c}_{\alpha,\beta}(\mathcal{L}) + \mathcal{O}\left(t^2\right) . \tag{1.259}$$

Again, since the contribution coming from the identity cancels, if $\tilde{c}_{\alpha,\beta}(\mathcal{L})$ is positive semidefinite, the same holds for $\tilde{c}_{\alpha,\beta}(\Phi_t)$. This concludes the proof. □

The discussion above takes care of the first condition for a map to be a quantum channel, namely complete positivity. Additionally, in order to conserve the total probability, one also needs to require that the map is trace preserving. A particularly compact expression for this is obtained by rewriting the requirement of being trace preserving explicitly:

$$\mathrm{Tr}\left[\Phi(\rho)\right] = \mathrm{Tr}\left[\Phi^\dagger[\mathbb{1}]\,\rho\right] = \mathrm{Tr}\left[\rho\right] . \tag{1.260}$$

Since this has to hold for any ρ, one can conclude that $\Phi^\dagger[\mathbb{1}] = \mathbb{1}$. The same condition for generators becomes $\mathcal{L}^\dagger[\mathbb{1}] = 0$, as it can be verified directly from Eq. (1.254). This kind of map admits a particularly canonical expression:

Theorem 7. *Let $\mathcal{L}$ be a Hermitian preserving (i.e., $\mathcal{L}[A^\dagger] = \mathcal{L}[A]^\dagger$) superoperator satisfying $\mathcal{L}^\dagger[\mathbb{1}] = 0$. Take an orthonormal basis with respect to the Hilbert-Schmidt scalar product $\{F_\alpha\}_{\alpha\in\{1,\dots,d^2\}}$ such that $F_1 = \frac{\mathbb{1}}{\sqrt{d}}$. Then, one can express $\mathcal{L}$ as:*

$$\mathcal{L}(\rho) = -i[H,\rho] + \sum_{\alpha,\beta=2}^{d^2} \tilde{c}_{\alpha,\beta}\left(F_\alpha\,\rho\,F_\beta^\dagger - \frac{1}{2}\{F_\beta^\dagger F_\alpha, \rho\}\right) , \tag{1.261}$$

where $\tilde{c}_{\alpha,\beta}$ is the reduced GKS matrix of $\mathcal{L}$ and H is a self-adjoint, traceless operator given by:

$$H = \frac{1}{2i\sqrt{d}} \sum_{\beta=1}^{d^2} c_{1,\beta} F_\beta^\dagger - c_{\beta,1} F_\beta \, . \tag{1.262}$$

Proof. A general Hermitian preserving superoperator can be rewritten as in Eq. (1.257), where the matrix $c_{\alpha,\beta}$ is self-adjoint, but not necessarily positive. Using this expression, and isolating the terms involving the identity we obtain:

$$\mathcal{L}(\rho) = \sum_{\alpha,\beta} c_{\alpha,\beta}\, F_\alpha\, \rho\, F_\beta^\dagger = \tag{1.263}$$

$$= \frac{c_{1,1}}{d}\,\rho + \frac{1}{\sqrt{d}} \sum_{\beta=2}^{d^2} \left(c_{\beta,1}\, F_\beta\, \rho + c_{1,\beta}\, \rho\, F_\beta^\dagger \right) + \sum_{\alpha,\beta=2}^{d^2} \tilde{c}_{\alpha,\beta}\, F_\alpha\, \rho\, F_\beta^\dagger = \tag{1.264}$$

$$= G\,\rho + \rho\, G^\dagger + \sum_{\alpha,\beta=2}^{d^2} \tilde{c}_{\alpha,\beta}\, F_\alpha\, \rho\, F_\beta^\dagger\,, \tag{1.265}$$

where we defined the operator $G := \frac{1}{2\,d} c_{1,1} + \frac{1}{\sqrt{d}} \sum_{\beta=2}^{d^2}\, c_{\beta,1}\, F_\beta$. Define the real and imaginary parts of G as $K = (G + G^\dagger)/2$ and $H = (G^\dagger - G)/(2i)$, so that $G = K - iH$. Both matrices are self-adjoint. Moreover, notice that H can be explicitly expressed as:

$$H = \frac{G^\dagger - G}{2i} = \frac{1}{2i\sqrt{d}} \sum_{\beta=1}^{d^2}\, c_{1,\beta} F_\beta^\dagger - c_{\beta,1} F_\beta\,, \tag{1.266}$$

where we implicitly make use of the fact that $c_{\alpha,\beta}$ is self-adjoint. Since the operators $\{F_\alpha\}$ are part of an orthonormal basis whose first element is the identity, F_α is traceless whenever $\alpha \neq 1$. This proves that H is also traceless.

Rewriting Eq. (1.265) using the matrices H and K we obtain:

$$\mathcal{L}(\rho) = -i[H, \rho] + K\,\rho + \rho\, K + \sum_{\alpha,\beta=2}^{d^2}\, \tilde{c}_{\alpha,\beta}\, F_\alpha\, \rho\, F_\beta^\dagger\,. \tag{1.267}$$

Thanks to $\mathcal{L}^\dagger(\mathbb{1}) = 0$, it follows that $K = -\frac{1}{2} \sum_{\alpha,\beta=2}^{d^2}\, \tilde{c}_{\alpha,\beta}\, F_\beta^\dagger\, F_\alpha$. Then, substituting this expression in the previous equation proves the claim. □

The operator in Eq. (1.261) is called Lindbladian in the literature. It should be noticed that its expression only depends on the fact that $\mathcal{L}$ is trace preserving, an assumption that is never dropped even when considering non-Markovianity. For this reason, this provides a unified

tool to study the Markovian character of an evolution. In particular, diagonalising the matrix $\tilde{c}_{\alpha,\beta}$, we obtain:

$$\mathcal{L}[\rho] = -i[H,\rho] + \sum_{\alpha}^{d^2} \lambda_\alpha \left(A_\alpha \, \rho \, A_\alpha^\dagger - \frac{1}{2}\{A_\alpha^\dagger \, A_\alpha, \rho\} \right) , \tag{1.268}$$

where $\{\lambda_\alpha\}$ are the eigenvalues of $\tilde{c}_{\alpha,\beta}$ and the A_αs go under the name of jump operators. Since Markovianity is in one to one relation with the positivity of the reduced GKS matrix (Thm. 6), then we also see that this condition can be reduced to the positivity of the rates $\{\lambda_\alpha\}$.

1.5.2 Flux of Fisher information

In this section we explain how the evolution of the Fisher information metric under divisible dynamics can be divided as a sum of independent fluxes. In particular, the object we want to consider is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t} H_g(\Phi_t(\pi)||\Phi_t(\pi + \varepsilon\,\delta\rho)) = \tag{1.269}$$

$$= \frac{\varepsilon^2}{2} \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}\left[\Phi_t(\delta\rho)\, \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(\delta\rho)] \right] + \mathcal{O}\left(\varepsilon^3\right) = \tag{1.270}$$

$$= \frac{\varepsilon^2}{2}\, \dot{\mathcal{F}}_t + \mathcal{O}\left(\varepsilon^3\right) , \tag{1.271}$$

where we denoted the trace in Eq. (1.270) by $\mathcal{F}_t$. In the following we use the notation $\pi_t := \Phi_t(\pi)$ and $\delta\rho_t := \Phi_t(\delta\rho)$. Then, we have the following result:

Theorem 8. *For any divisible dynamics, let $\{A_\alpha(t)\}$ and $\{\lambda_\alpha(t)\}$ be respectively the time dependent jump operators and time dependent rates, defined according to Eq.* (1.268)*. Then, the derivative of the Fisher information takes the form:*

$$\dot{\mathcal{F}}_t = \sum_{\alpha} \lambda_\alpha(t)\; \mathcal{I}_\alpha(t)\,, \tag{1.272}$$

where the current $\mathcal{I}_\alpha(t)$ are given by:

$$\mathcal{I}_\alpha(t) = -2 \int_0^1 \mathrm{d}N_g(s) \left(\mathrm{Tr}\left[\pi_t \left[A_\alpha(t), B_s(t)^\dagger\right]^\dagger \left[A_\alpha(t), B_s(t)^\dagger\right] \right] + \right.$$
$$\left. + \, s\, \mathrm{Tr}\left[\pi_t \left[A_\alpha(t), B_s(t)\right]^\dagger \left[A_\alpha(t), B_s(t)\right] \right] \right), \tag{1.273}$$

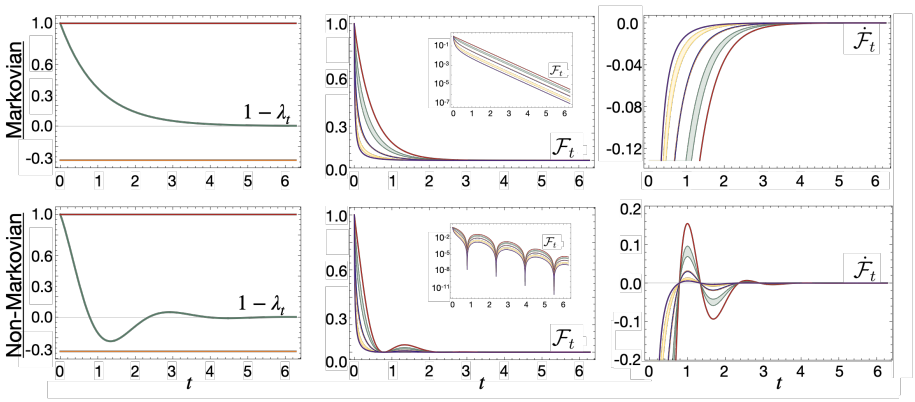


Figure 1.5: Evolution of the quantum Fisher informations under the action of the depolarising channel $\Delta_{\lambda_t}(\rho) = (1-\lambda_t)\rho + \lambda_t \frac{\mathbb{1}}{2}$ on a qubit. The time dependent contraction coefficient are $\lambda_t^M = 1 - e^{-t}$ and $\lambda_t^{NM} = 1 - e^{-t}\cos(2t)$ in the two cases. Notice that non-Markovianity is associated to a local increase in the value of $1-\lambda_t$. In the first panel we show the evolution of λ_t, in the second $\mathcal{F}_t$ (the inset is in logarithmic scale) and in the third the behaviour of $\dot{\mathcal{F}}_t$. Non monotonicity in λ_t are mirrored in the change of sign of $\dot{\mathcal{F}}_t$. The colour scheme is from Fig. 1.4.

and the measure $\mathrm{d}N_g(s)$ *is the one used in Eq.* (1.78)*, while the operators* $B_s(t)$ *are defined as:*

$$B_s(t) := (\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]\,. \tag{1.274}$$

The proof of this theorem is deferred to App. A.1. It should be noticed that the two traces in Eq. (1.273) are positive definite, so the currents $\mathcal{I}_\alpha(t)$ are always negative. This shows that the summands in Eq. (1.272) can become positive if and only if the corresponding rate $\lambda_\alpha(t)$ become negative, i.e., in the presence of non-Markovianity. In the same way, we see that $\dot{\mathcal{F}}_t$ will always be negative for Markovian dynamics, signalling the expected monotonic contraction. This change in behaviour is exemplified in Fig. 1.5, where we plotted the Fisher information and its derivative for a depolarising channel, both in the Markovian and non-Markovian regime. As it can be seen, in this case the oscillations in $\dot{\mathcal{F}}_t$ mirror the onset of non-Markovianity.

In order to give a more practical feeling about the expression in Eq. (1.273), we present here some specific examples for which the currents $\mathcal{I}_\alpha(t)$ take a particularly simple form. The first case that it is

interesting to study is the one of classical evolutions. Then, the jump operators are all of the form $A^{i\leftarrow j} = |i\rangle\langle j|$ and all the observables commute with π_t. Thanks to this fact $B_s(t)$ is simply given by $B_s(t) = B_s(t)^\dagger = \delta\rho_t/((1+s)\pi_t)$, where we use a slight abuse of notation to indicate the entrywise division. This allows us to rewrite Eq. (1.273) as:

$$\mathcal{I}_{(i\leftarrow j)}(t) = \tag{1.275}$$

$$= -2\int_0^1 \mathrm{d}N_g(s)\,(1+s)\,\mathrm{Tr}\left[\pi_t\left[|i\rangle\langle j|\,,\frac{\delta\rho_t}{(1+s)\pi_t}\right]^\dagger\left[|i\rangle\langle j|\,,\frac{\delta\rho_t}{(1+s)\pi_t}\right]\right] = \tag{1.276}$$

$$= -\int_0^1 \mathrm{d}N_g(s)\,\frac{2}{(1+s)}\left(\frac{(\delta\rho_t)_j}{(\pi_t)_j} - \frac{(\delta\rho_t)_i}{(\pi_t)_i}\right)^2(\pi_t)_j = \tag{1.277}$$

$$= -\left(\frac{(\delta\rho_t)_j}{(\pi_t)_j} - \frac{(\delta\rho_t)_i}{(\pi_t)_i}\right)^2(\pi_t)_j\,, \tag{1.278}$$

where in the last line we used the normalisation condition in Eq. (1.79). This result indeed coincides with the one obtained for classical stochastic dynamics [4].

Another case of particular interest is given by the flux of Bures metric. This corresponds to a measure of the form $\mathrm{d}N(s) = \frac{\delta(s-1)}{2}\,\mathrm{d}s$. Then, thanks to the self-adjointness relation $B_1(t) = B_1(t)^\dagger$, by carrying out the integration one obtains:

$$\mathcal{I}_\alpha(t) = -2\,\mathrm{Tr}\left[\pi_t\,[A_\alpha(t), B_1(t)]^\dagger\,[A_\alpha(t), B_1(t)]\right]. \tag{1.279}$$

Interestingly, in this case the flux is directly connected to the symmetric logarithmic derivative of the state. In fact, by inverting Eq. (1.274), one obtains:

$$B_1(t)\,\pi_t + \pi_t\,B_1(t) = \delta\rho_t\,. \tag{1.280}$$

Comparing this expression with the one in Eq. (1.138) for the symmetric logarithmic derivative, it is apparent that $B_1(t) = \frac{1}{2}\,L$. It should be pointed out that the expression of the Bures flux in terms of L was already found in [38].

At the other extreme, the smallest among the contrast functions corresponds to the measure $\mathrm{d}N(s) = \frac{\delta(s)}{2}\mathrm{d}s$. Then, the flows can be expressed as:

$$\mathcal{I}_\alpha(t) = -\,\mathrm{Tr}\left[\pi_t\,[A_\alpha(t), \delta\rho_t\,\pi_t^{-1}]^\dagger\,[A_\alpha(t), \delta\rho_t\,\pi_t^{-1}]\right]. \tag{1.281}$$

Two particularly symmetric cases are the one of the relative entropy, corresponding to $\mathrm{d}N(s) = \frac{1}{2(1+s)}\,\mathrm{d}s$, giving:

$$\mathcal{I}_\alpha(t) = -\int_0^1 \mathrm{d}s\, \frac{1}{1+s}\Bigg(\mathrm{Tr}\left[\pi_t\left[A_\alpha(t), B_s(t)^\dagger\right]^\dagger\left[A_\alpha(t), B_s(t)^\dagger\right]\right] + \\ + s\,\mathrm{Tr}\left[\pi_t\left[A_\alpha(t), B_s(t)\right]^\dagger\left[A_\alpha(t), B_s(t)\right]\right]\Bigg), \tag{1.282}$$

and the one of the square root, which has the measure $\mathrm{d}N(s) = \frac{1}{2\pi\sqrt{s}}\,\mathrm{d}s$, which when plugged in Eq. (1.274) gives:

$$\mathcal{I}_\alpha(t) = -\frac{1}{\pi}\,\mathrm{Re}\int_0^1 \mathrm{d}s\,\Bigg(\frac{1}{\sqrt{s}}\,\mathrm{Tr}\left[\pi_t\left[A_\alpha(t), B_s(t)^\dagger\right]^\dagger\left[A_\alpha(t), B_s(t)^\dagger\right]\right] + \\ + \sqrt{s}\,\mathrm{Tr}\left[\pi_t\left[A_\alpha(t), B_s(t)\right]^\dagger\left[A_\alpha(t), B_s(t)\right]\right]\Bigg). \tag{1.283}$$

Unfortunately, in the last two cases we did not find a way to analytically evaluate the integrals. Still, one can in principle compute the flux $\mathcal{I}_\alpha(t)$ for any Fisher information simply by solving a definite integral on a bounded interval.

1.5.3 Fisher information and Markovianity

In this section we give a complete characterisation of the relation between Markovianity and Fisher information metrics (see Fig. 1.6). As it was mentioned above, all Fisher information metrics monotonically decrease under Markovian evolutions. With the hindsight of Thm. 5 it is then not surprising that the reverse also holds:

Theorem 9. *A classical divisible evolution Φ_t acting on a d-dimensional state space is Markovian if and only if it induces a monotonic decrease in the Fisher information at all times and on the whole domain. In formulae, the Markovianity of the evolution Φ_t is equivalent to the condition:*

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\Phi_{s,t}(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\Phi_{s,t}(\rho)}[\Phi_{s,t}(\delta\rho)]\right]\Bigg|_{s=t} \leq 0 \quad \forall\, t\,,\rho\,,\delta\rho\,, \tag{1.284}$$

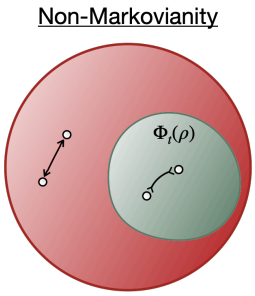
where ρ and $\rho+\delta\rho$ are two arbitrary close by states in the interior of the state space.

For quantum dynamics the same equivalence holds if one augments the space with a d-dimensional ancillary system on which the dynamics acts trivially, i.e., the evolution on the composite space is given by $\mathbb{I}_d \otimes \Phi_t$.

This result was initially proved in [4] with a different method, which we have included in App. A.4.1 for completeness. Still, the result above should be seen as a direct corollary of Thm. 5. Indeed, we can see that even for invertible non-Markovian dynamics, for ε small enough the intermediate maps $\Phi_{t+\varepsilon,t}$ satisfy the first two conditions of Thm. 5 (as it is ε-close to the identity superoperator), so the only way for non-Markovianity to exists is to drop the third condition, i.e., the contractivity of the Fisher information between any two points.

It should be noticed that the equivalence one can find for the contractivity of Fisher information is quite peculiar to this quantity. Indeed, one can contrast the result just obtained with the one for the trace distance, the most canonical quantity studied in the context of non-Markovianity (see App. A.3): in this case one can explicitly construct classical non-Markovian dynamics for which the trace distance between any two points contracts. For quantum evolutions, on the other hand, the use of a d-dimensional ancilla is necessary to separate positive preserving maps from completely positive maps: indeed, complete positivity of Φ_t is equivalent to the fact that $\mathbb{I}_d \otimes \Phi_t$ is positive preserving. It should be noticed that even in this case the same theorem does not hold for the trace distance, despite the addition of the d-dimensional ancillary system. Indeed, it can be shown that one needs at least an ancilla of dimension $d + 1$ for the trace distance to expand in the presence of non-Markovianity [39].

The difference with the trace distance is actually even sharper. Indeed, it is sufficient to adjoin ancillas of dimension high enough to the system (thanks to the construction in [39], $d + 1$ is enough) for the trace distance to operationally witness the non-Markovianity of the evolution. This means that an increase of trace distance can always be obtained on the image of Φ_t when a sufficient number of extra degrees of freedom are provided. Thm. 9, on the other hand, ensures that an expansion in the Fisher information metrics always happens in the presence of non-Markovianity, but it does not say anything about whether the states needed to do so can be physically prepared. If the violation of the monotonicity happens close to the boundary of the state space, for example, and the initial part of the evolution is particularly contracting, there is no way to actually detect the non-Markovian behaviour by looking at the Fisher information alone. Still, one could hope that by using a sufficient



	Contractivity implies Markovianity	Can witness non-Markovianity
Fisher information	YES	With ancillas and post-processing
Trace distance	NO	With ancillas

Figure 1.6: Summary of the results of this section. On the left we give a pictorial representation of Thm. 9: in red we depict the whole state space, while in green the image of the map Φ_t. Thm. 9 tells us that a map is non-Markovian if and only if there exists at least two points in the state space (not necessarily in the image of the map) for which the Fisher distance increases. On the right, we compare the Fisher information with the most canonical quantity to witness non-Markovianity, namely the trace distance. While the monotone contractivity of the Fisher information implies the Markovianity of the dynamics, this does not hold for the the trace distance (Thm. 9 and App. A.3). On the other hand, supplying ancillas to the system allows for the detection of non-Markovianity through the latter, while for the Fisher information one additionally needs some post-processing of the state (Thm. 10 and 11).

number of ancillas the Fisher information could provide witnesses for non-Markovianity. Somehow surprisingly, one can prove that this cannot be the case:

Theorem 10. *Given a divisible evolution Φ_t, no ancillary degree of freedom of finite dimensions or finite number of copy of the dynamics are sufficient to witness all possible non-Markovian evolutions via revivals of the Fisher distance between two initially prepared states.*

The proof of this theorem is provided in App. A.4.2. The difference in behaviour of the trace distance and the Fisher information metrics arises from the translational invariance of the first: thanks to this property, if a witness exists anywhere on the state space, then it can always be translated into the image of Φ_t. The Fisher information, on the other hand, has a strong dependence on the base-point, so that the same kind of argument cannot be applied.

Despite the negative result of Thm. 9, one can still define a witness

based on Fisher information by using post-processing:

Theorem 11. *Given an evolution Φ_t, for any state ρ and perturbation $\delta\rho$ defined on the system space and on some ancillary degrees of freedom, it is possible to implement a class of transformations $F^{(t)}_{\delta\rho}$ depending on Φ_t and $\delta\rho$ that can be used to witness non-Markovianity at time t through the use of Fisher information. This means that if the evolution $\Phi_{t+\mathrm{d}t,t}$ is Markovian, then*

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{F^{(t)}_{\delta\rho}\circ\Phi_s(\rho)}[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)]\right]\Big|_{s=t}\leq 0\,, \qquad (1.285)$$

whereas in the presence of non-Markovianity there always exists at least one $\delta\rho$ for which the inequality is reversed.

The minimal dimension of the ancilla for classical systems is $d_A = 2$, while for quantum maps one needs $d_A = d + 1$.

There is a shortcoming to this construction, though: in the definition of the post-processing $F^{(t)}_{\delta\rho}$ one needs to assume complete knowledge about the dynamics Φ_t until the onset of Markovianity. In this way, one either needs to try all the possible $\delta\rho$, or has to know in advance the structure of the dynamics in order to provide an explicit construction. Still, this example serves more as a proof of principle showing the possibility of designing post-processing filters to exploit the Fisher information for the detection of non-Markovianity.

This completes the characterisation of the relation between Markovianity and contractivity of the Fisher information, both on the image of Φ_t and on the rest of the state space, as summarised in Fig. 1.6. We point out once again the importance of Thm. 9: both Chentsov theorem and its quantum generalisation by Petz (Thm. 1 and Thm. 3) identify the defining property of the Fisher information metric to be its contractivity under stochastic maps or quantum channels. Thm. 9, on the other hand, could be read off as saying that the defining property of Markovianity is that it contracts the Fisher information monotonically at all times. This second implication shows how natural the concept of Fisher metric is in the context of open system dynamics.

1.5.4 Retrodiction and Fisher information

In the literature about non-Markovianity, one of the most used expression is *backflow of information*. Still, by looking at the usual quantifiers

defined to assess it, one might be surprised to discover that for the most part they are actually distinguishability measures. For example, even in the previous section the two objects considered are the trace distance and the Fisher information, both accounting for the statistical difference between states. Whereas one could in principle justify the interpretation of non-monotonicity for these quantities as actual backflow of information, here we take a different approach: we prove, in fact, that the contractivity of the Fisher information is in one-to-one correspondence with the monotonic degradation in the ability of an agent to retrodict the initial state of the system (Thm. 12).

Before doing so, we need some formalism. First of all, we present a way to simulate the Fisher scalar product at time t by the scalar product at time $t = 0$. This is easily obtained by the following rearranging:

$$\begin{aligned} \mathrm{Tr}\left[\Phi_t(A)\,\mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(B)]\right] &= \mathrm{Tr}\left[A\,\Phi_t^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t}[\Phi_t(B)]\right] = && (1.286)\\ &= \mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi\left[\left(\mathbb{J}_f\big|_\pi \circ \Phi_t^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}\right)\Phi_t(B)\right]\right] = && (1.287)\\ &= \mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi[\widetilde{\Phi}_{f,t}\,\Phi_t(B)]\right], && (1.288) \end{aligned}$$

where in the last line we implicitly defined the generalised Petz recovery map $\widetilde{\Phi}_{f,t}$ (the usual Petz recovery map corresponds to setting $f(x) = \sqrt{x}$). In this way, the evolution of any Fisher scalar products can be modelled without the need to actually evolve the state, at the cost of introducing a time dependent vector $\widetilde{\Phi}_{f,t}\Phi_t(B)$.

The generalised Petz recovery map satisfies a number of interesting features: first of all, it should be noticed that for classical systems it reduces to the Bayes rule. In fact, if Φ_t is a stochastic map, and π a diagonal state, the coordinate expression of $\widetilde{\Phi}_{f,t}$ is given by:

$$\widetilde{\Phi}_{f,t}(\cdot) = \sum_{i,j} \frac{(\Phi_t)_{j,i}\,\pi_i}{(\Phi_t(\pi))_j}\ |i\rangle\langle j|\cdot|j\rangle\langle i|\,, \tag{1.289}$$

where one can recognise the Bayes rule by interpreting π_i as the probability of obtaining the microstate i, $(\Phi_t(\pi))_j$ as the probability to be in the microstate j after the evolution and, finally, $(\Phi_t)_{j,i}$ as the conditional probability of the transition $i \to j$.

With this connection in mind, we call the state π the *prior*. Interest-

ingly, the generalised Petz recovery map perfectly recovers it:

$$\widetilde{\Phi}_{f,t}\,\Phi_t(\pi) = \mathbb{J}_f\big|_\pi \circ \Phi_t^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(\pi)] = \tag{1.290}$$

$$= \mathbb{J}_f\big|_\pi[\Phi_t^\dagger(\mathbb{1})] = \mathbb{J}_f\big|_\pi[\mathbb{1}] = \pi\,, \tag{1.291}$$

where we used the fact that $\mathbb{J}_f$ acts as the multiplication operator on commuting states, together with the fact that Φ_t is trace preserving, which implies $\Phi_t^\dagger(\mathbb{1}) = \mathbb{1}$. This equation shows that the evolution of the state π can be completely undone by applying $\widetilde{\Phi}_{f,t}$.

Additionally, $\widetilde{\Phi}_{f,t}$ is trace preserving if π is full-rank (which we will always assume to be the case in the rest of the section), as it can be verified by applying its adjoint on the identity matrix:

$$\widetilde{\Phi}_{f,t}^\dagger[\mathbb{1}] = \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)} \circ \Phi_t \circ \mathbb{J}_f\big|_\pi[\mathbb{1}] = \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(\pi)] = \mathbb{1}\,. \tag{1.292}$$

Complete positivity, on the other hand, does not hold in general, as at least one between $\mathbb{J}_f$ and $\mathbb{J}_f^{-1}$ is not CP. A special case in this context is given by the geometric mean, $f(x) = \sqrt{x}$, as it is the only case in which both operators are in fact CP (Sec. 1.4.8). This is the original definition by Petz, which we denote by $\widetilde{\Phi}_{P,t} := \widetilde{\Phi}_{\sqrt{x},t}$. It can be argued that it is the most natural generalisation of the Bayes' rule to non-commutative probability (see Sec. 3.3, or [2] and references therein), and for this reason it can be interpreted as a way to physically obtain information about the initial conditions of the system (even if is not always the most optimal method, as it is shown in App. C).

Still, if one drops the request that $\widetilde{\Phi}_{f,t}$ needs to be physically implementable, then all other features of the Petz' recovery map also hold for $\widetilde{\Phi}_{f,t}$, where f is a generic standard monotone function. For example, one of the defining property of $\widetilde{\Phi}_{P,t}$ is being involutive, which means that taking the Petz' map of the Petz' map gives the original channel (see Sec. 3.3). This is in fact a general feature. Choosing as a prior $\Phi_t(\pi)$ one can indeed see that:

$$\widetilde{(\widetilde{\Phi}_{f,t})}_{f,t} = \mathbb{J}_f\big|_{\Phi_t(\pi)} \circ \widetilde{\Phi}_{f,t}^\dagger \circ \mathbb{J}_f^{-1}\big|_{\widetilde{\Phi}_{f,t}\Phi_t(\pi)} = \tag{1.293}$$

$$= \mathbb{J}_f\big|_{\Phi_t(\pi)} \circ \left(\mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)} \circ \Phi_t \circ \mathbb{J}_f\big|_\pi\right) \circ \mathbb{J}_f^{-1}\big|_\pi = \Phi_t\,, \tag{1.294}$$

where in the second line we used Eq. (1.290) to obtain the prior state.

Interestingly, if Φ_t is a divisible dynamics, so is $\widetilde{\Phi}_{f,t}$. In fact, one can define the intermediate recovery maps as:

$$\widetilde{\Phi}_{f,(t,s)} = \mathbb{J}_f\big|_{\Phi_s(\pi)} \circ \Phi_{t,s}^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}\,. \tag{1.295}$$

Then, it is easy to show that $\widetilde{\Phi}_{f,t} = \widetilde{\Phi}_{f,s}\widetilde{\Phi}_{f,(t,s)}$, i.e., similarly to what happens for adjoint maps, generalised Petz' maps compose from left to right. In fact, one has:

$$\widetilde{\Phi}_{f,s}\,\widetilde{\Phi}_{f,(t,s)} = \mathbb{J}_f\big|_\pi \circ \Phi_s^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_s(\pi)} \circ \mathbb{J}_f\big|_{\Phi_s(\pi)} \circ \Phi_{t,s}^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)} = \tag{1.296}$$

$$= \mathbb{J}_f\big|_\pi \circ \Phi_t^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}\,, \tag{1.297}$$

proving the claim. In particular, in the case of $f(x) = \sqrt{x}$, the Markovian character of Φ_t is reflected in $\widetilde{\Phi}_{\sqrt{x},t}$, i.e., $\widetilde{\Phi}_{\sqrt{x},t}$ is Markovian if and only if the original map is Markovian.

Finally, before passing to the main theorem of this section, it is interesting to comment on the spectral properties of $\widetilde{\Phi}_{f,t}\Phi_t$. First, it should be noticed that it is self-adjoint with respect to $\mathbb{J}_f^{-1}\big|_\pi$, as it can be readily verified:

$$\mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi[\widetilde{\Phi}_{f,t}\,\Phi_t(B)]\right] = \mathrm{Tr}\left[\Phi_t(A)\,\mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(B)]\right] = \tag{1.298}$$

$$= \mathrm{Tr}\left[\left(\mathbb{J}_f\big|_\pi \circ \Phi^\dagger \circ \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}\right)\Phi_t(A)\,\mathbb{J}_f^{-1}\big|_\pi[(B)]\right] = \tag{1.299}$$

$$= \mathrm{Tr}\left[\widetilde{\Phi}_{f,t}\,\Phi_t(A)\,\mathbb{J}_f^{-1}\big|_\pi[B]\right]\,, \tag{1.300}$$

where we repeatedly used the self-adjointness of $\mathbb{J}_f$ to move the superoperators from right to left. This shows that its spectrum is real and its eigenoperators are hermitian. Moreover, it is also positive since it can be rewritten as

$$\widetilde{\Phi}_{f,t}\,\Phi_t = \mathbb{J}_f^{\frac{1}{2}}\big|_\pi \left((\mathbb{J}_f^{-\frac{1}{2}}\big|_{\Phi_t(\pi)}\Phi_t\,\mathbb{J}_f^{\frac{1}{2}}\big|_\pi)^\dagger(\mathbb{J}_f^{-\frac{1}{2}}\big|_{\Phi_t(\pi)}\Phi_t\,\mathbb{J}_f^{\frac{1}{2}}\big|_\pi\right)\mathbb{J}_f^{-\frac{1}{2}}\big|_\pi\,, \tag{1.301}$$

and similarity transformations preserve the spectrum. Indeed, one can also prove that the spectrum is actually contained in the interval $[0,1]$. Indeed, this follows from the contractivity of the Fisher information, since:

$$\frac{\mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi[\widetilde{\Phi}_{f,t}\,\Phi_t(A)]\right]}{\mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi[A]\right]} = \frac{\mathrm{Tr}\left[\Phi_t(A)\,\mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(A)]\right]}{\mathrm{Tr}\left[A\,\mathbb{J}_f^{-1}\big|_\pi[A]\right]} \leq 1\,. \tag{1.302}$$

Additionally, 1 is always an eigenvalue, as $\widetilde{\Phi}_{f,t}\,\Phi_t(\pi) = \pi$. Then, if one denotes by $\{E_i\}$ its other eigenoperators, it follows from the orthogonality between them, that they have to be traceless:

$$0 = \mathrm{Tr}\left[E_i\,\mathbb{J}_f^{-1}\big|_\pi[\pi]\right] = \mathrm{Tr}\,[E_i]\,. \tag{1.303}$$

In the following, we denote by $\perp_f$ the condition of being orthogonal with respect to $\mathbb{J}_f^{-1}\big|_\pi$. By using the min-max principle, one can obtain the second largest eigenvalue λ_2 as:

$$\lambda_2 = \sup_{A \perp_f \pi} \frac{\mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi[\widetilde{\Phi}_{f,t}\, \Phi_t(A)]\right]}{\mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi[A]\right]} = \sup_{\mathrm{Tr}[A]=0} \frac{\mathrm{Tr}\left[\Phi_t(A)\, \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(A)]\right]}{\mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi[A]\right]}\,. \tag{1.304}$$

The object on the right is called the Riemannian contraction coefficient of Φ_t at π, and it is denoted by $\eta_f^F(\Phi_t, \pi)$. It quantifies the minimum rate at which the Fisher information is lost in π. In fact, it follows from the definition that:

$$\mathrm{Tr}\left[\Phi_t(\delta\rho)\, \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(\delta\rho)]\right] \leq \eta_f^F(\Phi_t, \pi)\, \mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_\pi[\delta\rho]\right]\,, \tag{1.305}$$

for any traceless perturbation $\delta\rho$. One can define a similar object for contrast functions, by taking the superior over all the possible states:

$$\eta_g^{RE}(\Phi_t) = \sup_{\rho,\sigma} \frac{H_g(\Phi_t(\rho)||\Phi_t(\sigma))}{H_g(\rho||\sigma)}\,. \tag{1.306}$$

It is straightforward to verify that $\sup_\pi \eta_f^F(\Phi_t, \pi) =: \eta_f^F(\Phi_t) \leq \eta_g^{RE}(\Phi_t)$, where f and g are connected by Eq. (1.66), as $\eta_f^F(\Phi_t)$ just corresponds to restricting the superior in Eq. (1.306) to close-by states. The inequality is usually strict, but there are some special cases in which the equality can be proven, as in the case of the symmetrised relative entropy [32]. In general, whenever the inequality is not saturated, this implies that there exist two states $\rho \neq \sigma$ saturating the superior in Eq. (1.306).

In the discussion above we argued that the operator $\widetilde{\Phi}_{f,t}$ can be interpreted as a method to retrieve some information about the initial state of the system. This interpretation is further corroborated by the connection between $\widetilde{\Phi}_{f,t}\, \Phi_t$ and the contraction coefficients.

Consider now a state $\pi + \delta\rho$, for $\delta\rho$ a small perturbation, and evolve it according to Φ_t; at this point, $\widetilde{\Phi}_{f,t}$ is applied so to recover as much information about the initial state as possible. The quality of the retrieval can be quantified by the contrast function:

$$H_g(\pi + \delta\rho||\widetilde{\Phi}_{f,t}\, \Phi_t(\pi + \delta\rho)) = H_g(\pi + \delta\rho||\pi + \widetilde{\Phi}_{f,t}\, \Phi_t(\delta\rho)) = \tag{1.307}$$

$$= \frac{1}{2}\, \mathrm{Tr}\left[(\delta\rho - \widetilde{\Phi}_{f,t}\, \Phi_t(\delta\rho))\, \mathbb{J}_f^{-1}\big|_\pi[(\delta\rho - \widetilde{\Phi}_{f,t}\, \Phi_t(\delta\rho))]\right] + \mathcal{O}\left(|\delta\rho|^3\right). \tag{1.308}$$

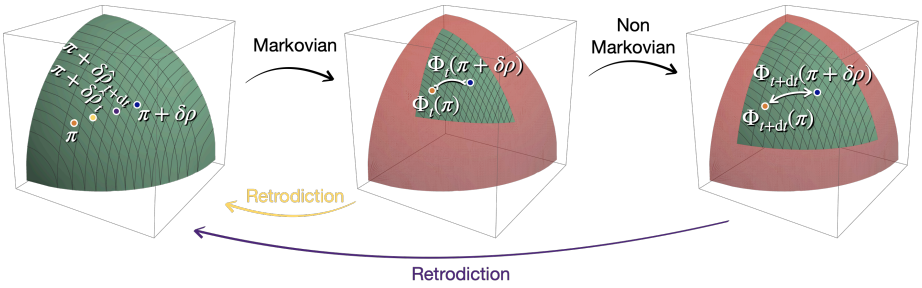


Figure 1.7: Depiction of the content of Thm. 12:consider a dynamics Φ_t which is Markovian till some time t. This implies that the Fisher distance between any two points on the state space contracts. If between time t and $t + \mathrm{d}t$ we observe non-Markovianity through the Fisher distance, i.e., two points in the image of Φ_t gets farther away, then retrodicting at time $t + \mathrm{d}t$ gives a better result than at time t (we use the notation $\delta\hat{\rho}_t := \widetilde{\Phi}_{f,t}\Phi_t(\delta\rho)$).

The interpretation of $\widetilde{\Phi}_{f,t}$ as a recovery map suggests that, if Φ_t is Markovian, the divergence above will increase with time, as the ability to retrodict the initial state deteriorates further and further. Still, from a mathematical point of view, it should be noticed that in principle there is no intuitive reason for this to be the case: in fact, $H_g(\rho||\sigma)$ is known to be contractive only if the same channel is applied on both states, whereas in the equation above $\widetilde{\Phi}_{f,t}\,\Phi_t$ is applied only on the right. The following theorem bridges the gap between the intuitive interpretation of Eq. (1.307) and its actual mathematical form:

Theorem 12. *Given a divisible dynamics Φ_t and a prior state π, define the generalised Petz' map $\widetilde{\Phi}_{f,t}$ according to Eq.* (1.288)*. Then, for any perturbation $\delta\rho$, it holds that:*

$$\frac{\mathrm{d}}{\mathrm{d}t} H_g(\pi + \delta\rho||\widetilde{\Phi}_{f,t}\,\Phi_t(\pi + \delta\rho)) > 0 \quad \Longleftrightarrow$$
$$\Longleftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Tr}\left[\Phi_t(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)}[\Phi_t(\delta\rho)]\right] < 0\,. \qquad (1.309)$$

where g and f are related by Eq. (1.66)*. Since the Fisher information is monotonically contractive under Markovian dynamics, this implies that the contrast function on the left is monotonically expanding in the same regime. Moreover, any backflow in the Fisher information is mirrored in an increased ability to retrodict the initial state $\pi + \delta\rho$.*

To the best of our knowledge, this result is the first that explicitly considers backflow of information at time $t = 0$ (in contrast with the usual approach at time t). Moreover, on the one hand, it legitimates the intuitive interpretation of $\widetilde{\Phi}_{f,t}$ as a state retrieval map. On the other, it directly connects the contractivity properties of the Fisher information at time t, which is a distinguishability measure, with the decrease in the ability of an agent to retrodict the initial state of the system.

Proof. Starting from the expansion of Eq. (1.307) in terms of the Fisher information metric, we can reassemble the terms as:

$$\frac{\mathrm{d}}{\mathrm{d}t} H_g(\pi + \delta\rho || \widetilde{\Phi}_{f,t}\, \Phi_t(\pi + \delta\rho)) = \tag{1.310}$$

$$= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\, \mathrm{Tr}\left[(\delta\rho - \widetilde{\Phi}_{f,t}\, \Phi_t(\delta\rho))\, \mathbb{J}_f^{-1}\big|_\pi [(\delta\rho - \widetilde{\Phi}_{f,t}\, \Phi_t(\delta\rho))] \right] = \tag{1.311}$$

$$= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\, \mathrm{Tr}\left[((\mathbb{I} - \widetilde{\Phi}_{f,t}\, \Phi_t)(\delta\rho))\, \mathbb{J}_f^{-1}\big|_\pi [((\mathbb{I} - \widetilde{\Phi}_{f,t}\, \Phi_t)(\delta\rho))] \right] = \tag{1.312}$$

$$= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\, \mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_\pi [((\mathbb{I} - \widetilde{\Phi}_{f,t}\, \Phi_t)^2(\delta\rho))] \right] = \tag{1.313}$$

$$= -\mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_\pi [((\mathbb{I} - \widetilde{\Phi}_{f,t}\, \Phi_t) \circ (\frac{\mathrm{d}}{\mathrm{d}t} \widetilde{\Phi}_{f,t}\, \Phi_t)(\delta\rho))] \right], \tag{1.314}$$

where in Eq. (1.313) we used the fact that both the identity operator and $\widetilde{\Phi}_{f,t}\, \Phi_t$ are self-adjoint with respect to $\mathbb{J}_f^{-1}\big|_\pi$. Thanks to Eq. (1.302) we also know that $(\mathbb{I} - \widetilde{\Phi}_{f,t}\, \Phi_t)$ is a positive operator. On the other hand, $-\frac{\mathrm{d}}{\mathrm{d}t}\, \widetilde{\Phi}_{f,t}\, \Phi_t$ has positive eigenvalues if and only if the corresponding Fisher information is contractive, as one can use the standard rewriting:

$$-\mathrm{Tr}\left[A\, \mathbb{J}_f^{-1}\big|_\pi [(\frac{\mathrm{d}}{\mathrm{d}t} \widetilde{\Phi}_{f,t}\, \Phi_t)(A))] \right] = -\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}\left[\Phi_t(A)\, \mathbb{J}_f^{-1}\big|_{\Phi_t(\pi)} [\Phi_t(A))] \right]. \tag{1.315}$$

Since the product of two positive operators has positive spectrum, this connection proves the claim. □

1.5.5 Fisher information and detailed balance

Chentsov theorem (Thm. 1) singles out the Fisher information metric in terms of the dynamical notion of contractivity under stochastic evolutions. Thm. 5, on the other hand, reverts the direction of the implication, proving the equivalence between complete positivity and contractivity of

Fisher information. This result demonstrates the intrinsically dynamical character of the Fisher information, showing that canonical properties of evolutions can be formulated in terms of their relation with the Fisher metric, as we have seen in the case of Markovianity (see Thm. 9).

Indeed, this is not an isolated case: in this section we explain how the notion of detailed balance evolution can be given in terms of self-adjointness with respect to the Fisher information scalar product. To this end, it is useful to start with the study of the classical case to avoid the complication coming from the variety of different quantum Fisher information.

As discussed in App. A.2, the dynamics of a classical Markov chain Φ_t are described by stochastic matrices, which can be equivalently characterised in terms of their rate matrices R_t, given by the limit:

$$R_t := \lim_{\varepsilon \to 0} \frac{\Phi_{t+\varepsilon,t} - \mathbb{I}}{\varepsilon} \,. \tag{1.316}$$

In particular, in the appendix it is also shown that these can be generically decomposed as:

$$R_t = \sum_{i \neq j} a^{(t)}_{i \leftarrow j} \left(|i\rangle\langle j| - |j\rangle\langle j| \right) , \tag{1.317}$$

where $a^{(t)}_{i \leftarrow j}$ are real coefficients which are non-negative for Markovian evolutions. Then, the canonical condition of being detailed balance with respect to the state π can be formulated in terms of the rates alone, as:

$$a^{(t)}_{i \leftarrow j} \, \pi_j = a^{(t)}_{j \leftarrow i} \, \pi_i \,, \tag{1.318}$$

for every i and j. It should be noticed that this condition directly implies that π is a steady state, as it can be readily verified by direct computation:

$$R_t(\pi) = \sum_{i \neq j} a^{(t)}_{i \leftarrow j} \, \pi_j \left(|i\rangle - |j\rangle \right) = \tag{1.319}$$

$$= \sum_{i<j} \Big(a^{(t)}_{i \leftarrow j} \, \pi_j \left(|i\rangle - |j\rangle \right) + a^{(t)}_{j \leftarrow i} \, \pi_i \left(|j\rangle - |i\rangle \right) \Big) = \tag{1.320}$$

$$= \sum_{i<j} a^{(t)}_{i \leftarrow j} \, \pi_j \left(\left(|i\rangle - |j\rangle \right) + \left(|j\rangle - |i\rangle \right) \right) = 0 \,. \tag{1.321}$$

Indeed, detailed balance corresponds to a stronger notion of equilibration: not only the dynamics has π as a fixed points, but it is also time symmetric

at equilibrium. In fact, if one interprets the rates $a^{(t)}_{j\leftarrow i}$ as the probability per unit of time of the transition $j \leftarrow i$, Eq. (1.318) can be read as the condition that the probability of observing the transition $j \leftarrow i$ is equal to the one for the reverse transition $j \rightarrow i$, when the system is at equilibrium. For this reason, detailed balance encodes the request of microscopic reversibility of the dynamics, i.e., the fact that at a molecular level the equations of motion are time symmetric.

We show now that Eq. (1.318) can be naturally formulated in terms of the Fisher information scalar product. In analogy with the quantum case (Eq. (1.67)), we use the notation for the scalar product:

$$K_\pi(\delta\rho, \delta\sigma) := \mathrm{Tr}\left[\delta\rho\, \mathcal{J}_\pi^{-1}[\delta\sigma]\right] = \mathrm{Tr}\left[\delta\rho\, \delta\sigma\, \pi^{-1}\right] , \tag{1.322}$$

where we implicitly defined $\mathcal{J}_\pi$ to be the component-wise multiplication by π, and all the operators involved commute. This scalar product naturally emerges when one is considering variations of states (as proven in Thm. 2), so from a differential geometry point of view the two vectors $\delta\rho$ and $\delta\sigma$ should be elements of the tangent space of the state space (i.e., hermitian, traceless operators).

Moreover, one can also interpret $\mathcal{J}_\pi$ as the scalar product on the cotangent space, i.e., the space of observables: indeed, a metric on the tangent space naturally induces one on its dual by taking the pointwise matrix inverse [40]. Hence, one can also define the Fisher scalar product on the space of observables as:

$$K^o_\pi(A, B) := \mathrm{Tr}\left[A\, \mathcal{J}_\pi[B]\right] = \mathrm{Tr}\left[A\, B\, \pi\right] , \tag{1.323}$$

where in this case A and B are not required to be traceless. It should be noticed that $K^o_\pi(A, B)$ is related to the two-point correlation function used in statistical mechanics and in linear response theory.

Now, the adjoint of a superoperator Φ with respect to K_π satisfies the equation:

$$K_\pi(\delta\rho, \Phi(\delta\sigma)) = \mathrm{Tr}\left[\delta\rho\, \mathcal{J}_\pi^{-1}[\Phi(\delta\sigma)]\right] = \mathrm{Tr}\left[\mathcal{J}_\pi^{-1}[[\delta\rho]\Phi(\delta\sigma)\right] = \tag{1.324}$$

$$= \mathrm{Tr}\left[\Phi^\dagger \circ \mathcal{J}_\pi^{-1}[[\delta\rho]\, \delta\sigma\right] = \tag{1.325}$$

$$= \mathrm{Tr}\left[\left(\mathcal{J}_\pi \circ \Phi^\dagger \circ \mathcal{J}_\pi^{-1}\right)(\delta\rho)\, \mathcal{J}_\pi^{-1}[\delta\sigma]\right] = K_\pi(\widetilde{\Phi}(\delta\rho), \delta\sigma) , \tag{1.326}$$

where we implicitly defined $\widetilde{\Phi} := \mathcal{J}_\pi \circ \Phi^\dagger \circ \mathcal{J}_\pi^{-1}$. Notice that since the equation above holds for generic vectors of the tangent space, we can

identify with $\widetilde{\Phi}$ the adjoint operation with respect to the Fisher information metric. Hence, the condition to be self-adjoint with respect to K_π is given by $\widetilde{\Phi} = \Phi$, i.e., $\Phi \circ \mathcal{J}_\pi = \mathcal{J}_\pi \circ \Phi^\dagger$. Writing this equation in coordinates as:

$$\Phi_{i,j}\, \pi_j = \bar{\Phi}_{j,i}\, \pi_i \,, \tag{1.327}$$

we obtain a condition akin to the one in Eq. (1.318). Indeed, it follows from the equation above that R_t is detailed balance exactly means that $\tilde{R}_t = R_t$, i.e., the generator of the dynamics is self-adjoint with respect to the Fisher information scalar product.

This construction is done in the Schrödinger picture, meaning that the states are the only evolving objects, while observables are static quantities. The dual situation, dubbed Heisenberg picture, is the one in which states are fixed in time, while the whole dynamics is relegated to observables. In this case it is well known that the generator of the dynamics is given by $R_t^\dagger$. Moreover, as it was argued above, the natural Fisher scalar product is the one given by K_π^o. We denote the adjoint with respect to this scalar product by $\widetilde{\Phi}^o := \mathcal{J}_\pi^{-1} \circ \Phi^\dagger \circ \mathcal{J}_\pi$, where this condition can be verified carrying out calculations completely analogous to the one that led to Eq. (1.326). It is easy to verify that:

$$\widetilde{\Phi} = \Phi \qquad \Longleftrightarrow \qquad (\widetilde{\Phi^\dagger})^o = \Phi^\dagger \,, \tag{1.328}$$

which implies that R_t is K_π-self-adjoint if and only if $R_t^\dagger$ is K_π^o-self-adjoint. This shows that one can formulate the condition of being detailed balance equivalently in the Heisenberg picture.

Putting these results together we obtain the theorem:

Theorem 13. *The following conditions are equivalent:*

1. *the dynamics is detailed balance, i.e., the rate matrix is characterised by coefficients satisfying*

$$a^{(t)}_{i\leftarrow j}\, \pi_j = a^{(t)}_{j\leftarrow i}\, \pi_i \,; \tag{1.329}$$

2. *the rate matrix R_t is self-adjoint with respect to the Fisher scalar product:*

$$\tilde{R}_t = R_t \,; \tag{1.330}$$

3. the rate matrix in the Heisenberg picture (i.e., $R_t^\dagger$) is self-adjoint with respect to the dual Fisher metric:

$$\widetilde{(R_t^\dagger)^o} = R_t^\dagger \,. \tag{1.331}$$

This characterisation shows how the condition of being detailed balance directly corresponds to the self-adjointness of the generator of the dynamics with respect to properly defined scalar products. We can now pass to give a generalisation of the definition of detailed balance evolution to the quantum regime based on this hindsight.

To this end, let us consider the dynamics induced by the Lindbladian operator in the form in Eq. (1.261), i.e.:

$$\mathcal{L}[\rho] = -i[H,\rho] + \sum_{\alpha,\beta=2}^{d^2} \tilde{c}_{\alpha,\beta} \left(F_\alpha \,\rho\, F_\beta^\dagger - \frac{1}{2}\{F_\beta^\dagger F_\alpha, \rho\} \right) . \tag{1.332}$$

Since the Fisher information is invariant under the action of a purely Hamiltonian dynamics, whereas it is contracting otherwise, we divide the two terms in the Lindbladian by introducing the notation $\mathcal{U}(\rho) := -i[H,\rho]$, and we denote by $\mathcal{L}_\mathcal{D} := \mathcal{L} - \mathcal{U}$, which is dubbed dissipator. It should be noticed that $\mathcal{U}$ is skew-Hermitian with respect to the Hilbert-Schmidt scalar product, meaning that:

$$\mathrm{Tr}\,[A\,\mathcal{U}(B)] = -\mathrm{Tr}\,[\mathcal{U}(A)\,B] \,. \tag{1.333}$$

Given the structure of the Lindbladian above, we can introduce the notion of quantum detailed balance. Historically, one of the first formalisations of this notion was provided by Alicki in [41], and it is based on the following scalar product on the space of observables:

$$K_\pi^o(A,B) := \mathrm{Tr}\,[AB\,\pi] \,, \tag{1.334}$$

where in this case A, B and π are not required to commute in general. Similarly to the case of classical systems, this scalar product is quite natural as it is related to two-points correlation functions, but it should also be kept in mind that it is not part of the Fisher family. Using the same notation as in the classical case, we denote by $\widetilde{\Phi}^o$ the adjoint of the map Φ with respect to the scalar product in Eq. (1.334). Then, the definition proposed by Alicki reads:

Definition 1 (Heisenberg picture [41]). *The dynamics generated in the Heisenberg picture by the operator $\mathcal{L}^\dagger$ is detailed balance if the three conditions are satisfied:*

1. *$\mathcal{L}^\dagger$ is normal with respect to the scalar product K_π^o:*

$$[\mathcal{L}^\dagger, (\widetilde{\mathcal{L}^\dagger})^o] = 0\,; \tag{1.335}$$

2. *the commutator $\mathcal{U}$ is skew-Hermitian with respect to K_π^o:*

$$\tilde{\mathcal{U}}^o = -\mathcal{U}\,; \tag{1.336}$$

3. *the dissipator $\mathcal{L}_\mathcal{D}^\dagger$ is self-adjoint with respect to K_π^o:*

$$(\widetilde{\mathcal{L}_\mathcal{D}^\dagger})^o = \mathcal{L}_\mathcal{D}^\dagger\,. \tag{1.337}$$

Interestingly, from this definition one can deduce a structural characterisation of detailed balanced Lindbladian (see [41]), which leads to the equivalent definition of detailed balance:

Definition 2 (Structural definition). *The dynamics generated by the Lindbladian operator $\mathcal{L}$ is detailed balance if its diagonal form can be written as:*

$$\mathcal{L}(\rho) = -i[H,\rho] + \sum_{\omega,i} \lambda_i^\omega \left(A_i^\omega\, \rho\, (A_i^\omega)^\dagger - \frac{1}{2}\{(A_i^\omega)^\dagger A_i^\omega, \rho\} \right), \tag{1.338}$$

and the following conditions hold:

1. *$[H,\pi] = 0$;*
2. *$(A_i^\omega)^\dagger = A_i^{-\omega}$;*
3. *$\pi\, A_i^\omega\, \pi^{-1} = e^\omega\, A_i^\omega$;*
4. *$\lambda_i^\omega = e^\omega\, \lambda_i^{-\omega}$.*

In the literature one usually finds this as the usual definition of detailed balance, as it mirrors the same structural properties we saw for classical systems. Indeed, from the condition 3 we know that A_i^ω are eigenoperators of the the auto-modular operator $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$, with eigenvalue e^ω. On the other hand, all the eigenvalues of $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$ are of the form

π_i/π_j, so the only values of ω that are allowed are the one satisfying the constraint $e^{\omega} = \pi_i/\pi_j$ for some i and j. Then, by substituting this expression into condition 4, one obtains the analogous characterisation at the level of the rates we found for classical detailed balanced systems (see Eq. (1.329)).

Despite this positive result, it should be noticed that the choice of the scalar product in Eq. (1.334) is somehow arbitrary, as in the passage from commuting observables to the non-commuting case there are many different possible orderings that can be used to extend the multiplication operator $\mathcal{J}_\pi$. Moreover, it should be noticed that the scalar product K^o_π in Eq. (1.334) is not monotone under CP-maps, a property that its classical counterpart had. For these reasons, in the following we show how one can define detailed balance through the help of quantum Fisher informations.

A first possible definition is given by imposing $\widetilde{\Phi}_f = \Phi$ for some Fisher scalar product $K_{f,\pi}$ (see Eq. (1.67)), where we used the definition of generalised Petz map in Eq. (1.288), i.e., $\widetilde{\Phi}_f := \mathbb{J}_f\big|_\pi \circ \Phi_t^\dagger \circ \mathbb{J}_f^{-1}\big|_\pi$. As it could be expected, the structure induced is much richer in this case. In fact, one can show that different $K_{f,\pi}$ induce inequivalent notions of detailed balance, by constructing dynamics that are detailed balanced with respect to one f but not with respect to others [42]. Moreover, all these conditions are all weaker than the one provided by Def. 1. This means that dynamics that are detailed balance with respect to K^o_π are also detailed balance with respect with $K_{f,\pi}$, but not the other way around.

Since in principle there is no preferred definition of quantum Fisher information, we choose to impose the condition of detailed balance for all the $K_{f,\pi}$ at the same time:

Definition 3 (Schroedinger picture)**.** *The dynamics generated by the Lindbladian operator $\mathcal{L}$ is detailed balance if for every standard monotone function f the following holds*

1. *$\mathcal{L}$ is normal with respect to all the scalar products $K_{f,\pi}$:*

$$[\mathcal{L}, \widetilde{\mathcal{L}}_f] = 0\,; \tag{1.339}$$

2. *the commutator $\mathcal{U}$ is skew-Hermitian with respect to $K_{f,\pi}$:*

$$\widetilde{\mathcal{U}}_f = -\mathcal{U}\,; \tag{1.340}$$

3. *the dissipator $\mathcal{L}_\mathcal{D}$ is self-adjoint with respect to $K_{f,\pi}$:*

$$(\widetilde{\mathcal{L}_\mathcal{D}})_f = \mathcal{L}_\mathcal{D}\,. \tag{1.341}$$

We arrived to three different definitions of what it means for a Markovian dynamics to be detailed balance. At this point, we can wrap everything together, providing a characterisation of their interdependency:

Theorem 14. *The following conditions are equivalent:*

1. *the generator of the dynamics in the Heisenberg picture $\mathcal{L}^\dagger$ satisfies the adjointness relations in Def. 1*

2. *the Lindbladian $\mathcal{L}$ satisfies the structural characterisation in Def. 2.*

These conditions imply:

3. *the generator of the dynamics in the Schroedinger picture $\mathcal{L}$ satisfies the adjointness relations in Def. 3.*

Moreover, if the Hamiltonian H is non-degenerate the three conditions are equivalent.

The proof of this result is provided in App. A.5. It shows that the definition using the Fisher information is weaker even when taking into considerations all the possible scalar products at the same time. This should be contrasted with the definition by Alicki, in which a single scalar product is used. The different between the two definitions arises in the way in which the coherences (in the eigenbasis of π) are handled. Still, it should be noticed that in both cases the evolution induced by the unitary part decouples from the dissipative dynamics. In fact, thanks to the normality of the generator, one has:

$$[\mathcal{U} + \mathcal{L}_\mathcal{D}, -\mathcal{U} + \mathcal{L}_\mathcal{D}] = 0 \qquad \Longrightarrow \qquad [\mathcal{U}, \mathcal{L}_\mathcal{D}] = 0\,. \tag{1.342}$$

This generic property can be used to further constrain $\mathcal{L}_\mathcal{D}$ in the case in which H is non-degenerate, allowing for the identification of Def. 1 and Def. 3 in this case.

If H is degenerate, on the other hand, we can provide the following structural characterisation of dissipators $\mathcal{L}_\mathcal{D}$ that are detailed balance according to Def. 3:

$$\mathcal{L}_\mathcal{D}[\rho] = \sum_{\omega,i} \lambda_i^\omega \left(A_i^\omega \, \rho \, (A_i^\omega)^\dagger - \frac{1}{2}\{(A_i^\omega)^\dagger A_i^\omega, \rho\} \right) + \mu_i^\omega \, B_i^\omega \, \rho^T (B_i^\omega)^\dagger \,, \tag{1.343}$$

where the following conditions are satisfied:

1. $(A_i^\omega)^\dagger = A_i^{-\omega}$ and $(B_i^\omega)^\dagger = B_i^{-\omega}$;
2. $\pi\, A_i^\omega\, \pi^{-1} = e^\omega\, A_i^\omega$ and $\pi\, B_i^\omega\, \pi^{-1} = e^\omega\, B_i^\omega$;
3. $\lambda_i^\omega = e^\omega\, \lambda_i^{-\omega}$ and $\mu_i^\omega = e^\omega\, \mu_i^{-\omega}$;
4. $\lambda_i^\omega \geq 0$ and $\sum_i \mu_i^\omega = 0\ \forall \omega$.

We defer the prove to App. A.5.3. It should be noticed that the additional term in Eq. (1.343) only acts on the off-diagonal elements (thanks to condition 4), so, even in this case, coherences evolve independently from the diagonal elements.

The discussions of this section, together with the one of about Markovianity, corroborates the claim that the Fisher information is an inherently dynamical quantity, a fact that is not completely acknowledged in the literature. Indeed, even if this relation was already hinted at by Chentsov theorem, it should be kept in mind that one can arrive to the same family of Fisher informations starting from the local expansion of contrast functions, which in principle has nothing to do with the nature of dissipative evolutions. The results presented in here, instead, show that the relation with the dynamics is so deep that one can even define physical evolutions as exactly the ones that contract the Fisher metric, showing the paramount importance of this family of scalar products.

Part II

Thermodynamics

Chapter 2

The first law

Experience teaches us that we live in a forgetful world, in which letters get lost and coffee gets cold. Fundamentally, though, we expect the information not to be erased from the universe, but rather to get hidden in subtle ways. The lost letter is just in the wrong post office, and the imperceivable increase in the temperature of the room witnesses the fact that the coffee was once warm. At the same time, despite the existence of this information, it is also easy to argue that this is of little help to any conceivable agent. This tension between a first principle that cannot be abandoned and its lack of applicability in real life was first noticed around two hundreds years ago in terms of energy: despite the pivotal role of the principle of conservation of energy in Newtonian mechanics, it is almost impossible in practice to get back the same energy we inputted in a system. The reconciliation of this paradox passed through the formulation of the first law of thermodynamics: this, on the one hand, recognises the existence of heat, a form of energy, or information, that is inaccessible to the operating agent; on the other, by defining work to be the energy that can be extracted from the system, it also imposes that all the energy injected in the system exactly equals the sum of heat and work, i.e., energy does not disappear. In this way, one accounts for the progressive degradation of the local information, while postulating its conservation on a global level. This is the perspective from which this chapter is written.

2.1 The thermodynamic framework

At the heart of the thermodynamic framework there are three main ingredients. The first one is the existence of a bath, a formally infinite system which has reached thermal equilibrium at a definite temperature. It is known from experience that thermalisation is a generic behaviour, but its justification is a difficult challenge of statistical mechanics, both technically and conceptually. Still, in most of thermodynamics one just assumes it, without having to dwell on the specific dynamics that led to the creation of a thermal bath.

The second ingredient is the definition of a system: this is the part of the universe under consideration. The state of the system can be arbitrary out of equilibrium and its Hamiltonian can depend on time through some externally driven parameters, as a magnetic field or the position of a piston. In order to highlight this dependence, it is customary to decompose the Hamiltonian as:

$$H_S(t) = \sum_{i=1}^{n} \lambda_t^i X_i \,, \tag{2.1}$$

where $\{\lambda_t^i\}$ are time-dependent scalar parameters, while $\{X_i\}$ are the associated time independent observables.

Lastly, one assumes the existence of an interaction between the system and bath that can be switched on and off at will, and which tends to thermalise the system at the temperature of the bath. The total Hamiltonian then reads:

$$H(t) = H_S(t) + gV + H_B \,. \tag{2.2}$$

where $H_S(t)$ and H_B are the Hamiltonians of system and bath, while gV is the interaction term. In this context, g is a scalar parameter quantifying the strength of the coupling. It is customary to require that $g \ll 1$, corresponding to the request that the interaction energy is small compared to the other energy scales (weak coupling regime). There are two main reasons to impose this constraint: first, in this way the total energy is additive, that is it equals the sum of the energy of the system and bath alone; secondly, whereas in general one needs to take into consideration at least a buffer region of the bath [43–45], in the weak coupling regime the thermodynamic description solely depends on the degrees of freedom of the system. These are two features that highly simplify the

thermodynamic treatment. Moreover, the weak coupling assumption is justified at least for macroscopic systems, as the interaction energy scales like the surface area, whereas the bulk energy scales like the volume. Unfortunately this argument does not apply to microscopic systems, for which surface and volumes have comparable contributions. Nonetheless, it is interesting to study the weak coupling regime first, in order to develop general insights on the theory without the technical and conceptual difficulties of having to deal with finite interactions.

2.2 Work and heat for classical systems

Now that the stage is set, we can speak about the energy balance. First, it should be noticed that the time dependency in the Hamiltonian implies a change in the energy of system and bath. In order for the energy of the universe to be conserved, we have to assume the existence of an external energy reservoir (i.e., a battery) from which energy can be extracted and inputted at will through the driving of the external parameters of the system Hamiltonian. We define work to be all the energy that gets exchanged with this external reservoir, leading to the equality:

$$\Delta U_S + \Delta U_B = w\,, \tag{2.3}$$

where $\Delta U_{S/B}$ is the change of energy in the system/bath, and w denotes the work. At the same time, we know that the driving is performed only on the system, so all the energy that is accumulated on the bath cannot be directly retrieved. This concept of unusable energy is identified with heat, leading to the definition:

$$\Delta U_B = Q\,, \tag{2.4}$$

where Q is the heat transferred to the environment[1]. Putting together Eq. (2.3) and Eq. (2.4), we obtain the equality:

$$\Delta U_S = w - Q\,. \tag{2.5}$$

This is the usual statement of the first law of thermodynamics.

In order to exemplify this relation, we consider a classical system undergoing a stochastic evolution. This is the most suited framework

[1]Here, and in the following, we only consider the case in which the system exchanges energy with the environment, but no matter

when treating microscopic classical systems and it is at the centre of the recently developed stochastic thermodynamics [46]. In this context, internal energy, work and heat become stochastic variables, by evaluating them at the trajectory level. For example, the probability associated with a change of internal energy ΔU_S is given by:

$$p(\Delta U_S) := \mu\left(\left\{x_t \,\middle|\, H_S(x_\tau) - H_S(x_0) = \Delta U_S\right\}\right), \tag{2.6}$$

where $\mu(\{x_t\})$ is the probability measure associated to the trajectory x_t, and τ is the final time of the driving. Since the internal energy is the only function of state in Eq. (2.5), this is also the only quantity for which the definition does not require stochastic integration or differentiation. Still, if we look at average quantities, we can split the internal energy as:

$$\langle \Delta U_S \rangle = \int \mathrm{d}x\, p(\Delta U_S = x) = \tag{2.7}$$

$$= \int \mathrm{d}\mu_\tau(\{x_\tau\})\; H_S(x_\tau) - \int \mathrm{d}\mu_\tau(\{x_0\})\; H_S(x_0) = \tag{2.8}$$

$$= \int_0^\tau \mathrm{d}t\; \left(\frac{\mathrm{d}}{\mathrm{d}t}\int \mathrm{d}\mu_\tau(\{x_t\})\; H_S(x_t)\right) = \tag{2.9}$$

$$= \sum_{i=1}^{n} \int \mathrm{d}\mu_\tau(\{x_t\}) \int_0^\tau \mathrm{d}t\; \dot{\lambda}_t^i\, X_i(x_t) + \tag{2.10}$$

$$+ \sum_{i=1}^{n} \int_0^\tau \mathrm{d}t\; \lambda_t^i\, \frac{\mathrm{d}}{\mathrm{d}t}\int \mathrm{d}\mu_\tau(\{x_t\})\; X_i(x_t)\,, \tag{2.11}$$

where $\mathrm{d}\mu_t$ is the probability density at time t. In Eq. (2.8) we used the fact that summing over all the possible outcomes allows to separate the average on the initial and final time, in Eq. (2.9) we exploited the smoothness of average trajectories to write the difference as the integral of a total derivative, and finally in the last equation we carried out the differentiation, and exchanged the order of integration in the first integral. Notice that these operations would be ill-defined without taking the average, as stochastic trajectories are almost everywhere non-differentiable.

When there is no driving in the external parameters, the change in internal energy equals minus the heat. For this reason, we can identify the term in Eq (2.11) with the average heat, as it is the only surviving term in the case that $\dot{\lambda}_t^i = 0$. Using the first law, then, we also have a definition for the average work, i.e., it corresponds to the integral in Eq (2.10).

The fact that the average change in internal energy equals the difference between the average energies at the end and at the beginning of the protocol (Eq. (2.8)) is almost trivial in this context. Another feature of the probability in Eq. (2.6) is that it can be rewritten as

$$p(\Delta U_S) = \sum_{E_\tau - E_0 = \Delta U_S} p(E_\tau|E_0)p(E_0)\,, \tag{2.12}$$

where we defined:

$$p(E_0) := \mu\left(\left\{x_t \,\middle|\, H_S(x_0) = E_0\right\}\right)\,; \tag{2.13}$$

$$p(E_\tau|E_0) := \mu\left(\left\{x_t \,\middle|\, H_S(x_\tau) = E_\tau \wedge H_S(x_0) = E_0\right\}\right)\,. \tag{2.14}$$

Eq. (2.12) means that the probability of obtaining a certain difference of internal energy is equal to the sum over all compatible energies of the probability of obtaining E_0 in the first measurement, multiplied by the probability of obtaining E_τ at the end conditioned on the first result.

Equations (2.8) and (2.12) are of course trivial statements in the context of classical thermodynamics. For quantum system, instead, the situation changes drastically.

2.3 Work is not an observable

In the quantum mechanics of isolated system it is known that for each measurable property of the system there exists a self-adjoint operator corresponding to it. Naively, then, one would expect to be able to associate an observable to the work output. This can be identified by imposing the validity of Eq. (2.8), i.e., that the change of average energy at the beginning and at the end of the protocol equals the average work (when considering system and bath together). In formulae, this reads:

$$\langle \hat{w} \rangle := \mathrm{Tr}\left[U_\tau\, \rho\, U_\tau^\dagger H_\tau\right] - \mathrm{Tr}\left[\rho\, H_0\right] = \mathrm{Tr}\left[\rho\left(U_\tau^\dagger H_\tau U_\tau - H_0\right)\right]\,, \tag{2.15}$$

where ρ is the state of the system and bath, $H_{0/\tau}$ is the total Hamiltonian at the beginning/end of the driving, and U_τ is the unitary describing the overall dynamics. Requiring that Eq. (2.15) holds for any initial state ρ constrains $\hat{w}$, the operator of work, to the form $\hat{w} := \left(U_\tau^\dagger H_\tau U_\tau - H_0\right)$.

Unfortunately, it is quite easy to see that this definition is unsatisfactory. Notice, in fact, that since by definition the work in this context

equals the difference in energy, one would also expect that the probability of obtaining an energy value E_0 at the beginning of the protocol would be connected to the one of obtaining E_τ at the end by the relation

$$p(w) = \sum_{E_\tau - E_0 = w} p_\tau(E_\tau|E_0) p_0(E_0)\,, \tag{2.16}$$

where $p(w)$ is the probability of the work output w and p_t the probability distribution for the energy at time t. Now consider the case in which $\hat{w}$ has a 0 eigenvalue, associated to the eigenvector $|\psi_0\rangle$. Then, initialising the system in the state $|\psi_0\rangle\langle\psi_0|$ would induce a delta centred in zero for the work probability, $p(w) = \delta(w)$. Together with Eq. (2.16), this implies that $p_0(E_0) = p_\tau(E_0)$. In particular, all the moments should be the same. This cannot be the case in general: consider for example a qubit system with Hamiltonian $H_0 = \sigma_x$ and $U_\tau^\dagger H_\tau U_\tau = (\mathbb{1} - \sigma_z + \sigma_x)$. The work operator has $|0\rangle$ as a zero eigenvector, but it is straightforward to verify that:

$$\mathrm{Tr}\left[\,|0\rangle\langle 0|\, U_\tau^\dagger\, H_\tau^n\, U_\tau\right] \neq \mathrm{Tr}\left[\,|0\rangle\langle 0|\, H_0^n\right]\,, \tag{2.17}$$

already for $n = 3$, for example. This means that the statistics of the energy were altered despite no energy was exchanged [47], contradicting the principle of conservation of energy. For this reason, this definition of work has to be discarded as unphysical.

One might be surprised by the absence of an operator of work. In fact, in quantum mechanics we are used to associate measurable properties to self-adjoint operators. There is however another notable example in which a measurable quantity is not associated to any observable: the Berry phase. There is a common cause to this: as it is argued in [48], since the Berry phase arises from the dependence of the Hamiltonian on some changing parameters, the system cannot be truly isolated (for example, in the case of molecular Berry phase, the external parameter is given by the position of the heavy nuclei in the Born-Oppenheimer approximation). In the same way, in thermodynamics the driving is modulated by an external battery, i.e., an effectively semi-classical object external to the system. In principle, taking into account the dynamics of these extra degrees of freedom would eliminate the time dependency in the Hamiltonian so that each measurable property would correspond to a self-adjoint operator, but this would severely deform our description of the system. In fact, as for molecules the dynamics of the nuclei should be neglected

due to the huge separation of time-scales between these and the electronic degrees of freedom, in the same way the ability of the external battery to deterministically store and provide energy implies that it should be treated semi-classically. Even if one could in principle solve the exact equations of motion, these would be hardly very informative, as their complicated expression would hide important features of the dynamics. This justifies the use of time-dependent Hamiltonians in this context.

Going back to the definition of work, since starting directly from Eq. (2.8) is problematic, one has to find another founding principle to follow. This was identified in [49] with the operational definition of Eq. (2.12). In this context, in order to define the work, one first needs to measure the initial energy, obtaining the value $E_0^{(i)}$ with probability:

$$p(E_0^{(i)}) = \mathrm{Tr}\left[\Pi_0^{(i)}\,\rho\right] , \tag{2.18}$$

where $\Pi_0^{(i)} := \left|E_0^{(i)}\right\rangle\!\left\langle E_0^{(i)}\right|$ is the projector into the $E_0^{(i)}$ eigenspace. The measurement changes the state of the system to $\Pi_0^{(i)}$. At this point, the system is evolved unitarily according to U_τ. Finally, one measures again the energy, obtaining $E_\tau^{(j)}$ with probability:

$$p(E_\tau^{(j)}|E_0^{(i)}) = \mathrm{Tr}\left[\Pi_\tau^{(j)}\,U_\tau\,\Pi_0^{(i)}\,U_\tau^\dagger\right] , \tag{2.19}$$

where $\Pi_\tau^{(j)} := \left|E_\tau^{(j)}\right\rangle\!\left\langle E_\tau^{(j)}\right|$ is the projector into the $E_\tau^{(j)}$ eigenspace. The total probability of extracting work w is then given by:

$$p(w) = \sum_{E_\tau^{(j)}-E_0^{(i)}=w} p(E_\tau^{(j)}|E_0^{(i)})p(E_0^{(i)}) = \tag{2.20}$$

$$= \sum_{E_\tau^{(j)}-E_0^{(i)}=w} \mathrm{Tr}\left[\Pi_\tau^{(j)}U_\tau\Pi_0^{(i)}U_\tau^\dagger\right]\mathrm{Tr}\left[\Pi_0^{(i)}\rho\right] = \tag{2.21}$$

$$= \sum_{E_\tau^{(j)}-E_0^{(i)}=w} \mathrm{Tr}\left[\Pi_\tau^{(j)}\,U_\tau\Pi_0^{(i)}\,\rho\,\Pi_0^{(i)}U_\tau^\dagger\right] = \tag{2.22}$$

$$= \sum_{E_\tau^{(j)}-E_0^{(i)}=w} \mathrm{Tr}\left[U_\tau^\dagger\,\Pi_\tau^{(j)}\,U_\tau\,\mathcal{D}_0^{(i)}[\rho]\right] , \tag{2.23}$$

where in the last equation we introduced the dephasing operator $\mathcal{D}_0^{(i)} := \Pi_0^{(i)}\,\rho\,\Pi_0^{(i)}$.

The expression of the probability distribution in Eq. (2.23) highlights a limitation of this scheme: all the coherences contained in the state ρ are destroyed by the initial measurement. Still, this definition of the work output is particularly appealing for its clear operational meaning, in addition to having a well defined classical limit [50] and satisfying fluctuations theorems (see Sec. 3.2). For these reasons, the procedure just presented has become the most commonly used definition of fluctuating work, known under the name of TPM (two point energy measurement) scheme.

Another interesting property of the TPM scheme is that it can be expressed in terms of well-known information theoretic quantities. In particular, for a system initially in thermal equilibrium (i.e., whose density matrix is given by $\pi_\beta(H) := \frac{e^{-\beta H}}{\mathcal{Z}_H}$, where $\mathcal{Z}_H$ is the partition function $\mathcal{Z}_H := \mathrm{Tr}\left[e^{-\beta H}\right]$) one can write the cumulant generating function of the random variable $(-\beta w)$ as:

$$K^{-\beta w}(\lambda) = \log \int \mathrm{d}w\; p(w) e^{-\beta w \lambda} = \tag{2.24}$$

$$= \log \sum_{\substack{w,\\ E_\tau^{(j)} - E_0^{(i)} = w}} \mathrm{Tr}\left[\Pi_\tau^{(j)}\, U_\tau \Pi_0^{(i)}\, U_\tau^\dagger\right] \frac{e^{-\beta E_0^{(i)}}}{\mathcal{Z}_{H_0}}\, e^{-\beta w \lambda} = \tag{2.25}$$

$$= \log \sum_{i,j} \mathrm{Tr}\left[\Pi_\tau^{(j)}\, U_\tau \Pi_0^{(i)}\, U_\tau^\dagger\right] \frac{e^{-\beta(1-\lambda)E_0^{(i)}}}{\mathcal{Z}_{H_0}^{1-\lambda}} \frac{e^{-\beta\lambda\, E_\tau^{(j)}}}{\mathcal{Z}_{H_\tau}^{\lambda}} \frac{\mathcal{Z}_{H_\tau}^{\lambda}}{\mathcal{Z}_{H_0}^{\lambda}} = \tag{2.26}$$

$$= \lambda \log \frac{\mathcal{Z}_{H_\tau}}{\mathcal{Z}_{H_0}} + \log \mathrm{Tr}\left[\pi_\beta(H_\tau)^\lambda \left(U_\tau \pi_\beta(H_0) U_\tau^\dagger\right)^{1-\lambda}\right] = \tag{2.27}$$

$$= -\lambda\beta\Delta F + (\lambda - 1)\, S_\lambda(\pi_\beta(H_\tau) || U_\tau \pi_\beta(H_0) U_\tau^\dagger)\,. \tag{2.28}$$

Notice that in Eq. (2.26) we expressed explicitly the value of w and exploited the sum over it to decouple $E_\tau^{(j)}$ from $E_0^{(i)}$. Moreover, in Eq. (2.27) we used the definition $\pi_\beta(H_0) := \sum_{E_0^{(i)}} \frac{e^{-\beta\, E_0^{(i)}}}{\mathcal{Z}_{H_0}} \Pi_0^{(i)}$ (and similarly for $\pi_\beta(H_\tau)$) to express the sum in terms of thermal states. Finally, in Eq. (2.28) we defined the difference of free energy ΔF as $\beta\Delta F := -\log \frac{\mathcal{Z}_\tau}{\mathcal{Z}_0}$, and we remind the reader about the definition of the λ-Rényi divergences (see Eq. (1.168)):

$$S_\lambda(\rho||\sigma) := \frac{1}{\lambda - 1} \log \mathrm{Tr}\left[\rho^\lambda \sigma^{1-\lambda}\right]\,. \tag{2.29}$$

Then, Eq. (2.28) shows how the statistics of work in the TPM scheme are directly connected to distinguishability between the final thermal state and the evolved state, as quantified by the λ-Rényi divergences. This gives an insight on the nature of dissipation and it will be further explored in Chapter 4.

In fact, consider a general contrast function $H_g(\rho||\sigma)$. Following the same steps as above we have that:

$$H_g(U_\tau \pi_\beta(H_0) U_\tau^\dagger \,||\, \pi_\beta(H_\tau)) =$$

$$= \sum_{i,j} \frac{e^{-\beta E_0^{(i)}}}{\mathcal{Z}_{H_0}} \, g\left(\frac{e^{-\beta(E_\tau^{(j)} - E_0^{(i)})}}{\mathcal{Z}_{H_\tau} \mathcal{Z}_{H_0}^{-1}} \right) \left| \left\langle E_\tau^{(j)} \right| U_\tau \left| E_0^{(i)} \right\rangle \right|^2 = \tag{2.30}$$

$$= \sum_{\substack{w, \\ E_\tau^{(j)} - E_0^{(i)} = w}} g(e^{-\beta(w - \Delta F)}) \, \mathrm{Tr}\left[\Pi_\tau^{(j)} \, U_\tau \Pi_0^{(i)} \, U_\tau^\dagger \right] \, \frac{e^{-\beta E_0^{(i)}}}{\mathcal{Z}_{H_0}} = \tag{2.31}$$

$$= \int \mathrm{d}w \; p(w) \, g(e^{-\beta(w - \Delta F)}) \, , \tag{2.32}$$

where in the first line we used the coordinate expression from Eq. (1.40). The quantity $\beta^{-1}\Sigma = w - \Delta F$ is dubbed entropy production, and quantifies the dissipation along the protocol (for more details, see Sec. 3.2). Then, thanks to the identity in Eq. (2.32), one can interpret different contrast functions between the evolved state and the thermal state at the final time as different functionals of the dissipation, where the probability distribution is the TPM one.

Finally, it should be noticed that for divisible dynamics we can obtain an expression akin to Eq. (2.11). In fact, the change of total energy for systems starting at equilibrium is given by:

$$\langle \Delta U \rangle = \mathrm{Tr}\left[H_\tau \, \rho_\tau \right] - \mathrm{Tr}\left[H_0 \, \rho_0 \right] = \int_0^\tau \mathrm{d}t \; \frac{\mathrm{d}}{\mathrm{d}t} \, \mathrm{Tr}\left[\rho_t H_t \right] \, , \tag{2.33}$$

where we denote by ρ_t the effective state of the system and bath at time t. Since the Hamiltonian of the bath is fixed, we have that $\langle \Delta U \rangle = \langle \Delta U_S \rangle$, as the only change in energy arises from the driving in the system Hamiltonian. Then, the total derivative above can be split as:

$$\langle \Delta U_S \rangle = \int_0^\tau \mathrm{d}t \; \mathrm{Tr}\left[\rho_t \, \dot{H}_t \right] + \int_0^\tau \mathrm{d}t \; \mathrm{Tr}\left[\dot{\rho}_t \, H_t \right] = \tag{2.34}$$

$$= \langle w \rangle - \langle Q \rangle \, , \tag{2.35}$$

where we identified the first integral with the average work, and the second with the heat. Indeed, if the Hamiltonian does not vary with time, the change of energy only arises from the exchange of heat. This argument, completely analogous to the one in Eq. (2.11) justifies the identification of the work with the first integral.

Wrapping up, the TPM scheme is well justified, satisfies most of the properties one would expect from a work distribution and has a neat connection with informational theoretical quantities. Unfortunately though, it also suffers from a severe drawback: if coherences are present in the initial state, the average work is not equal in general to the average change of energy (that is Eq. (2.15) does not hold!). Compared to the plethora of other nice features of the TPM statistics, this can appear as a small detail, so that one could naively hope for an extension of the TPM scheme which allows to recover Eq. (2.15) without radically changing the measurement procedure. Finding such a scheme would finally give an unproblematic definition of work in the quantum regime. However this is impossible.

2.3.1 There is no quantum analogue of the classical work probability

The failure of the two most natural proposals for a definition of quantum work distribution raises the question whether a satisfactory answer exists at all. This doubt turns out to be well motivated. In fact, if one associates to a thermodynamic protocol $\mathcal{P}$ a general work distribution $p(w|\mathcal{P})$, the following no-go theorem can be proved:

Theorem 15. *There is no function $p(w|\mathcal{P})$ associated to the protocol $\mathcal{P}$ that satisfies the following three conditions:*

1. *(Linear probability) $p(w|\mathcal{P})$ is a probability distribution linear under convex mixing of different protocols. This means that by randomly choosing one protocol between $\mathcal{P}_1$ and $\mathcal{P}_2$ with probability α, the work distribution changes as*

$$p(w|\alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2) = \alpha\, p(w|\mathcal{P}_1) + (1-\alpha)\, p(w|\mathcal{P}_2)\,;$$

2. *(Classicality 1) $p(w|\mathcal{P})$ agrees with the TPM scheme on incoherent states (with respect to the initial Hamiltonian);*

3. *(Classicality 2) The average energy change equals the average work.*

This theorem, proven in [51], shows that Condition 2 and 3, which correspond to Eq. (2.12) and Eq. (2.8), cannot be satisfied together by any probability distribution.

Proof. Thanks to Gleason's theorem, if condition 1 holds, there exists a POVM $\{M^w\}_{w\in\mathbb{R}}$ which depends on the protocol $\mathcal{P}$, but not on the initial state ρ, such that $p(w|\mathcal{P}) = \mathrm{Tr}\left[\rho M^w\right]$ [51]. Then, we can compare it with the POVM corresponding to the TPM scheme, given by:

$$M^w_{\mathrm{TPM}} := \sum_{i,j} \delta(w - (E^{(j)}_\tau - E^{(i)}_0))\, p(E^{(j)}_\tau|E^{(i)}_0)\, \Pi^{(i)}_0\,, \tag{2.36}$$

where $p(E^{(j)}_\tau|E^{(i)}_0)$ is the one defined in Eq. (2.19).

Condition 2 imposes that M^w agrees with M^w_{TPM} on the set of density matrices with no coherences, i.e., the convex mixture of states of the form $\rho^{(k)} := \Pi^{(k)}_0$. This allows to completely characterise M^w. In fact, first it should be noticed that whenever w is not an energy gap then $M^w \equiv 0$, due to the δ-function in Eq. (2.36). This allows to index the work POVM as $M^{i,j} := M^{(E^{(i)}_\tau - E^{(j)}_0)}$. Operating over the density matrices of form $\rho^{(k)}$ one obtains:

$$\left\langle E^{(k)}_0 \right| M^{i,j} \left| E^{(k)}_0 \right\rangle = \left\langle E^{(k)}_0 \right| M^{i,j}_{\mathrm{TPM}} \left| E^{(k)}_0 \right\rangle = \delta^i_k\, p(E^{(j)}_\tau|E^{(i)}_0)\,. \tag{2.37}$$

Since there is only one non-zero element on the diagonal of $M^{i,j}$, the positivity condition for POVMs implies that all the off-diagonal terms are zero, so that M^w is specified completely. Summarising, condition 1 implies that there exists a POVM $\{M^w\}$, and condition 2 ensures that $M^w = M^w_{\mathrm{TPM}}$ for all w.

On the other hand condition 3 reads:

$$\sum_w w\, \mathrm{Tr}\left[\rho M^w\right] = \mathrm{Tr}\left[\rho \left(U^\dagger_\tau H_\tau U_\tau - H_0\right)\right]\,. \tag{2.38}$$

Due to the genericity of ρ this is equivalent to asking $\sum_w w\, M^w = \left(U^\dagger_\tau H_\tau U_\tau - H_0\right)$. This condition cannot be always satisfied by the TPM scheme. For example, consider a qubit with initial Hamiltonian $H_0 = \epsilon\, |1\rangle\langle 1|$ and final Hamiltonian $H_1 = \epsilon'\, |1\rangle\langle 1|$, which undergoes the evolution $U = |0\rangle\langle +| + |1\rangle\langle -|$ (the Hadamard gate). In this case $\sum_w w\, M^w = \frac{1}{2}\left(\epsilon'\, |0\rangle\langle 0| + (\epsilon' - 2\epsilon)\, |1\rangle\langle 1|\right)$, while $\left(U^\dagger_\tau H_\tau U_\tau - H_0\right) = \epsilon'\, |-\rangle\langle -| - \epsilon\, |1\rangle\langle 1|$. Hence, condition 3 cannot be satisfied. This concludes the proof. □

Theorem 15 shows how straightforward properties that hold in classical thermodynamics are incompatible in the quantum regime. There is however a loophole in the argument, as noted in [52]. In fact, assumption 1, which is a formulation of preparation non-contextuality, is not justified for quantum systems (for a brief review about the concept of contextuality see App. B). Indeed, it can be proven that:

Theorem 16. *If $p(w|\mathcal{P})$ satisfies condition 1 and 2 of Theorem 15 then there exists a universally non-contextual ontological model which reproduces it.*

Proof. In the proof of Theorem 15 it is shown that condition 1 and 2 together imply the existence of a POVM set, i.e., the TPM one, with the property $p(w|\mathcal{P}) = \mathrm{Tr}\,[\rho\, M^w_{\mathrm{TPM}}]$. This induces a non-contextual ontological model. In fact, take the ontological space Λ to be the space of the initial energy eigenstates, and the preparation and measurement procedures are given by:

$$P(\lambda|\rho) := \langle\lambda|\,\rho\,|\lambda\rangle\ ; \tag{2.39}$$

$$P(w|\lambda,\mathcal{P}) := \mathrm{Tr}\,[\,|\lambda\rangle\langle\lambda|\,M^w_{\mathrm{TPM}}]\ . \tag{2.40}$$

Then, as the measurement output can be computed from the equation:

$$p(w|\mathcal{P}) = \int_\Lambda \mathrm{d}\lambda\ P(\lambda|\rho)\,P(w|\lambda,\mathcal{P})\,, \tag{2.41}$$

that is, this construction gives the required non-contextual ontological model. □

Theorem 16 strongly suggests that in order to probe genuinely quantum effects one has to drop the assumption that the work distribution is a probability density. This can be done by extending the TPM scheme from projective measurements to weak ones.

2.3.2 Work quasiprobability from weak measurements

In the TPM scheme the initial dephasing of the state erases any effects arising from initial coherences. For this reason, it is interesting to explore what happens when one considers minimally disturbing measurements. This naturally leads to the study of weak measurements: in this context, the initial energy measurement is carried out by coupling the system to

a pointer, which is then projectively measured. As explained in Sec. B.3, it is customary to associate to the pointer a Gaussian state:

$$\big|\psi_{\nearrow}\big\rangle = \frac{1}{\sqrt[4]{\pi s^2}} \int \mathrm{d}x \; e^{-\frac{x^2}{2s^2}} \; |x\rangle \,, \tag{2.42}$$

where the spread s of the wave function parametrises how precise the measurement is, allowing one to continuously switch from a projective measurement ($s \to 0$) to a weak measurement ($s \to \infty$).

The measurement protocol of the work is then given by the following steps: (*i*) first, the pointer interacts with the system through the Hamiltonian $H_{\rm int} = g(t)\, \Pi_0^{(i)} \otimes P$, so to gather information about $\rho\, \Pi_0^{(i)}$; (*ii*) the pointer state is projectively measured in the position basis; (*iii*) the system is decoupled from the pointer, and evolved according to the protocol $\mathcal{P}$; (*iv*) finally, $\Pi_\tau^{(j)}$ is measured projectively. For simplicity we assume non degenerate energy gaps, so that one can associate to any work value w a unique gap $E_\tau^{(j)} - E_0^{(i)}$. This procedure should be compared with the one in Sec. B.3, where $\Pi^{(i)} := \Pi_0^{(i)}$ and $\tilde{\Pi}^{(j)} = U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau$, where the additional unitary accounts for the presence of the transformation defined by $\mathcal{P}$.

Then, following the steps of Sec. B.3, we define $\rho^{11} := \Pi_0^{(i)}\, \rho\, \Pi_0^{(i)}$ and $\rho^{01} := (\Pi_0^{(i)})^\perp\, \rho\, \Pi_0^{(i)}$. This allows to express the average position of the pointer upon post-selection as (see Eq. (B.21)):

$$\langle X \rangle_{|j} = \frac{\mathrm{Tr}\Big[U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\, \rho^{11}\Big] + e^{-1/4s^2} \mathrm{Re}\Big[\mathrm{Tr}\Big[U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\, \rho^{01}\Big]\Big]}{\mathrm{Tr}\Big[U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\, \rho\Big] - 2\,(1 - e^{-1/4s^2}) \mathrm{Re}\Big[\mathrm{Tr}\Big[U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\, \rho^{01}\Big]\Big]} \,. \tag{2.43}$$

We can now explore the two limits $s \to 0/\infty$. In the first case, i.e., in the limit $s \to 0$, the pointer starts in an eigenstate of the position operator, and Eq. (2.43) reduces to:

$$\lim_{s\to 0} \langle X \rangle_{|j} = \frac{\mathrm{Tr}\Big[U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\, \Pi_0^{(i)}\, \rho\, \Pi_0^{(i)}\Big]}{q_j} = \frac{p_{\rm TPM}(w|\mathcal{P})}{q_j} \,. \tag{2.44}$$

Since the first measurement becomes effectively projective, we obtain the statistics given by the TPM scheme.

In the opposite limit, we have a weak measurement of the initial energy value. Then, the average position conditioned on the output j is

given in this case by:

$$\lim_{s\to\infty}\langle X\rangle_{|j}=\frac{\mathrm{Re}\Big[\mathrm{Tr}\left[U_\tau^\dagger\,\Pi_\tau^{(j)}\,U_\tau\,\Pi_0^{(i)}\,\rho\right]\Big]}{\mathrm{Tr}\left[\tilde{\Pi}_\tau^{(j)}\,\rho\right]}=\frac{p_{\mathrm{weak}}(w|\mathcal{P})}{q_j}\,. \tag{2.45}$$

It is interesting to point out the similarity with the TPM distribution. This definition of work was first introduced by Allahverdyan in [53], and it will be denoted here by the acronym WWS (weak work statistics). The quantity in Eq. (2.45) was already studied in the literature for its connection to non-classicality: it takes the name of generalised weak value and, interestingly, it can also be negative (in which case it is called anomalous, as explained in Sec. B.3). This corresponds to a negativity in the WWS, making $p_{\mathrm{weak}}(w|\mathcal{P})$ a quasiprobability. Hence the WWS does not satisfy condition 1 of Theorem 15. In fact, this must be the case. First, notice that for incoherent states one has:

$$p_{\mathrm{weak}}(w|\mathcal{P})=\mathrm{Re}\Big[\mathrm{Tr}\left[U_\tau^\dagger\,\Pi_\tau^{(j)}\,U_\tau\,\Pi_0^{(i)}\,\rho\right]\Big]= \tag{2.46}$$

$$=\mathrm{Tr}\left[\Pi_\tau^{(j)}\,U_\tau\,\Pi_0^{(i)}\rho\,\Pi_0^{(i)}\,U_\tau^\dagger\right]=p_{\mathrm{TPM}}(w|\mathcal{P})\,, \tag{2.47}$$

so that condition 2 is satisfied. Moreover, the average work is given by:

$$\langle w\rangle_{\mathrm{weak}}=\int\mathrm{d}w\;w\,p_{\mathrm{weak}}(w|\mathcal{P})= \tag{2.48}$$

$$=\sum_{E_\tau^{(j)},E_\tau^{(i)}}(E_\tau^{(j)}-E_\tau^{(i)})\,\mathrm{Re}\Big[\mathrm{Tr}\left[U_\tau^\dagger\,\Pi_\tau^{(j)}\,U_\tau\,\Pi_0^{(i)}\,\rho\right]\Big]= \tag{2.49}$$

$$=\mathrm{Tr}\left[\rho\,U_\tau^\dagger H_\tau U_\tau\right]-\mathrm{Tr}\left[\rho\,H_0\right]\,, \tag{2.50}$$

where one uses the two identities $\sum_k\Pi_{0/\tau}^{(k)}=\mathbb{1}$ and $\sum_k\Pi_{0/\tau}^{(k)}\,E_{0/\tau}^{(k)}=H_{0/\tau}$ to pass from Eq. (2.49) to Eq. (2.50). Hence condition 3 is also satisfied, so that condition 1 cannot hold. This shows that the assumption of Theorem 15 are tight, i.e., by dropping any of them one can explicitly construct a counterexample.

Since Theorem 16 heavily depends on condition 1, one could wonder if dropping it would allow to detect genuinely quantum phenomena. This turns out to be true, and it motivates the interest in the WWS [52]:

Theorem 17. *If the probability of work according to WWS is negative for some work value, then it always exists a pointer state with spread s large enough so that there is no measurement non-contextual, outcome deterministic ontological model reproducing the statistics of work.*

In fact, this result is a particular case of Thm. 26 in App. B, so it doesn't need an additional proof.

As it was shown in Eq. (2.47), in the limit in which there is no coherence in the initial energy basis, the WWS reduces to the TPM statistics. For this reason, one could hope to recover (at least up to corrections depending on the coherence) the information theoretic interpretation assigned to the TPM scheme. For example, assuming the state of the system to be initially given by $\rho = \pi_\beta(H_0) + \chi$, where χ has only off-diagonal contributions, one can write the cumulant generating function for the WWS as:

$$\begin{aligned} K^{-\beta w}(\lambda) &= \log \int \mathrm{d}w\; p(w) e^{-\beta w \lambda} = \\ &= -\lambda\beta\Delta F + (\lambda - 1)\, S_\lambda(\pi_\beta(H_\tau) || U_\tau \pi_\beta(H_0) U_\tau^\dagger) + \\ &\quad + \log\left(1 + \mathrm{Re}\left[\frac{\mathrm{Tr}\left[\pi_\beta(H_\tau)^\lambda U_\tau \chi\, \pi_\beta(H_0)^{-\lambda} U_\tau^\dagger\right]}{\mathrm{Tr}\left[\pi_\beta(H_\tau)^\lambda U_\tau \pi_\beta(H_0)^{1-\lambda} U_\tau^\dagger\right]}\right]\right), \end{aligned} \tag{2.51}$$

where we omit the lengthy but straightforward calculations. This shows that a result similar to Eq. (2.28) holds also for the WWS, but at the cost of an additional term depending on the coherences.

The discussion in this chapter shows the difficulty of defining the statistics of work for quantum systems, or rather a probability distribution able to manifest purely quantum behaviours. Indeed, the most used measurement scheme, the TPM one, always admits a non-contextual ontological model (Thm. 16), and only if one drops the condition of positivity of the probability it is possible to probe genuinely quantum phenomena (Thm. 17). Still, one of the merits of quantum thermodynamics is to highlight the connection between statistical mechanics and information theory which, even if no quantum phenomena were to be found, is enough to justify the interest in the field.

Chapter 3

The second law

Thermodynamics was originally devised to answer practical problems about the functioning of macroscopic classical engines. It is for this reason absolutely remarkable that, despite its initial narrow scope of definition, the second law became over time one of the most overreaching principles of physics, with applications ranging from microscopic quantum systems to black holes. As it encodes a preferred direction of time, seemingly in contradiction with the time reversal symmetry of microscopic dynamics, the second law has historically had a more controversial existence than the first. When Boltzmann laid down its statistical mechanical foundation through his famous H theorem, this attracted criticisms from personalities as Zermelo and Poincaré. Even more so, its extension to the microscopic regime, when only a handful of particles are involved, or to biological systems, where the order of the system seems to increase over time, was highly debated. Despite all of this, after almost two hundreds years we can say that the second law resisted stolidly. As pointed out in a famous quote by Eddington[1], its conceptual power is at the heart of its impermeability to apparent violations. In fact, the second law dictates how information gets transferred from a system to another, and it shows how in this passage information can become irretrievable (despite its global conservation). In order to make this point precise, the abstract

[1] *"The law that entropy always increases, holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations — then so much the worse for Maxwell's equations. If it is found to be contradicted by observation — well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation." [54]*

mathematical concept of information has to be given a physical ground. This task naturally leads to the analysis of one of the most fertile thought experiment ever designed: the Maxwell's demon.

3.1 Information is physical: the Maxwell's demon

In 1866, at the age of 22, Boltzmann published a tentative foundation of equilibrium thermodynamics starting from the Hamiltonian principle. A very similar construction appeared five years later, this time authored by Clausius, who had already been a professor for more than 20 years. The consequent debate between the two about the paternity of the idea reached Scotland, as it was commented by Maxwell in a letter to Tait: *"It is rare sport to see those learned Germans contending for the priority of the discovery that the 2nd law of thermodynamics is the Hamiltonsche Princip [...] The Hamiltonsche Princip, the while, soars along in a region unvexed by statistical considerations, while the German Icari flap their waxen wings in nephelococcygia amid those cloudy forms which the ignorance and finitude of human science have invested with the incommunicable attributes of the invisible Queen of Heaven."* [55]. For him trying to give a foundation to the second law without resting on probabilistic arguments was clearly total nonsense (ironically, in the meantime, Boltzmann had already amended his mistake, and had already started laying down the foundations of statistical mechanics).

In the same spirit, in order to highlight this misconception, Maxwell formulated one of the most famous thought experiments in physics: in another letter to Tait dated 1867 he proposed a finite being with the ability to detect and operate on the dynamics of a gas at the level of single atoms, and that, by doing so, could reversibly induce a temperature gradient in a closed box. In particular, he considered the following set-up: a box containing a uniform distribution of gas is divided in two partitions by a wall with a small aperture that can be opened or closed in a reversible manner and without exerting any work. Then, the finite being could be instructed to open the aperture whenever it would detect a fast particle moving from right to left, and do the same for slow particles moving left to right. As temperature is connected with the average kinetic energy of the gas, repeating this procedure many times would induce a temperature gradient at the cost of zero work, in contradiction with the second law of thermodynamics.

The nickname of demon was promptly popularised by Thomson (Lord Kelvin) [56], whereas Maxwell had the care in another letter to specify that this neat-fingered being could also just be thought as a well designed valve [55]. For Maxwell the possibility of violating the second law highlighted its statistical nature and, most importantly, its independence from mechanical laws.

The ability of the second law to stir controversy, though, created a long and heated debate around Maxwell's demon, leading to a continuous quest for loopholes in the argument. First, it was unclear whether an intelligent being had to operate the system in order for a violation to appear, and, most importantly, if empirical laws of thermodynamics could be applied to systems capable of reasoning (i.e., biological systems). As a matter of fact, even Smochulowski, who proposed a possible physical realisation of Maxwell's demon and showed how this would fail, was tricked by this argument. In his minimalistic model, the aperture between the two partitions of the box is given by a door that can open only in one direction, and that keeps bouncing back thanks to the action of a spring. This system could induce a pressure gradient, if only the spring didn't heat up: the successive atomic collisions with the door would make it fluctuate so much that it would quickly become unusable. In order to continue to operate it some kind of reset mechanism dampening the oscillations has to be used, but this inevitably leads to dissipation. Hence, Smochulowski argued that mechanical beings are prevented by thermal fluctuations from violating the second law, but, interestingly, he left open this possibility for "intelligent beings".

In this context, the landmark work *"On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings"* by Szilárd (dated 1929, seven years before the formalisation of Turing machines) is almost revolutionary. His idea was to show that even mechanical systems could reproduce the main ingredients needed from an intelligent being to violate the second law: in particular, information acquisition, procession and storage. Famously, he explored these phenomena on a single atom gas[2]. He showed how ignoring the information acquired during the initial measurement would prevent any violation of the second law, demonstrating how the use of information was a necessary ingredient. Finally, he argued that, in order for the second law to hold, there must be an entropy increase of at least $k_B T \log 2$ associated

[2]He also studied them on a gas made of different types of molecules, but this approach is less innovative than the first and, for this reason, less well known.

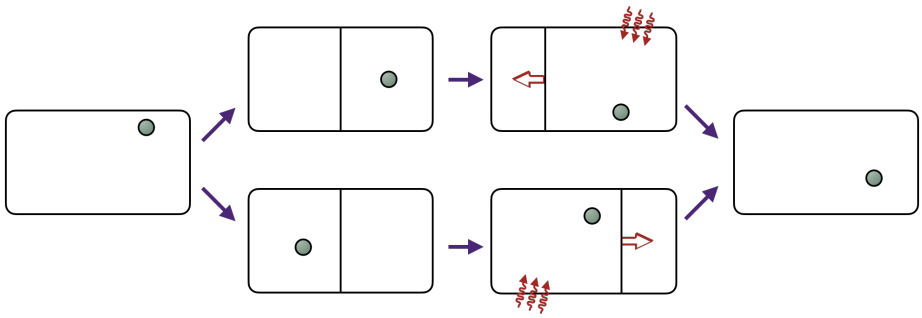


Figure 3.1: A simple depiction of a Szilárd engine. A single molecule is at equilibrium inside a box. At the beginning of the cycle, a partition is inserted adiabatically (i.e., without dissipation) and the demon measures whether the molecule is on the right or on the left of the partition. This step can be realised completely mechanically by coupling the partition to a switch that changes its state depending on whether there are collisions coming from the right or from the left. At this point the partition is coupled to a mechanism that lets it move only in one direction, and its motion raises a weight. Finally, the partition is adiabatically removed. This procedure allows to extract work from a single heat source, in contradiction with the second law of thermodynamics. There is a loophole in the argument though: the switch stores a bit of information about the initial state of the system, so the protocol is not truly cyclic.

with binary measurements, or to the memory of the outcome, or to the erasure thereof.

With the hindsight of modernity, all these intuitions are incredibly ahead of time: first, he noticed how the process of information processing and decision making could be automatised; second, the reduction of information units to binary outcomes resembles closely the bits of digital computers; finally, his radical stance on the second law made him recognise the physicality of information and its effects on entropy production.

Szilárd did not specify exactly which part of the information processing should be associated with dissipation. The first to extensively study this issue was Brillouin, who tried to reduce the entropy increase to the measurement procedure connected with electromagnetic interactions [57]. This approach was rebutted by Bennet, who showed that entropy increase during measurement is not a general feature, so it is not enough for saving the second law [58].

Moreover, not just information acquisition, but also information processing could be made completely dissipationless: in fact, Bennet also proved that reversible computation is Turing complete, and that it can efficiently simulate irreversible computers [59] (for a proof of principle physical implementation using billiard balls see the work of Toffoli and Fredkin [60]). Hence, also the information processing and decision making associated with a demon can be performed without necessarily producing entropy.

The last possible loophole was then connected with the information storage. In particular, the progressive identification of logical irreversibility with the thermodynamic one suggested a natural candidate that could save the second law: the only process that is by nature irreversible, that is information erasure. This was made precise by Landauer, who showed that there is a fundamental physical bound on the minimal heat that has to be dissipated in order to erase a bit of information, given by $\Delta S = k_B T \log 2$ [61]. The famous Landauer's bound then agrees with the analysis carried out by Szilárd, finally giving the long awaited exorcism of Maxwell's demon.

3.1.1 An exactly solvable system

The discussion above can be made more concrete by analysing an exactly solvable model proposed by Mandal and Jarzynski in [62] (also see [63] for a proposal of a physical implementation using a spin valve and quantum dots). The system is described as follows: a particle on a ring thermally hops between three microstate (A, B and C). Connected to the transition $C \to A$ there is a mechanism that allows to lift a weight of mass m by a certain height Δh. In order for the dynamics to be thermodynamically consistent, we need to impose detailed balance on the rates:

$$\frac{r_{A\to B}}{r_{B\to A}} = \frac{r_{B\to C}}{r_{C\to B}} = 1\,, \qquad \frac{r_{A\to C}}{r_{C\to A}} = e^{-\beta m g \Delta h}\,. \tag{3.1}$$

Moreover, we assume that the height of the weight is constrained to the interval $[0, \Delta h]$, meaning that it can be lifted only if it hadn't been raised before, and it can be lowered only in the opposite situation. In this way, the weight effectively behaves as a two level system.

A Maxwell's demon is introduced in the system as a mechanism continuously monitoring the microstate of the particle and weight, and registering this information as a pair (X, Z), where $X \in \{A, B, C\}$ is the coordinate of the particle and $Z \in \{0, 1\}$ corresponds to whether the

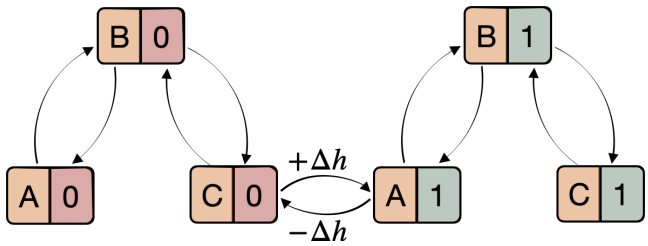


Figure 3.2: Depiction of the possible transitions during a single time step between the internal states of the demon. The first variable represents the state of the system, while the second binary variable corresponds to whether the weight is down or up.

weight is down or up. In this way, the dynamics of the system is completely represented by the internal state of the demon. Its memory is modelled by a finite number of cells containing one bit each, together with a counter that can switch between A, B and C. The difference in treatment between the two variables mirrors the fact that the only useful information in order to extract work is the one about the weight, so there is no need to keep track of the particle's history, but only of its current position.

In order to avoid any loopholes in the treatment, we assume a minimalistic functioning for the demon: at the beginning of the protocol, the weight is initialised according to the bit in the demon's memory, and starts interacting with the particle for a period of time $\tau \gg \tau_{eq}$ (the thermalisation timescale of the system); then, the weight is decoupled from the system, the work contained in the weight is stored into a battery, and the demon passes to the next memory cell.

Suppose the bits of memory have a probability p_0 to be in the state 0 (and similarly for p_1). In what follows, it is useful to define the quantity $\delta := p_0 - p_1$, called excess parameter. After tracking the dynamics of the system for long enough, each memory bit reaches a steady state characterised by the new probability distribution $p'_{0/1}$ and excess parameter δ'. The average change of height per bit is given by:

$$\langle \Delta h \rangle = \Delta h \, (p'_1 - p_1) = \frac{\Delta h}{2}(\delta - \delta') = \tag{3.2}$$

$$= \Delta h \, \frac{\delta - \delta^{eq}}{2} \left[1 - K(\tau) + J(\tau, \delta \cdot \delta^{eq})\right] , \tag{3.3}$$

where $\delta^{eq} := p_0^{eq} - p_1^{eq} = \tanh\left(\frac{\beta m g \Delta h}{2}\right)$, and $K(\tau)$ and $J(\tau, \delta \cdot \delta^{eq})$ admit

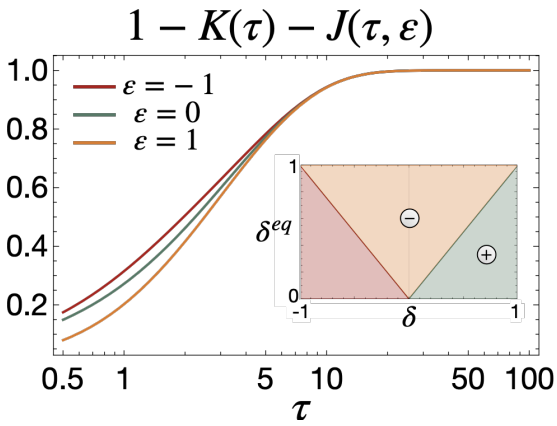


Figure 3.3: Behaviour of the quantity in the brackets of Eq. (3.3) as a function of the observation time τ, and for different $\varepsilon := \delta \cdot \delta^{eq}$. We refer to [62] for the analytical expression of $K(\tau)$ and $J(\tau, \varepsilon)$ (notice that the difference in notation, namely $K(\tau) = \frac{1}{3}\tilde{K}(\tau)$ and $J(\tau, \varepsilon) = \frac{\varepsilon}{6}\tilde{J}(\tau, \varepsilon)$, where the functions with the tilde are the one presented in [62]). In the inset we present the different regimes of the demon as characterised by the excess parameter δ and δ^{eq}: if $0 < \delta < \delta^{eq}$ (green region) the demon acts as an engine, extracting work at the cost of increasing the randomness of its memory; if $|\delta| > \delta^{eq}$ (orange region) then the demon acts as an eraser, resetting its memory tape at the cost of lowering the weight. Lastly, if $0 < -\delta < \delta^{eq}$ (the red region) the demon doesn't do anything useful: the weight is lowered and the entropy of its memory increases.

an analytical form, which we omit for brevity (see Fig. 3.3 for their behaviour). The most important feature of $\langle \Delta h \rangle$ is that in the limit $\tau \to \infty$ it rapidly approaches the asymptotic value of $\langle \Delta h_\infty \rangle = \frac{\delta - \delta^{eq}}{2}$.

This model allows for a clean study of the internal functioning of the Maxwell's demon. Suppose all the bits in the demon's memory are initialised to 0, i.e., $\delta = 1$. Then, in the long interaction limit $\tau \gg \tau_{eq}$, the demon is able to lift the weight by $\frac{1-\tanh\left(\frac{\beta m g \Delta h}{2}\right)}{2} N_m$ (where N_m is the number of bits in the demon's memory) extracting work from a unique thermal source, in contradiction with the second law of thermodynamics. To compensate for this, its memory gets filled with random bits distributed according to $p^{eq}_{0/1}$, which retrace the weight's history. There are two reasons to say that the second law is not violated: for one thing, the process cannot be said to be really cyclic, while the second law only

applies to cyclic processes; secondly, it should be noticed that the demon cannot continue to operate: in fact, once its memory is filled according to $p_{0/1}^{eq}$, the excess parameter will also become thermal, corresponding to an average change in height of $\langle \Delta h \rangle = 0$ (as it can be verified by inspection of Eq. (3.3)). Hence, in order for the demon to continue to extract work, it has to re-initialise its memory to 0, causing a dissipation lower bounded by the Landauer's limit. Interestingly, this is the fundamental reason that prevents Smochulowski's trapdoor from working.

Additionally to this regime, there is another one that deserves to be examined: consider the case in which $\Delta h \gg 1$, so that $\delta^{eq} \sim 1$. Then, if the demon's memory is initially in an equal mixture of zeros and ones ($\delta = 0$) , the system will act as an eraser: the energetic cost paid by lowering the weight, which will be lowered on average by $-\frac{\tanh\left(\frac{\beta m g \Delta h}{2}\right)}{2} N_m$, is translated into a reinitialisation of the bits in the demon's memory, which are all set to zero.

The two possible regimes that the demon can undergo show the interplay between the randomisation of memory bits and the energy stored in the weight: these are two resources that can be interchanged through the action of the demon. In fact, the model above shows that an empty memory is a resource that can be dissipated to extract energy; in the same way, one can use energetic resources to produce a derandomised memory tape. This close relationship shows to which great degree the information can be regarded as physical.

3.1.2 The physicality of intrinsic information

On this long route towards the exorcism of the demon the physicality of information and logical operations become evident to a deeper and deeper extent: first, one is led to consider the cost of information acquisition; then, the cost of information processing; finally, the cost of information erasure. This progressively led to the identification of logical irreversibility with the thermodynamic one.

As a matter of fact, the relationship between physics and intrinsic information is even more stringent. Let us consider again the operation of the Maxwell's demon above. As it was discussed, after a full cycle the memory of the demon is completely randomised, so it is impossible for it to continue to extract work without freeing up some of its memory tape. Denote by x the string describing its memory content (which, for inessential technical reasons, we suppose to be part of a prefix-free

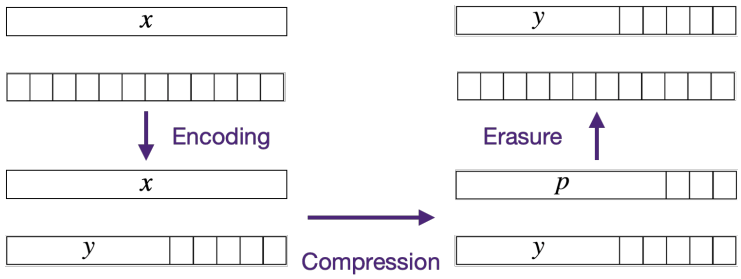


Figure 3.4: In order to reduce the amount of bits that it needs to erase, the demon can first compress the content of its memory. This is done in three steps: first, the string x describing its internal state is encoded into a shorter element y. Then, one can substitute x with a program p that on input y outputs x. Finally, p is erased. This is the only step that is not reversible.

language). Suppose now that one can provide the demon with an extra empty tape that needs to be returned in the same state at the end of the protocol. Then, in principle, in order to free up part of its memory, the demon could encode x into a new, shorter string y, which is written on the new tape. At this point, we can use y to compress x: one can, in fact, reversibly replace it with a program p that on input y gives back x. The minimum length of p is by definition $K(x|y)$, the conditional Kolmogorov complexity of x with respect to y (see App. D for the precise definition of this object). After this step, the state of the two tapes is given by (y, p). At this point the demon erase p, i.e., a minimum of $K(x|y)$ bits, and transfers back y on its memory, freeing up the ancillary tape.

This protocol hence can be used to free $\ell^*(x) - \ell^*(y)$ memory bits at the cost of dissipating at least $k_B T\, K(x|y)$ amount of heat. In particular, if one wants to completely erase the memory, corresponding to the transition $x \to \varepsilon$ (where ε denotes the empty string), the minimal amount of heat dissipated is $K(x|\varepsilon) \equiv K(x)$, the Kolmogorov complexity of x. Since $K(x) \leq \ell^*(x)$, by repeating this reinitialisation cycle, one could in principle observe macroscopic violations of the second law.

Considerations along these lines led Zurek to propose the following modification of the thermodynamic entropy [64, 65]:

$$S_Z = H + K\,, \tag{3.4}$$

where H is the usual Shannon or von Neumann entropy of the system, while K is the Kolmogorov complexity of the string in the measurement

register. Let us now re-examine Maxwell's demon in view of the definition above. At the beginning, the system starts with a certain amount of statistical entropy ($H > 0$), while the memory tape of the demon is completely blank ($K = 0$), so the initial thermodynamic entropy is given by $S_Z \equiv H$. When the demon starts operating, some of the statistical entropy H is harvested to produce work. At the same time, though, this operation results in a change of its memory to a state with finite K, so that globally $\Delta S_Z = 0$ (which also implies that $\Delta H = -\Delta K$). This should be contrasted with the usual definition of thermodynamic entropy: in this case, by identifying $S := H$, it could seem that the demon induces a momentary decrease of entropy, and that only the finiteness of its memory prevents it from actually producing macroscopic violations. Using S_Z on the other hand, shows that since every step is reversible, the thermodynamic entropy does not change during the transformation.

Interestingly, the compression procedure described at the beginning of this section does not affect the thermodynamic entropy. Indeed, since y is a recursive encoding of x, it holds that $K(x, y) = K(x)$ (see Eq. (D.4)), so the first step does not increase the algorithmic entropy; similarly, for the same reason we also have that $K(p, y) = K(x, y)$, showing that K stays constant during the overall process (see Eq. (D.5)).

Finally, during the erasure step one cancels the program for x generating $K_B T\, \ell^*(p)$ amount of heat. On the other hand, by applying Eq. (D.12), one can see that the algorithmic entropy decreases by $K(y) - K(x, y) = -K(x|y^*)$, where y^* is the minimal program for y. Then, the total dissipation during this last step can be quantified as:

$$Q = -K_B T\,(-\ell^*(p) - \Delta S_Z) = K_B T\,(\ell^*(p) - K(x|y^*)) \geq 0\,, \qquad (3.5)$$

since by definition $\ell^*(p) \geq K(x|y)$ and $K(x|y) \geq K(x|y^*)$. Since all the other steps were realised reversibly, Eq. (3.5) is sufficient to prove that the dissipation during the whole process is positive, in agreement with the second law of thermodynamics. This shows the connection between the randomness stored in the memory of the demon (or equivalently the intrinsic information contained in it) and the minimal amount of dissipation needed to clear it.

The proposal in Eq. (3.4) not only is able to give an elegant and organic treatment to Maxwell's demon, but it also explains what happens to the entropy when a (classical) measurement is performed: without the inclusion of the algorithmic complexity it would seem that carrying out a measurement on a classical system would reduce the total entropy of

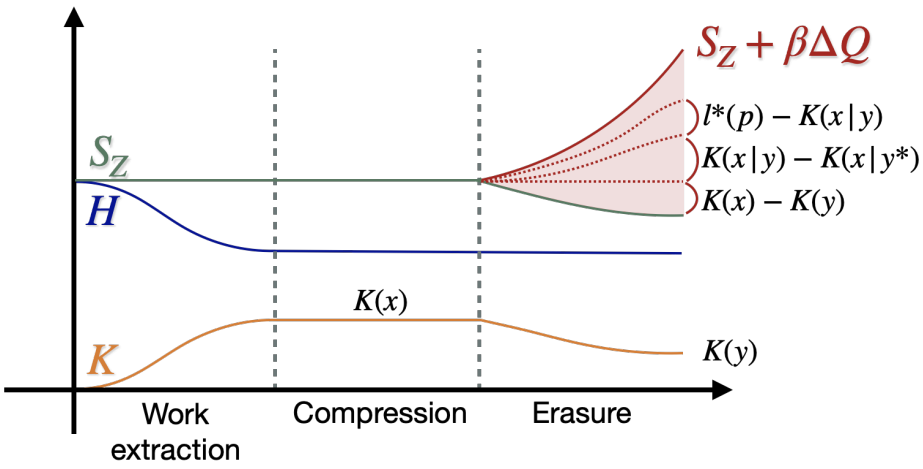


Figure 3.5: Evolution of the entropy S_Z and dissipation βQ during the three stages of the demon, together with the individual contributions coming from the statistical entropy H and the algorithmic complexity K. During the first phase the demon harvests information about the system and uses it to decrease its statistical entropy in order to extract work. In order to realise this operation reversibly, it needs to keep track of its observations, so the complexity of its memory record increases. Then, the compression depicted in Fig. 3.4 is realised. As discussed in the main text, despite the decrease in the number of memory bits used, this step does not affect K. Finally, during the erasure step, while K decreases, the amount of heat βQ released always satisfies $\beta Q \geq \beta \ell^*(p) \geq K(x|y) \geq K(x|y^*)$, so there is an overall increase in the total entropy.

the system, since a point in the phase space has zero statistical entropy. However, when the complexity of the register is taken into account one can see that the entropy simply changes its form, transforming from statistical to algorithmic.

Regarding the inequality in Eq. (3.5), it should be noticed that K is not computable. Hence, Landauer's limit ($Q = 0$) in this formulation can only be attained by sheer luck, randomly guessing the minimal encoding of the string. Ironically, even if this were the case, one wouldn't be able to tell in general how lucky it was, for the very same reason that prevented finding the optimal program in the first place. Moreover, it should also be noticed that there are two inequalities that need to be saturated in order to reach Landauer's limit: namely, one should have

$\ell^*(p) = K(x|y) = K(x|y^*)$. This means that not only p needs to be the optimal program for x, but that also y has to be minimal: in fact, the second inequality can only be satisfied if y is the shortest program for some string showing the importance of optimal encodings in order to avoid any form of dissipation.

It is also important to point out that in the case of macroscopic systems one can neglect the contribution that the algorithmic complexity gives to Eq. (3.5). In fact, given an ensemble of strings $\{s_i\}$ distributed according to $\{p(s_i)\}$, one can prove the double inequality [66]:

$$H(\{p(s_i)\}) \leq \sum_i p(s_i)K(s_i) \leq H(\{p(s_i)\}) + K(\{s_i, p(s_i)\}) \,, \qquad (3.6)$$

where $K(\{s_i, p(s_i)\})$ is the Kolmogorov complexity of the ensemble. For macroscopic systems one can expect that $H(\{p(s_i)\}) \gg K(\{s_i, p(s_i)\})$, since H is an extensive quantity, while only a coarse-grained description of the ensemble is usually needed. For this reason, one can approximate $S_Z \propto H$, which explains the virtually perfect agreement between the predictions of the usual thermodynamics in which the entropy function is given by H only.

The discussion above concludes the review of the theoretical understanding of the Maxwell's demon present in the literature so far. Possibly, though, there might be an extra step that one can take. It should be noticed, in fact, that no reference to time was made at any point. Each of the operations of the demon can take in principle infinite time. Still, the theoretical foundations of the second law coming from statistical mechanics (and the H-theorem in particular) suggest that one should not apply the thermodynamic formalisms when treating timescales comparable with the Poincaré recurrence period, as in this case there are provable violations of the second law that happen even without external interventions.

This discussion becomes particularly relevant when one treats information processing tasks. Suppose, for example, that the compression of the original content of the demon's memory x is done by encoding it into the solution of a formula, i.e., the Boolean satisfiability problem. In this case, whereas the encoding procedure is quite straightforward, retrieving the actual solution is a hard problem, most surely requiring a super-polynomial amount of steps. In this way, even if the information is completely conserved, it becomes effectively irretrievable in the timescales of interest. In order to prevent this kind of pathological situations it is

sensible, then, to modify Eq. (3.4). A minimal modification is given by substituting Kolmogorov complexity with Levin's one, defined as [67]:

$$Kt(x) := \{\ell(p) + \log t(p) |\, U(p) = x \text{ in } t(p) \text{ steps}\}\,, \tag{3.7}$$

where U is a fixed prefix-free universal Turing machine, and $\ell(p)$ and $t(p)$ are respectively the length and the number of steps of the program p. Thanks to the extra contribution coming from the running time of the program, one can see some kind of trade-off between reducing the length of the string x, and allowing it to be still retrieved in reasonable time. This proposal is still tentative, but it hints at the fact that maybe Maxwell's demon hasn't finished yet to bring new insights to our physical understanding of the world.

3.1.3 Quantum Landauer principle

The key role of information in thermodynamics is reflected by the simplicity with which thermodynamic laws can be proved using tools from quantum information theory. In particular, it was shown in [68] how the second law of thermodynamics and Landauer's limit can be derived as a simple consequence of the monotonicity of the relative entropy. Consider, in fact, a system in contact with an external environment undergoing an arbitrary evolution described by a unitary U. The initial state is supposed to be given by the product[3] $\rho_{SE} := \rho_S \otimes \rho_E$ and the final state is denoted by $\rho'_{SE} := U(\rho_S \otimes \rho_E)\, U^\dagger$. Then we have the following theorem:

Theorem 18 (Second law)**.** *The sum in the change of entropy of the system and of the environment is always positive:*

$$\Delta S_S + \Delta S_E = (S(\rho'_S) - S(\rho_S)) + (S(\rho'_E) - S(\rho_E)) \geq 0\,. \tag{3.8}$$

In the theorem we implicitly used the definition of reduced state $\rho'_S := \mathrm{Tr}_E\,[\rho'_{SE}]$ (and similarly for what regards ρ'_E), and we identify the thermodynamic entropy with the von Neumann one $S(\rho) := -\mathrm{Tr}\,[\rho \log \rho]$.

[3]Notice in fact that any correlation between system and environment can be thought as information that the latter has about the first. In this case, we already know from the Maxwell's demon example that one can generate an apparent violation of the second law by exploiting this correlation. However, if one also considers the dynamics that led to the creation of the correlation, then we are back to the scenario presented here, and there is no threat to the second law.

Proof. There are a number of simple results from quantum information that will be used. First notice that thanks to the tensor product structure of ρ_{SE}, the entropy of the total state equals the entropy of the reduced ones, i.e., $S(\rho_{SE}) = S(\rho_S) + S(\rho_E)$ (see Eq. (1.201) and Eq. (1.202)). Moreover, thanks to the invariance of the von Neumann entropy under the action of global unitaries, it also holds that $S(\rho_{SE}) = S(\rho'_{SE})$. Using this two facts, we can rewrite Eq. (3.8) as:

$$\begin{aligned} \Delta S_S + \Delta S_E &= (S(\rho'_S) + S(\rho'_E)) - (S(\rho_S) + S(\rho_E)) = && (3.9)\\ &= (S(\rho'_S) + S(\rho'_E)) - S(\rho_{SE}) = && (3.10)\\ &= S(\rho'_S) + S(\rho'_E) - S(\rho'_{SE}) = I(S' : E') \geq 0\,, && (3.11) \end{aligned}$$

where in the last step we used the definition of mutual information in Eq. (1.203). We remind the reader that this quantity is always positive unless ρ_{AB} is a product state thanks to equality in Eq. (1.205), namely $I(A : B) = S(\rho_{AB}||\rho_A \otimes \rho_B)$. This also shows that the equality in Eq. (3.8) can be saturated if and only if $\Delta S_S = \Delta S_E = 0$. □

The proof of the theorem actually offers insights on the nature of the second law: the appearance of the mutual information in Eq. (3.11) shows how the irreversibility arises from the spreading of information across the system-environment partition, which makes it irretrievable through local operations. Still, the information is conserved at a global level, as it can be noticed from the equality $S(\rho'_{SE}) = S(\rho_{SE})$. Moreover, reading Eq. (3.11) from right to left, one can also interpret it as the fact that creating correlation between system and environment always corresponds to a global increase of local entropy.

So far, all the quantities used were information theoretic, and the consideration made were quite general. Notice, for example, that in order to derive Thm. 18 we don't need to assume a thermal environment. Suppose now that this is the case, i.e., $\rho_E \equiv \pi_\beta(H_E) = \frac{e^{-\beta H_E}}{\mathcal{Z}_{H_E}}$, and remember that the heat is given by the change of energy in the environment:

$$Q := \mathrm{Tr}\left[H_E\, \rho'_E\right] - \mathrm{Tr}\left[H_E\, \rho_E\right]\,. \tag{3.12}$$

Then, the following result can also be proven:

Theorem 19 (Landauer's principle as equality)**.** *The heat transferred to the environment can be decomposed as follows:*

$$\beta Q = -\Delta S_S + I(S' : E') + S(\rho'_E||\rho_E)\,. \tag{3.13}$$

This result implies that the decrease in the entropy of the system always causes a corresponding release of heat bounded by:

$$\beta Q \geq -\Delta S_S \,. \tag{3.14}$$

Thm. 19 can be seen as a refinement of Landauer's principle to an equality: in fact, if one considers the erasure of a maximally mixed bit the corresponding change in entropy is given by $\Delta S_S = -\log 2$.

Proof. If $\rho_E \equiv \pi_\beta(H_E)$ it follows that:

$$\begin{aligned}
\beta Q &= \beta(\mathrm{Tr}\left[H_E\,\rho'_E\right] - \mathrm{Tr}\left[H_E\,\rho_E\right]) = && (3.15)\\
&= \mathrm{Tr}\left[\log e^{-\beta H_E}\,\rho_E\right] - \mathrm{Tr}\left[\log e^{-\beta H_E}\,\rho'_E\right] = && (3.16)\\
&= \mathrm{Tr}\left[\rho_E \log \rho_E\right] - \mathrm{Tr}\left[\rho'_E \log \rho_E\right] = && (3.17)\\
&= \mathrm{Tr}\left[\rho_E \log \rho_E\right] - \mathrm{Tr}\left[\rho'_E \log \rho'_E\right] + \mathrm{Tr}\left[\rho'_E(\log \rho'_E - \log \rho_E)\right] = && (3.18)\\
&= \Delta S_E + S(\rho'_E||\rho_E)\,, && (3.19)
\end{aligned}$$

where in the second line we used the functional form of the Gibbs state to make ρ_E appear in the logarithms. Moreover, from the proof of Thm. 18, we also know that:

$$\Delta S_S + \Delta S_E = I(S' : E')\,. \tag{3.20}$$

Then, putting together Eq. (3.19) and (3.20) we obtain the equality in Eq. (3.13). Thanks to the positivity of $I(S' : E')$ and $S(\rho'_E||\rho_E)$, also Eq. (3.14) follows easily. □

From Eq. (3.13) one can also deduce what are the conditions to saturate Landauer's bound. In fact, $I(S' : E')$ is zero if and only if ρ'_{SE} is a product state, and $S(\rho'_E||\rho_E)$ only in the case for $\rho'_E \equiv \rho_E$. Interestingly, whereas the first condition can be satisfied also for finite dimensional systems, the latter requires an infinite bath [68].

3.2 Fluctuation theorems

At its birth, thermodynamics was especially successful in providing universal results expressed by equation of states. The most famous example is, of course, the equation of state of the ideal gas, but in the same category we can also find the characterisation of the efficiency of Carnot

engine or of the Otto one. These are called universal, as they are independent of the working medium chosen, as long as it stays at equilibrium during the whole cycle. This level of generality is lost for non-equilibrium transformations: equalities get replaced by inequalities as it happens, for example, for the Carnot efficiency, which becomes a lower bound on the performance of any engine working between two given temperatures. In the same family, we have the Clausius expression of the second law:

$$\beta Q \geq -\Delta S\,. \tag{3.21}$$

For more than a century the study of the quantity $\Sigma := \Delta S + \beta Q$, dubbed entropy production, seemed hopeless[4]. Remarkably, the situation unblocked when microscopic systems, for which fluctuations become relevant, were started to be studied. As we sketched in Chapter 2, in this context quantities like work and heat become stochastic variables, which can be computed at the level of the single trajectory. There are a number of complications coming from the stochasticity of the quantity at hand, but, maybe counterintuitively, this generalisation allowed to obtain elegant universal results about the entropy production.

3.2.1 The Jarzynski equality

The first and among the most important results found in this direction is given by Jarzynski equality [69]:

$$\left\langle e^{-\beta w} \right\rangle = e^{-\beta \Delta F}\,. \tag{3.22}$$

This equation connects the average of the exponential of the work, i.e., a stochastic variable, to the exponential of a function of state, namely the difference of free energy $\Delta F := \Delta U - \beta^{-1}\Delta S$. The connection with Σ can be obtained through the first law, noticing that $w = \Delta U + Q = \Delta F + \beta^{-1}\Sigma$, so that Eq. (3.22) can be rewritten as:

$$\left\langle e^{-\beta(w-\Delta F)} \right\rangle = \left\langle e^{-\Sigma} \right\rangle = 1\,, \tag{3.23}$$

also called Kawasaki identity. It is quite exceptional that a quantity like the entropy production which, in general, depends both on the trajectory in the space of parameters and on the speed at which the transformation

[4]Indeed, the fact that the Clausius inequality is not presented as an equality, i.e., $\beta Q = -\Delta S + \Sigma$ is very emblematic of this.

is performed, can be put in relation through an equality with a function that only depends on the value of the parameters at the endpoints of the trajectory. Moreover, the only key assumption in deriving Eq. (3.22) is that the system is initially in thermal equilibrium at inverse temperature β, whereas the final distribution can be generic.

There are a number of frameworks in which Jarzynski equality can be proven [70, 71]. In particular, one can prove Eq. (3.22) in the context of Hamiltonian dynamics. Suppose that the system is initially in equilibrium with a thermal bath at inverse temperature β and that, before the beginning of the protocol, it is detached from the environment and evolved according to some driven Hamiltonian dynamics. Since the system is in isolation, the change of internal energy coincides with the work performed on it. This allows us to write the work of a trajectory $\{\mathbf{x}_t\}_{t\in[0,\tau]}$ as

$$w(\{\mathbf{x}_t\}_{t\in[0,\tau]}) \equiv w(\mathbf{x}_\tau, \mathbf{x}_0) = H(\mathbf{x}_\tau, \tau) - H(\mathbf{x}_0, 0)\,, \tag{3.24}$$

where $H(\mathbf{x}, t)$ denotes the Hamiltonian evaluated at time t and generalised coordinates $\mathbf{x}$. Then, the average in Eq. (3.22) can be rewritten as:

$$\left\langle e^{-\beta w} \right\rangle = \frac{1}{\mathcal{Z}_0} \int \mathrm{d}\mathbf{x}_0\, e^{-\beta H(\mathbf{x}_0,0)} e^{-\beta w(\mathbf{x}_\tau,\mathbf{x}_0)} = \frac{1}{\mathcal{Z}_0} \int \mathrm{d}\mathbf{x}_0\, e^{-\beta H(\mathbf{x}_\tau,\tau)} = \tag{3.25}$$

$$= \frac{1}{\mathcal{Z}_0} \int \mathrm{d}\mathbf{x}_\tau \left|\frac{\partial \mathbf{x}_\tau}{\partial \mathbf{x}_0}\right|^{-1} e^{-\beta H(\mathbf{x}_\tau,t)} = \frac{\mathcal{Z}_\tau}{\mathcal{Z}_0} = e^{-\beta \Delta F}\,, \tag{3.26}$$

where $\mathcal{Z}_\tau := \int \mathrm{d}\mathbf{x}_\tau\, e^{-\beta H(\mathbf{x}_\tau,t)}$ is the partition function, and we used the canonical definition of free energy in the context of statistical mechanics given by $F_t := -\beta^{-1} \log \mathcal{Z}_t$. Passing from the first to the second line we changed variables, and we used the notation $\left|\frac{\partial \mathbf{x}_\tau}{\partial \mathbf{x}_0}\right|$ for the Jacobian of the transformation. Thanks to Liouville's theorem, this is equal to one for Hamiltonian evolutions, leading to Eq. (3.13).

An alternative proof follows from Feynman-Kac formula for stochastic processes. In particular consider an evolution described by the master equation:

$$\dot{p}_t(\mathbf{x}) = \mathcal{L}_t\, p_t(\mathbf{x})\,, \tag{3.27}$$

where $\mathcal{L}_t$ is a time-dependent transition matrix satisfying the property:

$$\mathcal{L}_t\, e^{-\beta H(\mathbf{x},t)} = 0\,. \tag{3.28}$$

This connects the stochastic dynamics to the Hamiltonian parameters, and it corresponds to the request that for each t the thermal state is a fixed point of the instantaneous dynamics. Then, define the density:

$$g_t(\mathbf{x}) := \frac{e^{-\beta H(\mathbf{x},t)}}{\mathcal{Z}_0} \,. \tag{3.29}$$

This coincides with the thermal state at time $t = 0$, and satisfies the following transport equation:

$$\dot{g}_t(\mathbf{x}) = \mathcal{L}_t \, g_t(\mathbf{x}) - \beta \, \frac{\partial H}{\partial t}(\mathbf{x}, t) g_t(\mathbf{x}) \,. \tag{3.30}$$

At this point we can apply Feynman-Kac formula[5]: this tells us that $g_t(\mathbf{x})$ is also the solution to:

$$g_t(\mathbf{x}) = \left\langle \delta(\mathbf{x} - \mathbf{x}_t) \, e^{-\beta \int_0^t \mathrm{d}s \, \frac{\partial H}{\partial s}(\mathbf{x},s)} \right\rangle = \left\langle \delta(\mathbf{x} - \mathbf{x}_t) \, e^{-\beta w_t} \right\rangle \tag{3.31}$$

where the average is on all trajectories evolving according to Eq. (3.27) and that are initially distributed as $g_0(\mathbf{x})$, while $\mathbf{x}_t$ is the coordinate of the trajectory at time t. Notice that we also used the definition of the work in Eq. (2.10):

$$w_t := \int_0^t \mathrm{d}s \, \frac{\partial H}{\partial s}(\mathbf{x}_s, s) \,. \tag{3.32}$$

Then, we can interpret $g_t(\mathbf{x})$ as the propagator of $\mathcal{L}_t$, and w_t as the analogous of an imaginary action in the path integral formalism. We now have all the definitions that we need to prove Jarzynski equality. Setting the time to τ and integrating over all final configurations we obtain:

$$\left\langle e^{-\beta w} \right\rangle = \int \mathrm{d}\mathbf{x} \, \left\langle \delta(\mathbf{x} - \mathbf{x}_\tau) \, e^{-\beta w_\tau} \right\rangle = \int \mathrm{d}\mathbf{x} \, g_\tau(\mathbf{x}) = \tag{3.33}$$

$$= \int \mathrm{d}\mathbf{x} \, \frac{e^{-\beta H(\mathbf{x},\tau)}}{\mathcal{Z}_0} = \frac{\mathcal{Z}_\tau}{\mathcal{Z}_0} = e^{-\beta \Delta F} \,. \tag{3.34}$$

The possibility of deriving Eq. (3.22) in so radically different frameworks shows the exceptional generality of Jarzynski equality. Notably, the only assumption at the heart of all the derivations is the thermality of the initial state.

[5]This can be thought as application of the path integral formalism in imaginary time to the solution of stochastic equations.

Apart from the remarkable result of connecting properties out of equilibrium with equilibrium ones, the Jarzynski equality can also be used to give quantitative bounds on the magnitude of the dissipation. In particular, from Jensen's inequality we have that $\langle e^x \rangle \geq e^{\langle x \rangle}$. Applying it to Eq. (3.22) and taking the logarithm gives:

$$-\beta \Delta F = \log \left\langle e^{-\beta w} \right\rangle \geq -\beta \langle w \rangle \qquad \Longrightarrow \qquad \Delta F \leq \langle w \rangle \,, \tag{3.35}$$

which is equivalent to the second law presented in Eq. (3.21), just in terms of the work. Hence, Clausius inequality continues to hold in this framework for average quantities. Moreover, from Kawasaki identity (Eq. (3.23)), this inequality can also be expressed as $\langle \Sigma \rangle \geq 0$, i.e., the average entropy production is always positive.

In fact, due to the stochasticity of Σ, its positivity does not hold in general, but only after taking the average. Still, one can give an estimate on the probability of obtaining negative dissipation as follows:

$$P(\Sigma \leq -\zeta) = \int_{-\infty}^{-\zeta} \mathrm{d}\Sigma \; p(\Sigma) \leq \int_{-\infty}^{-\zeta} \mathrm{d}\Sigma \; p(\Sigma) e^{-(\zeta+\Sigma)} \leq \tag{3.36}$$

$$\leq e^{-\zeta} \int_{-\infty}^{\infty} \mathrm{d}\Sigma \; p(\Sigma) e^{-\Sigma} = e^{-\zeta} \,, \tag{3.37}$$

where $P(\Sigma \leq -\zeta)$ is the probability of obtaining a value of Σ smaller than $-\zeta$, and in the second line we used Kawasaki identity (also notice that $-(\zeta + \Sigma) \geq 0$). This shows that even if violations of the second law are possible in principle, they are exponentially suppressed in their magnitude.

3.2.2 Time reversal and Crooks fluctuation theorem

One of the central questions in statistical mechanics is how a preferred direction of time can arise from time reversible equations of motion. Crooks relations show that if one can assume the thermalisation of a system, then a time arrow naturally emerges [72]. Consider, in fact, a Hamiltonian which changes in a series of discrete quenches $H_0 \to H_1 \to \cdots \to H_N$. When the Hamiltonian is not changing, the system undergoes a microscopic reversible dynamics: this concept is the extension of time symmetry to non-conservative dynamics and can be summarised by the following equation:

$$p(\mathbf{x} \to \mathbf{y} \,|\, H_i) = p(\mathbf{y} \to \mathbf{x} \,|\, H_i) \,, \tag{3.38}$$

where we implicitly defined the probability at equilibrium of the transition $\mathbf{x} \to \mathbf{y}$ and we highlighted the dependency on the particular Hamiltonian H_i. If the steady state is thermal, Eq. (3.38) can be rewritten as:

$$p(\mathbf{y} \,|\, \mathbf{x},\, H_i)\, e^{-\beta H_i(\mathbf{x})} = p(\mathbf{x} \,|\, \mathbf{y},\, H_i)\, e^{-\beta H_i(\mathbf{y})}\,, \tag{3.39}$$

where $p(\mathbf{y} \,|\, \mathbf{x},\, H_i)$ is the conditional probability of evolving to the state $\mathbf{y}$ starting from $\mathbf{x}$ and with Hamiltonian H_i. Rearranging the above equation, one obtains the classical definition of detailed balance (see Eq. (1.329)).

Crooks fluctuation theorem puts in relation the probability of trajectories evolved according to the forward evolution $H_0 \to H_1 \to \cdots \to H_N$ with their reversed, evolved according to $H_N \to H_{N-1} \to \cdots \to H_0$. In particular, for each trajectory $\{\mathbf{x}_i\}_{i\in 1,\dots,N}$ we define its reversed version $\{\bar{\mathbf{x}}_i\}_{i\in 1,\dots,N}$ to satisfy $\bar{\mathbf{x}}_i = \mathbf{x}_{N-i}$. Moreover, denote by $p^F(\{\mathbf{x}_i\})$ the probability of observing the trajectory $\{\mathbf{x}_i\}$ starting at thermal equilibrium and evolving according to the forward evolution; in the same way $p^R(\{\mathbf{x}_i\})$ indicates the probability for $\{\mathbf{x}_i\}$ starting at equilibrium (this time the thermal state is the one associated with the Hamiltonian H_N) and evolving according to the reversed dynamics[6]. Then, a simple algebraic manipulation shows that:

$$\frac{p^F(\{\mathbf{x}_i\})}{p^R(\{\bar{\mathbf{x}}_i\})} = \frac{e^{-\beta H_0(\mathbf{x}_0)}}{\mathcal{Z}_0}\, \frac{\mathcal{Z}_N}{e^{-\beta H_0(\mathbf{x}_N)}}\, \frac{\prod_{i=1}^N p(\mathbf{x}_i \,|\, \mathbf{x}_{i-1},\, H_i)}{\prod_{i=1}^N p(\mathbf{x}_{N-i} \,|\, \mathbf{x}_{N-i+1},\, H_{N-i+1})} = \tag{3.40}$$

$$= \frac{\mathcal{Z}_N}{\mathcal{Z}_0}\, \frac{e^{-\beta H_0(\mathbf{x}_0)}}{e^{-\beta H_N(\mathbf{x}_N)}} \prod_{i=1}^N \frac{p(\mathbf{x}_i \,|\, \mathbf{x}_{i-1},\, H_i)}{p(\mathbf{x}_{i-1} \,|\, \mathbf{x}_i,\, H_i)} = \tag{3.41}$$

$$= e^{-\beta \Delta F} e^{\beta \sum_{i=0}^{N-1} (H_{i+1}(\mathbf{x}_i) - H_i(\mathbf{x}_i))}\,, \tag{3.42}$$

where in the second line we rearranged the transition probabilities, and used the microscopic reversibility together with the definition of ΔF to obtain the final expression. Consider now the sum:

$$\sum_{i=0}^{N-1} (H_{i+1}(\mathbf{x}_i) - H_i(\mathbf{x}_i))\,. \tag{3.43}$$

[6] An additional difference between the forward and reversed process is that the first begins with a quench, while we define the latter to start with a change of the internal state. This technical assumption becomes increasingly less relevant as $N \to \infty$.

This expression equals the work performed during the forward evolution, which we denote by w_F. In fact, since during a quench the system is effectively isolated, the work performed equals the change of energy; summing over every quench, we obtain exactly the quantity expressed in Eq. (3.43). Wrapping everything together, gives the famous Crooks relations:

$$\frac{p^F(\{\mathbf{x}_i\})}{p^R(\{\bar{\mathbf{x}}_i\})} = e^{\beta(w_F - \Delta F)} \,. \tag{3.44}$$

This result tells us that if a trajectory has a positive entropy production, then its probability is exponentially bigger than its time reversed version.

It should be noticed that Jarzynski equality is not independent from Eq. (3.44). In fact, we can use Crooks fluctuation theorem to prove it:

$$\left\langle e^{-\beta w} \right\rangle = \int \mathrm{d}\{\mathbf{x}_i\}\; p^F(\{\mathbf{x}_i\}) e^{-\beta w_F} = \tag{3.45}$$

$$= \int \mathrm{d}\{\bar{\mathbf{x}}_i\}\; p^R(\{\bar{\mathbf{x}}_i\}) e^{\beta(w_F - \Delta F)} e^{-\beta w_F} = e^{-\beta \Delta F} \,, \tag{3.46}$$

where we used Eq. (3.44) to pass from p^F to p^R, and in the last step we integrated over $\{\bar{\mathbf{x}}_i\}$, exploiting the fact that $p^R(\{\bar{\mathbf{x}}_i\})$ is normalised.

Moreover, one can find a similar relation to Eq. (3.44) for the probability of work. Denote by $p^{F/R}(w)$ the probability of obtaining a work value of w along the forward/reversed trajectory. Then, it follows from Eq. (3.44) that the two are related as:

$$p^F(w) = \int_{w(\{\mathbf{x}_i\})=w} \mathrm{d}\{\mathbf{x}_i\}\; p^F(\{\mathbf{x}_i\}) \tag{3.47}$$

$$= e^{\beta(w-\Delta F)} \int_{w(\{\mathbf{x}_i\})=w} \mathrm{d}\{\mathbf{x}_i\}\; p^R(\{\bar{\mathbf{x}}_i\}) = \tag{3.48}$$

$$= e^{\beta(w-\Delta F)} \int_{w(\{\bar{\mathbf{x}}_i\})=-w} \mathrm{d}\{\bar{\mathbf{x}}_i\}\; p^R(\{\bar{\mathbf{x}}_i\}) \tag{3.49}$$

$$= e^{\beta(w-\Delta F)}\, p^R(-w) \,, \tag{3.50}$$

where in Eq. (3.49) we used the fact that the work along the inverse trajectory is the negative of the one along the forward evolution. This shows that the probability of obtaining a positive dissipation in the forward evolution is exponentially bigger than the one of getting back the same energy spent during the backwards evolution.

One of the most interesting features of Crooks fluctuation theorem is that it makes quantitative the relation between entropy production and

the asymmetry in the direction of time. In fact, by taking the logarithm of both sides of Eq. (3.44) and integrating over $\{\mathbf{x}_i\}$ we obtain:

$$\langle \Sigma \rangle = \int \mathrm{d}\{\mathbf{x}_i\}\; p^F(\{\mathbf{x}_i\})\, \beta(w_F - \Delta F) = \tag{3.51}$$

$$= \int \mathrm{d}\{\mathbf{x}_i\}\; p^F(\{\mathbf{x}_i\}) \log \frac{p^F(\{\mathbf{x}_i\})}{p^R(\{\bar{\mathbf{x}}_i\})} = S(p^F || p^R)\,. \tag{3.52}$$

The average entropy production is exactly connected to the distinguishability between the forward and reversed probability distribution. Hence, $\langle \Sigma \rangle$ is always positive and zero if and only if $p^F \equiv p^R$. Indeed, other quantifiers for the asymmetry in the direction of time can be introduced (as the symmetrised relative entropy or the Jensen-Shannon divergence [71, 73]) showing how Crooks fluctuation theorem implies the presence of a preferred direction of time. Still, it should be noticed that this relation is not sufficient to give a statistical mechanical foundation to the time arrow: in fact, as noted in [71], following a remark by Gibbs, assuming the existence of a thermal state is enough to prove the second law of thermodynamics in the form in Eq. (3.21). Since the existence of thermalisation is a key assumption in the derivation of both Jarzynski equality and Crooks theorem, neither of them can be used to explain how systems reach thermal equilibrium. Nonetheless, their importance is not diminished from this, but rather amplified: they actually make explicit how the existence of a thermal state has strong physical consequences explaining the time asymmetry of dissipative transformations.

3.2.3 Fluctuation theorems for quantum systems: Jarzynski equality and the TPM scheme

Both Jarzynski equality and Crooks fluctuation theorem can be deduced from Eq. (2.28), the expression of the cumulant generating function (CGF) for the work statistics in the TPM scheme. This reads:

$$K^{-\beta w}(\lambda) = \log \int \mathrm{d}w\; p(w) e^{-\beta w \lambda} = \tag{3.53}$$

$$= -\lambda \beta \Delta F + (\lambda - 1)\, S_\lambda(\pi_\beta(H_\tau) || U_\tau \pi_\beta(H_0) U_\tau^\dagger)\,. \tag{3.54}$$

Then, setting λ to 1 we obtain:

$$K^{-\beta w}(1) = \log \int \mathrm{d}w\; p(w) e^{-\beta w} \tag{3.55}$$

$$= \log \left\langle e^{-\beta w} \right\rangle = -\beta \Delta F\,, \tag{3.56}$$

which is equivalent to the Jarzynski equality. The Crooks fluctuation theorem can be proven just as easily. First, notice that the reverse protocol is the one which starts from the thermal state $\pi_\beta(H_\tau)$ and evolves according to the unitary $U_\tau^\dagger$. Moreover, from the definition of the λ-Rényi divergence it easily follows that:

$$(\lambda-1)\, S_\lambda(\pi_\beta(H_0)||U_\tau^\dagger \pi_\beta(H_\tau)U_\tau) = \tag{3.57}$$

$$= \log \mathrm{Tr}\left[\pi_\beta(H_0)^\lambda (U_\tau^\dagger \pi_\beta(H_\tau)U_\tau)^{1-\lambda}\right] = \tag{3.58}$$

$$= \log \mathrm{Tr}\left[U_\tau\, \pi_\beta(H_0)^\lambda\, U_\tau^\dagger\, \pi_\beta(H_\tau)^{1-\lambda}\right] = \tag{3.59}$$

$$= ((1-\lambda)-1)\, S_{1-\lambda}(\pi_\beta(H_\tau)||U_\tau \pi_\beta(H_0)U_\tau^\dagger)\,. \tag{3.60}$$

Then, we can rewrite the CGF for the reversed protocol as:

$$K_R^{-\beta w}(\lambda) = \lambda\beta\Delta F + (\lambda-1)\, S_\lambda(\pi_\beta(H_0)||U_\tau^\dagger \pi_\beta(H_\tau)U_\tau) = \tag{3.61}$$

$$= \beta\Delta F - (1-\lambda)\beta\Delta F - \lambda S_{1-\lambda}(\pi_\beta(H_\tau)||U_\tau \pi_\beta(H_0)U_\tau^\dagger) = \tag{3.62}$$

$$= \beta\Delta F + K_F^{-\beta w}(1-\lambda)\,, \tag{3.63}$$

where we used the subscripts F/R to distinguish the forward and reversed CGF. Eq. (3.63) is in fact equivalent to Crooks theorem. Indeed, the probability of the forward process can be rewritten as the following Fourier transform:

$$p^F(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\; e^{i\,k\beta w}\, \hat{p}^F(k) = \tag{3.64}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\; e^{i\,k\beta w} \int_{-\infty}^{\infty} \mathrm{d}\tilde{w}\; e^{-i\,k\beta\tilde{w}} p^F(\tilde{w}) = \tag{3.65}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\; e^{i\,k\beta w}\, e^{K_F^{-\beta w}(ik)} = \tag{3.66}$$

$$= \frac{e^{-\beta\Delta F}}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\; e^{i\,k\beta w}\, e^{K_R^{-\beta w}(1-ik)} = \tag{3.67}$$

$$= \frac{e^{-\beta\Delta F}}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k \int_{-\infty}^{\infty} \mathrm{d}\tilde{w}\; e^{i\,k\beta(w+\tilde{w})}\, e^{-\beta\tilde{w}}\, p^R(\tilde{w}) = \tag{3.68}$$

$$= e^{\beta(w-\Delta F)}\, p^R(-w)\,, \tag{3.69}$$

extending Eq. (3.50) to quantum systems. Since there is no straightforward definition of quantum trajectory, this is the only fluctuation theorem we derive here.

The fact that Crooks fluctuation theorem implies Jarzynski equality can also be deduced at the level of the CGF: in fact, since for any probability distribution $p(w)$ one has $K^{-\beta w}(0) = 0$ (as it corresponds to the normalisation of $p(w)$), one can use Eq. (3.63) to prove that:

$$\beta\Delta F + K_F^{-\beta w}(1) = K_R^{-\beta w}(0) = 0\,, \tag{3.70}$$

which is equivalent to Eq. (3.56).

There is a strong relation between Jarzynski equality and the TPM work distribution: in fact, any non-contextual probability distribution for the work that (*i*) satisfies Jarzynski equality for any thermal state and (*ii*) admits a classical limit when evaluated on thermal states, must coincide with the TPM scheme [74]. More precisely, we have:

Theorem 20. *Suppose that the probability distribution $p(w|\mathcal{P})$ associated to a thermodynamic protocol $\mathcal{P}$ satisfies the three conditions:*

1. *(Linear probability) $p(w|\mathcal{P})$ is a probability distribution linear under convex mixing of different protocols. This means that by randomly choosing one protocol between $\mathcal{P}_1$ and $\mathcal{P}_2$ with probability α, the work distribution changes as*

$$p(w|\alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2) = \alpha\, p(w|\mathcal{P}_1) + (1-\alpha)\, p(w|\mathcal{P}_2)\,;$$

2. *(Jarzynski equality) For any thermal initial distribution $p(w|\mathcal{P})$ satisfies the Jarzynski equality;*

3. *(Classicality 3) The average energy change equals the average work on initial thermal states.*

Then $p(w|\mathcal{P})$ coincides almost everywhere in the Hamiltonian space with the statistics coming from the TPM scheme.

Proof. For simplicity, we assume that the initial and final Hamiltonian are non-degenerate. Moreover, we also assume that the energy gaps are non-degenerate, meaning that $(E_\tau^{(j)} - E_0^{(i)}) = (E_\tau^{(l)} - E_0^{(k)})$ if and only if $i = k$ and $j = l$. This reduces the scope of applicability of the theorem by a measure zero set, as any degenerate Hamiltonian can be made non-degenerate by an ε-perturbation.

As noted in the proof of Theorem 15, condition 1 implies the existence of a POVM $\{M^w\}_{w\in\mathbb{R}}$ which depends on the protocol $\mathcal{P}$ and the final measurement, but not on the initial state ρ, such that $p(w|\mathcal{P}) = \mathrm{Tr}\,[\rho M^w]$.

Hence, we can rewrite Jarzynski equality in this context as:

$$\left\langle e^{-\beta w} \right\rangle = \int \mathrm{d}w \; p(w|\mathcal{P}) e^{-\beta w} = \int \mathrm{d}w \; \mathrm{Tr}\left[\frac{e^{-\beta H_0}}{\mathcal{Z}_0} M^w \right] e^{-\beta w} = \quad (3.71)$$

$$= e^{-\beta \Delta F} = \frac{\mathcal{Z}_\tau}{\mathcal{Z}_0} , \quad (3.72)$$

where in the second line we highlighted the dependence of the free energy on the partition functions. Multiplying both sides by $\mathcal{Z}_0$ and expressing the trace in coordinates leads to the equality:

$$\int \mathrm{d}w \; \sum_i \left\langle E_0^{(i)} \right| M^w \left| E_0^{(i)} \right\rangle \; e^{-\beta(w + E_0^{(i)})} = \sum_j e^{-\beta E_\tau^{(j)}} . \quad (3.73)$$

Moreover, if one expands the exponentials in Taylor series, by noticing that Eq. (3.73) should hold for any β we can equate the single terms in the series as follows:

$$\int \mathrm{d}w \; \sum_i \left\langle E_0^{(i)} \right| M^w \left| E_0^{(i)} \right\rangle \; \frac{(-\beta(w + E_0^{(i)}))^n}{n!} = \sum_j \frac{(-\beta E_\tau^{(j)})^n}{n!} . \quad (3.74)$$

Denote by $E_\tau^{(N)}$ the maximum eigenvalue of H_τ. Then, focusing on the n-th term in the sums in Eq. (3.74), by dividing both sides of the equation by $(E_\tau^{(N)})^n$ we obtain:

$$\int \mathrm{d}w \; \sum_i \left\langle E_0^{(i)} \right| M^w \left| E_0^{(i)} \right\rangle \left(\frac{w + E_0^{(i)}}{E_\tau^{(N)}} \right)^n = \sum_j \left(\frac{E_\tau^{(j)}}{E_\tau^{(N)}} \right)^n . \quad (3.75)$$

Since the relation above has to hold for arbitrarily large n, we can take the limit $n \to \infty$. On the right hand side, this gives 1, as all the ratios with eigenvalues smaller than $E_\tau^{(N)}$ go to zero, so that only the last term in the sum survives. On the left hand side we have the following three possibilities:

$$\lim_{n\to\infty} \left(\frac{w + E_0^{(i)}}{E_\tau^{(N)}} \right)^n = \begin{cases} 0 & \text{if } w + E_0^{(i)} < E_\tau^{(N)} , \\ 1 & \text{if } w + E_0^{(i)} = E_\tau^{(N)} , \\ \infty & \text{if } w + E_0^{(i)} > E_\tau^{(N)} . \end{cases} \quad (3.76)$$

First, this shows that values of the work $w > (E_\tau^{(N)} - E_0^{(i)})$ are not allowed, as they would cause the left hand side of the equation to diverge.

Secondly, since only the terms corresponding to the second case survive, we also obtain the equality:

$$\sum_{w} \sum_{\substack{i \\ E_0^{(i)}=(E_\tau^{(N)}-w)}} \left\langle E_0^{(i)} \right| M^w \left| E_0^{(i)} \right\rangle = 1 \,. \tag{3.77}$$

Plugging this back into Eq. (3.74) makes the term corresponding to $E_\tau^{(N)}$ disappear. Hence, we can repeat the same reasoning iteratively, to obtain that M^w has non-zero entries only for a discrete set of outcomes w indexed by $E_0^{(i)}$ and $E_\tau^{(j)}$:

$$M^w_{i,j} = 0 \qquad \text{if} \;\; w \neq (E_\tau^{(j)} - E_0^{(i)}) \,. \tag{3.78}$$

We are now ready to use condition 3. This reads in formulae:

$$\langle w \rangle_{\pi_\beta(H_0)} = \sum_w w \,\mathrm{Tr}\left[M^w \pi_\beta(H_0)\right] = \tag{3.79}$$

$$= \sum_{w,i,j} (E_\tau^{(j)} - E_0^{(i)}) \frac{e^{-\beta E_0^{(i)}}}{\mathcal{Z}_0} \mathrm{Tr}\left[M^w_{i,j}\, \Pi_0^{(i)}\right] = \tag{3.80}$$

$$= \mathrm{Tr}\left[\pi_\beta(H_0)\left(U_\tau^\dagger H_\tau U_\tau - H_0\right)\right] = \tag{3.81}$$

$$= \sum_{i,j} (E_\tau^{(j)} - E_0^{(i)}) \frac{e^{-\beta E_0^{(i)}}}{\mathcal{Z}_0} \mathrm{Tr}\left[\Pi_0^{(i)}\, U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\right] \,, \tag{3.82}$$

where we remind the reader that $\Pi_0^{(i)} := \left|E_0^{(i)}\right\rangle\!\left\langle E_0^{(i)}\right|$ is the projector into the $E_0^{(i)}$ eigenspace, and in the last line we used the fact that $\sum_j \Pi_\tau^{(j)} = \mathbb{1}$ to group together the two contributions. Again, since this relation should be valid for any β, we can equate the i-index in the two sums:

$$\sum_{w,j} (E_\tau^{(j)} - E_0^{(i)}) \mathrm{Tr}\left[M^w_{i,j}\, \Pi_0^{(i)}\right] = \sum_j (E_\tau^{(j)} - E_0^{(i)}) \mathrm{Tr}\left[\Pi_0^{(i)}\, U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\right] \,. \tag{3.83}$$

This step can be understood as follows: sending $\beta \to \infty$ collapses the population of the state to the ground state. Then, the only surviving term in Eq. (3.80) and Eq. (3.82) are the one corresponding to $E_0^{(1)}$. Then, iteratively applying the same argument we obtain that one can

impose the same relation for all the energy levels of H_0, which proves the expression in Eq. (3.83).

Finally, we make one last physically motivated assumption. We suppose that under arbitrary small perturbations of $E_\tau^{(j)}$, the corresponding $M_{i,j}^w$ does not change too much: we note, in fact, that any quench at the end of the protocol does not affect the state of the system, so the dependency of M^w on the final Hamiltonian cannot be too strong. In particular, if the perturbation is small enough such that no level crossing happens, it is sensible to assume that $M_{i,j}^w$ does not change. With this assumption, we can isolate the j-index as well, obtaining:

$$\mathrm{Tr}\left[M_{i,j}^w\right] = \mathrm{Tr}\left[\Pi_0^{(i)}\, U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\right] . \tag{3.84}$$

Using Eq. (3.77) and Eq. (3.78), together with the gap non-degeneracy assumption, we can deduce that $M_{i,j}^w = \Pi_0^{(i)}\, M_{i,j}^w\, \Pi_0^{(i)}$. This implies that $M_{i,j}^w \propto \Pi_0^{(i)}$ (thanks to the non-degeneracy of the spectrum of H_0). Hence, in order to satisfy Eq. (3.84) we must have

$$M_{i,j}^w = \Pi_0^{(i)}\, \mathrm{Tr}\left[\Pi_0^{(i)}\, U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\right] . \tag{3.85}$$

Summing over all the indexes, we then obtain:

$$M^w := \sum_{i,j} \delta(w - (E_\tau^{(j)} - E_0^{(i)}))\, \mathrm{Tr}\left[\Pi_0^{(i)}\, U_\tau^\dagger\, \Pi_\tau^{(j)}\, U_\tau\right]\, \Pi_0^{(i)} = \tag{3.86}$$

$$= \sum_{i,j} \delta(w - (E_\tau^{(j)} - E_0^{(i)}))\, p(E_\tau^{(j)}|E_0^{(i)})\, \Pi_0^{(i)} = M_{\mathrm{TPM}}^w\,, \tag{3.87}$$

proving the claim. □

It directly follows from the result just presented that the no-go theorem on the existence of a probability distribution of work (Theorem 15) can equivalently be stated without the need to explicitly refer to the TPM scheme:

Theorem 21. *There is no function $p(w|\mathcal{P})$ associated to the protocol $\mathcal{P}$ that satisfies the following three conditions:*

1. *(Linear probability) $p(w|\mathcal{P})$ is a probability distribution linear under convex mixing of different protocols. This means that by randomly choosing one protocol between $\mathcal{P}_1$ and $\mathcal{P}_2$ with probability α, the work distribution changes as*

$$p(w|\alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2) = \alpha\, p(w|\mathcal{P}_1) + (1-\alpha)\, p(w|\mathcal{P}_2)\,;$$

2. *(Jarzynski equality) For any thermal initial distribution $p(w|\mathcal{P})$ satisfies the Jarzynski equality;*

3. *(Classicality 2) The average energy change equals the average work.*

This theorem can also be read as the incompatibility between energy conservation and Jarzynski equality for quantum thermodynamic protocols. It should be noticed that since Jarzynski equality imposes weaker constraints than Crooks relations (as the latter implies the first), if one restricts the consideration to non-contextual descriptions, it also follows that fluctuation theorems for quantum work always imply the TPM scheme.

3.2.4 Crooks relations for algorithmic complexity

The wide range of applicability of fluctuation theorems can be considered a hint for some deep underlying principle from which they can be derived, much in the same way that many results in statistical mechanics are applications of probability theory, and in particular of the law of large numbers. For example, in [75] fluctuation theorems were connected to Bayesian retrodiction, with arguments similar to the ones presented here to derive Eq. (3.96). Then, as it is explained in Appendix D, algorithmic complexity gives a foundation to probability theory, a more radical stance would be to derive fluctuation theorems directly in that context. In this section we then present a very incipient tentative in this direction.

Indeed, a theorem resembling Crooks relations can be derived directly from the coding theorem of algorithmic complexity (see Sec. D.2). This states that:

$$K(y|x) = \log \frac{1}{Q_U(x \to y)}\,, \tag{3.88}$$

up to a constant a factor, where the logarithm is in base 2, and we remind the reader that the semi-measure Q_U is defined as (see Eq. (D.14)):

$$Q_U(x \to y) := \sum_{p|\, U(p,x)=y} 2^{-\ell(p)}\,. \tag{3.89}$$

This quantity can be interpreted as follows: it is the unnormalised probability of obtaining y from x running a random program, sampled by generating its bits by sequentially tossing a fair coin (hence, the probability of obtaining each p is given by $2^{-\ell(p)}$). This result can be interpreted as

the fact that the probability of generating a given string by running a program at random on a computer roughly equals how difficult it is to describe it.

Consider now a Maxwell's demon given by a universal Turing machine taking the program from random fluctuations in the environment. Whenever it detects the end of the program, it carries out the computation changing the content of its memory to $y := U(p, x)$. This process is dissipationless, as no energy exchange is needed to carry out the process, and since any transition can in principle move in both direction (there is no time asymmetry). This situation is quite similar to the molecular motors made out of RNA suggested by Bennet.

Then, in principle, the demon could start with a string x in its memory and end up with the string y satisfying $\ell(x) > \ell(y)$, effectively clearing $\ell(x) - \ell(y)$ bits of memory. As discussed above, this would lead to violations of the second law. Fortunately, it also holds that:

$$\frac{Q_U(x \to y)}{Q_U(y \to x)} = 2^{-K(y|x)+K(x|y)}\,, \tag{3.90}$$

up to a constant multiplicative factor. If $K(y|x) = \mathcal{O}(1)$, as it is the case when y is the output of a computation starting from x, we can specialise the above relation to:

$$\frac{Q_U(x \to y)}{Q_U(y \to x)} = 2^{K(x|y)}\,, \tag{3.91}$$

up to a constant multiplicative factor. This tells us that the forward transition is exponentially more probable than the reverse one, so there is an effective emergence of a time arrow. In fact, it should be noticed that the quantity in the exponent of Eq. (3.91) coincides exactly with the minimal dissipation found in Sec. 3.1.2.

This result is still quite incipient, as for example there is no clear definition of thermal equilibrium and thermal states for algorithmic complexity (and there hasn't been so far any compelling motivation to introduce them) and the probabilities in Eq. (3.91) are actually unnormalised. Still, the straightforward connection that can be carried out between the two fields is subtle, possibly cunning, but for sure fascinating.

3.3 How to revert an open system dynamics?

In order to define Crooks relations, one explicitly uses the time reversal of a dissipative dynamics. Whereas for evolutions that preserve the in-

formation it is straightforward to define a concept of time reversal (for example, for a unitary U, the time reversed evolution is just the one induced by $U^\dagger$), the meaning of inverse evolution is more ambiguous for open system dynamics. Just to give an example, it follows from Stinespring's dilation theorem that for any channel Φ there exists a unitary evolution and an ancillary state σ such that [17]:

$$\Phi(\rho) \equiv \mathrm{Tr}_A \left[U(\rho \otimes \sigma) U^\dagger \right] , \tag{3.92}$$

where the partial trace is on the ancillary space. Then, a possible proposal of an inverse channel is given by:

$$\tilde{\Phi}(\rho) \equiv \mathrm{Tr}_A \left[U^\dagger(\rho \otimes \sigma) U \right] . \tag{3.93}$$

It is easy to see that this inverse map is ill-defined. In fact, to each Φ the corresponding dilation is not unique, and different dilations give inequivalent time inversions. For this reason, this naive definition cannot work.

An alternative route is to define the time reversed evolution of a classical stochastic dynamics as the one obtained through Bayes' retrodiction: in this way, one could interpret the inverse map more as an artificial tool for an agent to guess the initial state of an evolution than an actual physical dynamics [76]. Interestingly, as it was pointed out in [75], this construction allows for a straightforward derivation of Crooks relations. Consider, for example, a classical stochastic map Φ. Notice that the coordinates $\Phi_{j,i}$ can be interpreted as the probability of the transition $p(i \to j)$. Moreover, suppose that the prior is given by the thermal state $\pi_\beta(H_0)$. This is a sensible request: one can expect that a generic system, left unperturbed, will eventually thermalise. Encoding this information gives the prior $\pi_\beta(H_0)$. Moreover, since we want to introduce a notion of microreversibility (as it is necessary for Crooks relations to hold), we impose that the inverse transition $\tilde{p}(j \leftarrow i)$ satisfies detailed balance with respect to the equilibrium state:

$$(\pi_\beta(H_0))_i \, p(i \to j) = (\pi_\beta(H_0))_j \, \tilde{p}(j \leftarrow i) \, , \tag{3.94}$$

meaning that, at equilibrium, the probability of observing the forward transition (the left hand side of the equation) should be balanced by the probability of the backwards transition (the right hand side). This is the standard definition of detailed balance for Markov chains (see Eq. 1.329)

and it induces a time symmetric dynamics on the trajectories. As it was briefly mentioned in Sec. 1.5.4, this condition is connected to Bayes' theorem by solving for $\tilde{p}$:

$$\tilde{p}(j \leftarrow i) = \frac{(\pi_\beta(H_0))_i \, p(i \to j)}{(\pi_\beta(H_0))_j} \,. \tag{3.95}$$

In order to interpret the above equation as Bayes' theorem, we also need to demand that $\pi_\beta(H_0)$ is a fixed state of the dynamics. Assuming now that the Hamiltonian of the system quenches to H_τ, we can assume the prior for the reverse transformation to be given by $\pi_\beta(H_\tau)$. Then, the ratio between the probabilities of forward and backwards trajectory are given by:

$$\frac{p_F(i \to j)}{p_R(j \leftarrow i)} = \frac{(\pi_\beta(H_0))_i \, p(i \to j)}{(\pi_\beta(H_\tau))_j \, \tilde{p}(j \leftarrow i)} = \frac{(\pi_\beta(H_0))_j}{(\pi_\beta(H_\tau))_j} = e^{\beta(w_j - \Delta F)} \,, \tag{3.96}$$

where we defined $w_j = (H_\tau - H_0)_j$, the work coming from moving the energy level of the j-th microstate. Iteratively repeating this argument (similarly to what happened in Eq. (3.42)) we obtain the Crooks relations on the full trajectory[7].

The downside of this construction is that there is no universally accepted extension of Bayes theorem to quantum mechanics. For this reason in [2] we suggested an axiomatic approach to single out Bayes' inversion from physical principles. We denote by $\tilde{\Phi}$ a possible inversion of the map Φ. Then, the first two principles demanded are:

1. $\tilde{\Phi}$ is physically implementable;
2. if Φ^{-1} is physically implementable then $\tilde{\Phi} \equiv \Phi^{-1}$.

In the first condition, by *physically implementable* we mean that $\tilde{\Phi}$ is either a stochastic map or a CPTP, depending on the context. This request is needed since we want the reverse of a physical state to be a physical state itself. Moreover, the second condition connects to the discussion we made at the beginning of the section: when Φ completely preserves the information of the initial state (i.e., for permutations or unitaries), Φ^{-1} is physical. Then, we have an unambiguous definition of

[7] This kind of reasoning can be applied to quantum dynamics in the context of quantum trajectories to derive fluctuations theorems similar to the ones in Eq. (3.96). Since this formalism was not used elsewhere in the work, we did not include it here, but we refer the interested reader to [77, 78] for an alternative take on the problem.

time reversal, namely $\tilde{\Phi} = \Phi^{-1}$. This is hardly the general behaviour, as generically Φ^{-1} will map probability distributions to quasiprobability distributions.

The second set of principles we demand characterises the way $\tilde{\Phi}$ behaves with respect to the prior π. We want to demand some version of microreversibility (as we know that Bayes inversion can be used to reconstruct Crooks relations), but rather than asking that Φ and $\tilde{\Phi}$ alone satisfy detailed balance, we impose this condition on the forth-and-back evolution $\tilde{\Phi}\Phi$. This is motivated by two reasons: first, even if asking Φ and $\tilde{\Phi}$ to be connected by detailed balance directly isolates Bayes', it also imposes that π should be a fixed state of the dynamics. For this reason, in order to explore also out-of-equilibrium scenarios, we only demand the fiducial state to be perfectly retrieved by the forth and back evolution, $\tilde{\Phi}\Phi(\pi) = \pi$. The second reason to keep this level of generality is connected to the first: since π and $\Phi(\pi)$ are now different states, it could be possible for them not to commute in general. At the heart of Bayes' theorem there is the definition of a joint probability distribution of the form $P(A, B)$, together with the use of the rule of conditional probability $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$. If π and $\Phi(\pi)$ do not commute, one cannot define in general a joint probability distribution on their microstates. On the other hand, since $\tilde{\Phi}\Phi$ sends π into itself, there is no problem of incompatibility in the latter case. Hence, we demand:

3. $\tilde{\Phi}\Phi$ is detailed balance with respect to the prior π;

4. $\tilde{\Phi}\Phi$ has only positive eigenvalues.

Before explaining what we mean by detailed balance in condition 3, we just point out that the last axiom is more technical in nature, and for this reason we refer the interested reader to the explanation in App. C. On the other hand, whereas there is a universally accepted definition of detailed balance for classical stochastic systems (namely $\tilde{\Phi}\Phi$ should satisfy $(\tilde{\Phi}\Phi)_{i,j}\pi_j = (\tilde{\Phi}\Phi)_{j,i}\pi_i$), in the case of quantum dynamics the issue becomes a bit subtler (see Sec. 1.5.5 for an in-depth explanation). In particular, we only demand detailed balance with respect to $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$, which is a weaker version of the most commonly used definition, and corresponds to the condition:

$$(\tilde{\Phi}\Phi)\,\mathbb{J}_{\sqrt{x}}\big|_{\pi} = \mathbb{J}_{\sqrt{x}}\big|_{\pi}(\tilde{\Phi}\Phi)^{\dagger}\,, \tag{3.97}$$

where $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$ in this context is the superoperator defined by $\mathbb{J}_{\sqrt{x}}\big|_{\pi}[\rho] = \sqrt{\pi}\,\rho\,\sqrt{\pi}$. It should be noticed that it is CP, and reduces to the multipli-

cation by the state π if $[\rho, \pi] = 0$. This choice of $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$ is made for its connection to the Petz' recovery map, defined as:

$$\tilde{\Phi}_P := \mathbb{J}_{\sqrt{x}}\big|_{\pi} \, \Phi^{\dagger} \, \mathbb{J}^{-1}_{\sqrt{x}}\big|_{\Phi(\pi)} \,, \tag{3.98}$$

which is the most usual extension of Bayes' reverse to the quantum regime [79]. Notice that in the same way in which requiring detailed balance on Φ and $\tilde{\Phi}$ for classical dynamics basically defines the Bayes' inversion, the same happens for the Petz' recovery map. Moreover, Eq. (3.97) reduces to the usual definition of detailed balance for stochastic maps if all the quantities in the equation are classical. Finally, also notice that condition 3 also implies that $\tilde{\Phi}\Phi(\pi) = \pi$ (see Sec. 1.5.5), so this requirement also encodes the fact that π is preserved by the composed dynamics.

It came as a surprise that these conditions are not sufficient to single out Bayes' reverse or the Petz recovery. In fact, we found in [2] that one can define a whole class of possible recovery maps, and optimise over them to obtain the best state retrieval channel. This topic is explored in App. C, where it is shown that the optimal state retrieval map outperforms both Bayes' and Petz' maps on the task of guessing the initial state of an evolution.

In order to single out Bayes' reverse map, we propose the following extra axiom:

5. The reversal is an involutive operations, meaning that $\tilde{\tilde{\Phi}} = \Phi$.

This request encodes the usual intuition that the reverting twice the time direction should leave the evolution unaffected. The optimal map we mentioned above does not satisfy this last axiom, but both Bayes' and Petz' maps do, so if one could prove that these are the only two cases in which condition 5 holds we would have a physically motivated foundation of time reversals for open system dynamics. Unfortunately, we were only able to find strong numerical evidences that this is the case for classical dynamics (see App. C) but we weren't able to present an analytical proof of this fact. Still, the evidences collected so far are enough to hope for a positive solution of the problem.

Chapter 4

Thermodynamic transformations close to equilibrium

Once the tendency of systems to thermalisation is established, one can study the way in which energy transfers from one system to another using the thermodynamic formalism. Assuming the continuous thermality of the system during a transformation, for example, one can find universal results which allow the expression of the second law in terms of the maximum efficiency of an engine (the famous Carnot formulation). When finite time effects are included, though, the situation becomes significantly more involved, as non-equilibrium processes usually depend on the specific initial state of the system, the trajectory in the Hamiltonian space and the speed at which the transformation is realised. Midway between these two extreme situations lies the slow driving regime: in this case, by assuming that the transformation is realised in very large but finite time, one finds that the system is always close to equilibrium. As a consequence, this allows to significantly reduce the complexity of the problem, which turns out to be completely describable in terms of differential geometrical quantities defined on the space of thermal states. In fact, it can be shown that the cumulant generating function of the work in the TPM scheme can be expressed solely in terms of Fisher information metrics. Moreover, the slow driving expansion of the average entropy production defines a metric structure whose geodesics correspond to minimally dissipating protocols. In the same way, also the fluctuations in the entropy production define a metric, which coincides with the first only

for classical systems, providing a signature of the creation of quantum coherence during the protocol. Even if these effects cannot be contextual, as they are defined in the TPM scheme, we also show that restricting the study to the slow driving regime is not an impediment to the appearance of genuinely quantum effects.

4.1 Ideal thermodynamics

The natural space in which thermodynamic protocols take place is given by the set of pairs (H, ρ), where H is the Hamiltonian of the system and ρ is its state. Then, thermodynamics transformations can be thought as trajectories in this space, where the pair (H_t, ρ_t) describes the state of the system and its Hamiltonian for each time. In this context, ideal transformations are those for which the only points in the trajectories are of the form $(H, \pi_\beta(H))$.

In particular, define the isothermal slice at inverse temperature β as the set $\bigcup_H \{(H, \pi_\beta(H))\}$. Then, an isothermal transformation is one that lies completely in a single isothermal slice, and can be thought as the the zeroth order approximation to protocols whose duration τ is much greater than any thermalisation timescale of the system. Interestingly, in this case the work becomes a function of state, i.e., it does not depend on the specific trajectory in the thermodynamic space, but only on its endpoints. In fact, by looking at the CGF (Eq. (2.28)) in this case one finds:

$$K^{-\beta w}(\lambda) = \log \int \mathrm{d}w \; p(w) e^{-\beta w \lambda} = \tag{4.1}$$

$$= -\lambda \beta \Delta F + (\lambda - 1)\, S_\lambda(\pi_\beta(H_\tau) || U_\tau \pi_\beta(H_0) U_\tau^\dagger) = \tag{4.2}$$

$$= -\lambda \beta \Delta F \,, \tag{4.3}$$

where in the last line we used the assumption that τ is long enough for the evolved state to be effectively thermal. If one takes the inverse Fourier transform of the last equation, it is easy to show that the corresponding work distribution is a δ-function centred in ΔF, meaning that the work output becomes deterministic and only depends on H_τ and H_0.

The other fundamental type of ideal trajectory is given by the adiabatic transformations, corresponding to protocols $(H_t, \pi_{\beta_t}(H_t))$ such that the product $\beta_t H_t$ is constant along the trajectory. Since thermal states only depend on this product, and not on the specific value of β_t or H_t,

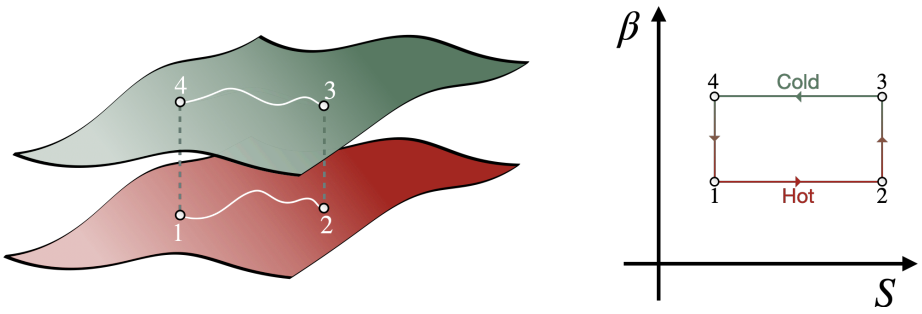


Figure 4.1: Depiction of a Carnot cycle in the thermodynamic space and in the $\beta - S$ plane (where β is the inverse temperature and S the entropy). Different isothermal slices are represented with different colours, and the dashed lines connecting them correspond to the adiabatic quench. In the $\beta - S$ plane horizontal lines correspond to isothermal transformations, while the vertical lines corresponds to the adiabatic quenches.

the state does not change during the transformation, and no heat is exchanged with the environment. Hence, the work can be identified with the change of energy, and there is no increase of entropy during the transformation. If the adiabatic is realised in a single quench $H_1 \to H_2$ the average work performed is given by:

$$\begin{aligned}
\langle w \rangle &= \mathrm{Tr}\left[H_2\, \pi_{\beta_2}(H_2)\right] - \mathrm{Tr}\left[H_1\, \pi_{\beta_1}(H_1)\right] = && (4.4)\\
&= \mathrm{Tr}\left[(H_2 - H_1)\, \pi_{\beta_1}(H_1)\right] = && (4.5)\\
&= \mathrm{Tr}\left[\left(\frac{\beta_1}{\beta_2} H_1 - H_1\right) \pi_{\beta_1}(H_1)\right] = && (4.6)\\
&= \left(\frac{\beta_1}{\beta_2} - 1\right) \mathrm{Tr}\left[H_1\, \pi_{\beta_1}(H_1)\right] = \left(\frac{\beta_1}{\beta_2} - 1\right) U_1\,, && (4.7)
\end{aligned}$$

where we used the condition $\pi_{\beta_1}(H_1) = \pi_{\beta_2}(H_2)$ to group the difference in average energies, together with the defining property of adiabatic transformations $\beta_1 H_1 = \beta_2 H_2$.

It should be noticed that every ideal transformation can be approximated arbitrarily well just in terms of isothermal and adiabatic transformations: in fact, one can rectify the original trajectory in terms of horizontal (within an isothermal slice) and vertical (connecting different slices) lines, obtaining an arbitrary good approximation. Then, it is just sufficient to notice that isothermal transformations completely lie in a single slice, while adiabatics connect different slices.

Since the isothermal work is a function of state, the only way of extracting work along a cycle is by connecting different isothermal slices (in fact, otherwise the cyclicity condition would imply $w = \Delta F = 0$, since the initial and final Hamiltonian coincide). Then, the minimal variation is obtained by connecting two isothermals through adiabatic quenches, realising the famous Carnot cycle. It is worthwhile to look at this protocol more closely. During the hot isothermal an amount $|Q_h|$ of heat is transferred from the bath to the system, while during the cold isothermal an amount Q_c passes from the system to the bath. Thanks to Eq. (4.3), by applying the first law one obtains $Q = \Delta F - \Delta U = -\beta^{-1}\Delta S$, as one would expect from the fact that there is no dissipation during the isothermal transformation. Moreover, thanks to the adiabaticity condition we also have that the heat exchanged during the isothermals satisfies:

$$\begin{aligned} \beta_h\, Q_h &= -\Delta S_h = -(S(\pi_{\beta_h}(H_2)) - S(\pi_{\beta_h}(H_1))) = && (4.8)\\ &= -\left(S(\pi_{\beta_c}(H_3)) - S(\pi_{\beta_c}(H_4))\right) = && (4.9)\\ &= \Delta S_c = -\beta_c\, Q_c\,, && (4.10) \end{aligned}$$

where we used the same convention for H_i as in Fig. 4.1. We also define w to be the overall average work absorbed by the system during the cycle (notice that the sign convention follows the one used in the definition of the first law, Eq. (2.5)). The first law implies that the total change of energy during the cycle is given by $\Delta U = w - Q_c - Q_h = 0$, where the last equality imposes the conservation of energy along the cycle. Then, by defining the efficiency of the Carnot engine to be the ratio between the work output (i.e., $-w$) and the heat absorbed (i.e., $-Q_h$), one obtains:

$$\begin{aligned} \eta_C := \frac{\langle w\rangle}{|Q_h|} &= \frac{Q_c + Q_h}{Q_h} = 1 + \frac{Q_c}{Q_h} = && (4.11)\\ &= 1 - \frac{\beta_h}{\beta_c}\,. && (4.12) \end{aligned}$$

The expression in Eq. (4.12) sets a limit to the maximum efficiency of any engine operating between two thermal reservoirs. As a matter of fact, this result is equivalent to the second law, and indeed constitutes its first formulation ever. Suppose in fact an engine operating between two baths at the same temperature whose efficiency is higher. From Eq. (4.11) and Eq. (4.12) this would mean that $Q_c + Q_h < Q_h - \frac{\beta_h}{\beta_c} Q_h$ (notice that the inequality get reversed since Q_h is negative). But then, one would have that the exchange of heat during the cycle satisfies $\beta_c\, Q_c + \beta_h\, Q_h < 0$ in contradiction with the formulation of the second law in Eq. (3.14).

Additionally, the Carnot engine can be used to define a scale of temperature, as it is apparent from Eq. (4.12). In fact, once the zero-th law of thermodynamics is established (the fact that thermal baths can be grouped in equivalence classes labelled by a parameter called temperature), Eq. (4.12) says that any Carnot engine operating between two baths has an efficiency that only depends on their temperatures. This means that the ratio between the heat transfer during the cold and hot isothermal can be expressed as:

$$\frac{Q_c}{Q_h} = f(T_c, T_h)\,, \tag{4.13}$$

where $f(T_c, T_h)$ is a general function of the temperatures $T_{c/h}$. Then, since the same efficiency should also be given by a cycle that uses an intermediate bath at temperature T_i, the following equality holds:

$$f(T_c, T_h) = \frac{Q_c}{Q_h} = -\frac{Q_c}{Q_i}\frac{Q_i}{Q_h} = -f(T_c, T_i) f(T_i, T_h)\,. \tag{4.14}$$

Whereas the left hand side depends only on T_c and T_h, the right hand side contains the additional variable T_i. This implies that the defining function must be of the form:

$$f(T_c, T_h) = -\frac{g(T_c)}{g(T_h)}\,, \tag{4.15}$$

where $g(T)$ is an arbitrary monotonic function. For simplicity, one choses the convention that g is linear, i.e., $g(T) := T$, which gives back the expression in Eq. (4.12) constructively. Moreover, once a conventional standard temperature is chosen (say the triple-point of the water), a scale can in principle be defined exactly through the use of a Carnot cycles between a bath at the standard temperature and the target temperature one wants to measure. On the one hand, this results demonstrates that the temperature scale is a derived concept, once the zero-th and second law of thermodynamics are established. On the other, it shows that optimising Carnot engines, even without any practical application for work extraction, can yield improvement in the way temperature is measured.

4.2 Thermodynamics close to equilibrium

The assumption that the system is always at equilibrium completely erases any dependency on the dynamics. In general, though, one needs

to know the complete evolution U_t in order to describe the statistics of work, as it is apparent from the expression of the CGF in Eq. (2.28). In this context, the assumption of slow driving regime is key to allow for results that are universal, but that do not depend on the totality of the previous history of the system. Indeed, whereas deriving the thermalisation of local variables directly from the unitary evolution U_t is quite a non-trivial task, once this fact is established many interesting results can be derived from this approximation, as the description of the system is at the right level of abstractness.

In the same way, if one is allowed to assume that the state of the system satisfies $\rho_t := U_t \pi_\beta(H_0) U_t^\dagger \simeq \pi_\beta(H_t) + \varepsilon \delta\rho_t$ for all times (with $\varepsilon \ll 1$), a number of universal results can be proven, which characterise the property of slowly driven systems. Indeed, to this end, let us consider again the CGF, which can be rewritten as:

$$K^{\mathrm{diss}}(\lambda) := K^{-\beta(w-\Delta F)}(\lambda) = (\lambda - 1)\, S_\lambda(\pi_\beta(H_\tau)||\rho_\tau) = \tag{4.16}$$

$$= (\lambda - 1) \int_0^\tau \mathrm{d}t \;\; \frac{\mathrm{d}}{\mathrm{d}t}\, S_\lambda(\pi_\beta(H_t)||\rho_t) = \tag{4.17}$$

$$= (\lambda - 1) \int_0^\tau \mathrm{d}t \;\; \left(\frac{\mathrm{d}}{\mathrm{d}s}\, S_\lambda(\pi_\beta(H_s)||\rho_t) + \frac{\mathrm{d}}{\mathrm{d}s}\, S_\lambda(\pi_\beta(H_t)||\rho_s) \right) \Big|_{s=t}, \tag{4.18}$$

where we isolated the dissipative component by considering the distribution of the variable $\beta(w - \Delta F) \equiv \Sigma$, and we passed to the integral expression so to be able to apply the slow driving expansion. Indeed, even if $\rho_t \simeq \pi_\beta(H_t) + \varepsilon\delta\rho_t$ for each t, the correction to this equality could in principle build up with time. For this reason, one can substitute ρ_t with its slow-driving approximation only when treating quantities local in time, as for example the integrand of Eq. (4.18). Now, it should be noticed that the second term in Eq. (4.18) does not contribute to the integral. Indeed, by explicitly carrying out the differentiation one has:

$$(\lambda - 1)\, \frac{\mathrm{d}}{\mathrm{d}s}\, S_\lambda(\pi_\beta(H_t)||\rho_s)\Big|_{s=t} = \frac{\mathrm{d}}{\mathrm{d}s}\, \log \mathrm{Tr}\left[\pi_\beta(H_t)^\lambda \rho_s^{1-\lambda}\right] = \tag{4.19}$$

$$= -i\, \frac{\mathrm{Tr}\left[\pi_\beta(H_t)^\lambda \left[H_t, \rho_t^{1-\lambda}\right]\right]}{\mathrm{Tr}\left[\pi_\beta(H_t)^\lambda \rho_t^{1-\lambda}\right]} = -i\, \frac{\mathrm{Tr}\left[\left[H_t, \pi_\beta(H_t)^\lambda\right] \rho_t^{1-\lambda}\right]}{\mathrm{Tr}\left[\pi_\beta(H_t)^\lambda \rho_t^{1-\lambda}\right]} = 0\,, \tag{4.20}$$

where we used the fact that $\rho_t^{1-\lambda} = U_t \pi_\beta(H_0)^{1-\lambda} U_t^\dagger$ to take the derivative of the state, and in the last step we used the cyclicity of the trace to

transfer the commutator on the thermal state, which clearly commutes with H_t.

We are now ready to use the slow-driving approximation. This is given by:

$$\frac{K^{\text{diss}}(\lambda)}{\lambda-1} \simeq \int_0^\tau \mathrm{d}t \ \frac{\mathrm{d}}{\mathrm{d}s} S_\lambda(\pi_\beta(H_s)||\pi_\beta(H_t)+\varepsilon\delta\rho_t)\Big|_{s=t} \simeq \tag{4.21}$$

$$\simeq \int_0^\tau \mathrm{d}t \ \varepsilon \frac{\partial^2}{\partial\varepsilon_1\partial\varepsilon_2} S_\lambda(\pi_\beta(H_{t+\varepsilon_1})||\pi_\beta(H_t)+\varepsilon_2\delta\rho_t)\Big|_{\varepsilon_1,\varepsilon_2=0} = \tag{4.22}$$

$$= \frac{\varepsilon}{\lambda-1}\int_0^\tau \mathrm{d}t \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y \ \mathrm{cov}_t^y(\mathbb{J}_{t,L}^{-1}\left[\frac{\mathrm{d}}{\mathrm{d}t}\pi_\beta(H_t)\right], \mathbb{J}_{t,L}^{-1}[\delta\rho_t])\,, \tag{4.23}$$

where in the last line we used the expansion of the λ-Rényi entropy obtained in Eq. (1.180), and the shorthand notation $\mathrm{cov}_t^y := \mathrm{cov}^y_{\pi_\beta(H_t)}$ and $\mathbb{J}_{t,L} := \mathbb{J}_L\big|_{\pi_\beta(H_t)}$ for the Fisher information operator associated with the expansion of the relative entropy. It should be noticed that the extra minus compared with the expression in Eq. (1.180) comes from differentiating in the two arguments of the Rényi entropy, see Eq. (1.60). Finally, it should also be noticed that the derivative of the thermal state takes the form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi_\beta(H_t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{e^{-\beta H_t}}{\mathrm{Tr}\left[e^{-\beta H_t}\right]}\right) = \tag{4.24}$$

$$= -\beta\left(\mathbb{J}_{t,L}[\dot{H}_t] - \frac{e^{-\beta H_t}}{\mathrm{Tr}\left[e^{-\beta H_t}\right]^2}\mathrm{Tr}\left[e^{-\beta H_t}\dot{H}_t\right]\right) = \tag{4.25}$$

$$= -\beta\,\mathbb{J}_{t,L}[\dot{H}_t - \left\langle \dot{H}_t\right\rangle_{\pi_\beta(H_t)}\mathbb{1}] := -\beta\,\mathbb{J}_{t,L}[\Delta\dot{H}_t]\,, \tag{4.26}$$

where we used the Dyson series for the exponential (as explained below Eq. (1.211)) and the cyclicity of the trace to transform the derivative of the partition function in an average, and finally we implicitly defined the operator $\Delta\dot{H}_t := \dot{H}_t - \left\langle \dot{H}_t\right\rangle_{\pi_\beta(H_t)}\mathbb{1}$.

Hence, the slow driving limit of the dissipative CGF takes the form:

$$K^{\text{diss}}(\lambda) = -\beta\int_0^\tau \mathrm{d}t \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y \ \mathrm{cov}_t^y(\Delta\dot{H}_t, \mathbb{J}_{t,L}^{-1}[\varepsilon\delta\rho_t])\,, \tag{4.27}$$

The one presented above serves as a more simple version of the derivation given in [7].

We can discuss now the two main frameworks in which one can use the slow driving expansion. First, consider an evolution given by a sequence of infinitesimal quenches followed by a complete thermalisation. This scenario is particularly appealing because it allows for an unambiguous separation between work and heat: in fact, since during the quench the system is effectively isolated, the change of energy during this step is constituted completely by work; similarly, during the thermalisation, as the Hamiltonian stay fixed, the change of energy is solely given by the transfer of heat. Moreover, since one always starts in thermal equilibrium, the work defined according to WWS coincides with the TPM scheme. Then, considering a sequence of quenches of the form $H_1 \to H_2 \to \cdots \to H_N$, where $H_{i+1} - H_i \simeq \frac{\dot{H}_i}{N}$, one can approximate the discrete trajectory by a continuous path H_t with $t \in [0, N]$. Since the thermalisation step completely erases all the informations about the previous step, the work output at each step is completely independent from all the other steps. For this reason, the CGF in this context is given by:

$$K^{-\beta w}(\lambda) = \sum_{i=1}^{N} \log \int \mathrm{d}w_i \; p(w_i) e^{-\beta w_i \lambda} = \tag{4.28}$$

$$= -\lambda\beta\Delta F + (\lambda - 1) \sum_{i=1}^{N} \; S_\lambda(\pi_\beta(H_{i+1}) || \pi_\beta(H_i)) = \tag{4.29}$$

$$= -\lambda\beta\Delta F + (\lambda - 1) \int_0^N \mathrm{d}t \; \frac{1}{N} \frac{\mathrm{d}}{\mathrm{d}s} \, S_\lambda(\pi_\beta(H_s) || \pi_\beta(H_t)) \Big|_{s=t} , \tag{4.30}$$

where in Eq. (4.28) we used the fact that the CGF of a sum of independent variables is just the sum of the single CGF, then in Eq. (4.29) one can read the sum of dissipative CGF of a single quench (and we highlighted the total difference in free energy), and in the last step we used the approximation $S_\lambda(\pi_\beta(H_{i+1}) || \pi_\beta(H_i)) \simeq \frac{1}{N} \frac{\mathrm{d}}{\mathrm{d}s} \, S_\lambda(\pi_\beta(H_s) || \pi_\beta(H_t)) \big|_{s=t}$ to pass to the integral. Without even the need to pass from Eq. (4.21) we can now give the expression for the dissipative CGF as:

$$K^{\mathrm{diss}}(\lambda) = (\lambda - 1) \int_0^N \mathrm{d}t \; S_\lambda(\pi_\beta(H_t) - \frac{\beta}{N} \mathbb{J}_{t,L}[\Delta\dot{H}_t] || \pi_\beta(H_t)) = \tag{4.31}$$

$$= -\frac{\beta^2}{2N^2} \int_0^N \mathrm{d}t \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y \; \mathrm{cov}_t^y(\Delta\dot{H}_t, \, \Delta\dot{H}_t) = \tag{4.32}$$

$$= -\frac{\beta^2}{2N} \int_0^1 \mathrm{d}s \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y \; \mathrm{cov}_{Ns}^y(\Delta\dot{H}_{Ns}, \, \Delta\dot{H}_{Ns}) \, , \tag{4.33}$$

where in the last line we changed the variables as $t \to (Ns)$ in order to highlight the scaling of the dissipative CGF, which goes to zero as $1/N$. Indeed, in the limit $N \to \infty$ the system is effectively always at equilibrium, so there should be no dissipation.

Another standard situation worth exploring is the one in which the state evolves according to the driven Lindbladian equation;

$$\dot{\rho}_t = \mathcal{L}_t(\rho_t)\,; \qquad\qquad \mathcal{L}_t(\pi_\beta(H_t)) = 0\,, \tag{4.34}$$

meaning we have a Markovian evolution such that the thermal state corresponding to the driven Hamiltonian H_t is the instantaneous fixed point of the dynamics. Moreover, we also assume that this is the unique equilibrium state of the evolution. Then, if the duration of the protocol goes to infinity, it is clear that for each H_t the system would be in the corresponding thermal state. With this hindsight in mind, it is useful to write the state of the system as $\rho_t := \pi_\beta(H_t) + \delta\rho_t$ and to pass to the variable $s := t/\tau$, so that $s \in [0, 1]$. Then, the Lindbladian equation transforms to:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\pi_\beta(H_{\tau s}) + \delta\rho_{\tau s}\right) = \tau\,\frac{\mathrm{d}}{\mathrm{d}t}\,\rho_t = \tau\,\mathcal{L}_{\tau s}(\delta\rho_{\tau s})\,. \tag{4.35}$$

We can now solve for $\delta\rho_t$. First, in order to invert the Lindbladian we need to introduce the concept of Drazin inverse $\mathcal{L}_t^+$. This is a pseudoinverse satisfying the following three conditions for any trace-class operator A:

1. given the fixed point $\pi_\beta(H_t)$ of $\mathcal{L}_t$, it holds that $\mathcal{L}_t^+(\pi_\beta(H_t)) = 0$;
2. it maps operators to traceless operators, i.e., $\mathrm{Tr}\left[\mathcal{L}_t^+(A)\right] = 0$;
3. for any trace class operator A, it holds that:

$$\mathcal{L}^+\mathcal{L}(A) = \mathcal{L}\mathcal{L}^+(A) = A - \pi_\beta(H_t)\mathrm{Tr}\left[A\right]\,. \tag{4.36}$$

Interestingly, if a Drazin inverse can be found, then it is unique [9]. Indeed, the following functional expression:

$$\mathcal{L}_t^+(A) := -\int_0^\infty \mathrm{d}\nu\; e^{\nu\mathcal{L}_t}(A - \pi_\beta(H_t)\mathrm{Tr}\left[A\right]) \tag{4.37}$$

satisfies all the conditions above, as it can be verified from direct calculation, proving the existence of the Drazin inverse. Then, by applying $\mathcal{L}_t^+$

on both sides of Eq. (4.35) and grouping the terms containing $\delta\rho_t$ on the left, we obtain:

$$\left(\mathbb{I} - \frac{1}{\tau}\mathcal{L}^+_{\tau s}\frac{\mathrm{d}}{\mathrm{d}s}\right)(\delta\rho_{\tau s}) = \frac{1}{\tau}\mathcal{L}^+_t\left(\frac{\mathrm{d}}{\mathrm{d}s}\,\pi_\beta(H_{\tau s})\right). \tag{4.38}$$

Finally, for τ long enough one can invert the operator on the left to obtain the perturbative expansion:

$$\delta\rho_t = \sum_{i=1}^{\infty}\left(\mathcal{L}^+_t\frac{\mathrm{d}}{\mathrm{d}t}\right)^n(\pi_\beta(H_t)) = -\beta\mathcal{L}^+_t(\mathbb{J}_{t,L}[\Delta\dot{H}_t]) + \mathcal{O}\left(\frac{1}{\tau^2}\right), \tag{4.39}$$

where in the limit $\tau \gg 1$ we can keep just the first term. Then, in this case the CGF in Eq. (4.27) takes the form:

$$K^{\mathrm{diss}}(\lambda) = \beta^2\int_0^\tau \mathrm{d}t\int_0^\lambda \mathrm{d}x\int_x^{1-x}\mathrm{d}y\,\mathrm{cov}^y_t(\Delta\dot{H}_t,\,\mathbb{J}^{-1}_{t,L}[\mathcal{L}^+_t(\mathbb{J}_{t,L}[\Delta\dot{H}_t])])\,. \tag{4.40}$$

It should be noticed that if the thermalisation timescales are all equal, then one obtains back the expression from Eq. (4.33). Indeed, this corresponds to setting $\mathcal{L}^+_t = -\tau^{\mathrm{eq}}\,\mathbb{1}_{\frac{1}{2}}$. This correspondence will be made more precise in the next section.

A standard assumption in this context is that the Lindbladian is detailed balance at each time, which implies that $\mathbb{J}^{-1}_{t,L}\mathcal{L}^+_t\mathbb{J}_{t,L} = (\mathcal{L}^\dagger_t)^+$. This reads:

$$(\mathcal{L}^\dagger_t)^+(A) := -\int_0^\infty \mathrm{d}\nu\; e^{\nu\mathcal{L}^\dagger_t}(A - \mathbb{1}\mathrm{Tr}\,[A\,\pi_\beta(H_t)])\,. \tag{4.41}$$

Since $\mathrm{Tr}\left[\Delta\dot{H}_t\,\pi_\beta(H_t)\right] = \mathrm{Tr}\left[\dot{H}_t\,\pi_\beta(H_t)\right] - \mathrm{Tr}\left[\dot{H}_t\,\pi_\beta(H_t)\right] = 0$, for detailed balance Lindbladians the expression in Eq. (4.40) simplifies to:

$$K^{\mathrm{diss}}(\lambda) = \beta^2\int_0^\tau \mathrm{d}t\int_0^\infty \mathrm{d}\nu\int_0^\lambda \mathrm{d}x\int_x^{1-x}\mathrm{d}y\,\mathrm{cov}^y_t(\Delta\dot{H}_t,\,e^{\nu\mathcal{L}^\dagger_t}(\Delta\dot{H}_t))\,, \tag{4.42}$$

where we can interpret the integrand as the autocorrelation function of $\Delta\dot{H}_t$ as measured by the y-covariance.

Now that the general structure of the statistics of work close to equilibrium has been laid down, we can start studying more fine-grained properties.

4.3 Average dissipation and the emergence of a metric structure

Average quantities play a central role in thermodynamics. Indeed, whenever a rule holds in classical thermodynamics (without fluctuations), then it can be translated to the stochastic framework by taking the average of each random variable. The reason for this is that for macroscopic systems the fluctuations around the average go to zero, and the system becomes effectively deterministic, so that all the quantities of interest collapse on their average value. Hence, any result in classical thermodynamics can be directly translated to the corresponding averaged equation.

In particular, we remind the reader that the first cumulant of a stochastic variable coincides with its average, so we can compute the average work from the expression:

$$\langle w \rangle = -\beta^{-1} \frac{\mathrm{d}}{\mathrm{d}\lambda} K^{-\beta w}(\lambda)\Big|_{\lambda=0} = \tag{4.43}$$

$$= \Delta F - \beta^{-1} \frac{\mathrm{d}}{\mathrm{d}\lambda} \left((\lambda - 1) S_\lambda(\pi_\beta(H_\tau)||\rho_\tau)\right)\Big|_{\lambda=0} . \tag{4.44}$$

Using the integral expression in Eq. (1.169) to express the Rényi entropy, we can compute the average dissipation as:

$$\langle \Sigma \rangle = \beta(\langle w \rangle - \Delta F) = \tag{4.45}$$

$$= -\frac{\mathrm{d}}{\mathrm{d}\lambda} \log \left(1 + \int_0^\lambda \mathrm{d}x \ \mathrm{Tr}\left[\pi_\beta(H_\tau)^x (\log \pi_\beta(H_\tau) - \log \rho_\tau) \rho_\tau^{1-x}\right]\right) = \tag{4.46}$$

$$= -\mathrm{Tr}\left[(\log \pi_\beta(H_\tau) - \log \rho_\tau)\rho_\tau\right] = S(\rho_\tau||\pi_\beta(H_\tau)) . \tag{4.47}$$

The expression just obtained connects the average irreversibility with the distinguishability between the evolved state and the final thermal state as measured by the relative entropy. We remind the reader that a similar application of the Kullback-Leibler divergence had been pointed out in Eq. (1.209), where it was used to divide the out of equilibrium free energy into an equilibrium contribution and a term quantifying the distance from the thermal state.

Thanks to the positivity of the relative entropy, Eq. (4.47) directly implies the second law:

$$\langle \Sigma \rangle = S(\rho_\tau||\pi_\beta(H_\tau)) \geq 0 , \tag{4.48}$$

with equality if and only if the final state is exactly thermal, i.e., only in the infinite time limit. Notice that all the information about the history of the evolution is completely contained in ρ_τ. Yet, if the dynamics is Markovian, this description is not satisfactory, and a better suited approach is to examine how work (or dissipation) is created instantaneously. To this end, we can use Eq. (2.35), to express the average work of a system starting at equilibrium as:

$$\langle w \rangle = \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\rho_t \dot{H}_t\right] = \tag{4.49}$$

$$= -\beta^{-1} \int_0^\tau \mathrm{d}t \ \left(\mathrm{Tr}\left[\rho_t \frac{\mathrm{d}}{\mathrm{d}t} \log \frac{e^{-\beta H_t}}{\mathcal{Z}_t}\right] - \frac{\mathrm{d}}{\mathrm{d}t} \log \mathcal{Z}_t\right) = \tag{4.50}$$

$$= \Delta F + \beta^{-1} \int_0^\tau \mathrm{d}t \ \frac{\mathrm{d}}{\mathrm{d}s} S(\rho_t || \pi_\beta(H_s))\Big|_{s=t}, \tag{4.51}$$

where now ρ_t is the effective state of the system at each time, and we performed the integral over the logarithm of the partition function to get the difference in free energy. Then, the average dissipation can be rewritten as:

$$\langle \Sigma \rangle = \int_0^\tau \mathrm{d}t \ \frac{\mathrm{d}}{\mathrm{d}s} S(\rho_t || \pi_\beta(H_s))\Big|_{s=t} = \tag{4.52}$$

$$= \int_0^\tau \mathrm{d}t \ \left(\frac{\mathrm{d}}{\mathrm{d}t} S(\rho_t || \pi_\beta(H_t)) - \frac{\mathrm{d}}{\mathrm{d}s} S(\rho_s || \pi_\beta(H_t))\Big|_{s=t}\right) = \tag{4.53}$$

$$= S(\rho_\tau || \pi_\beta(H_\tau)) + \int_0^\tau \mathrm{d}t \ \left(-\frac{\mathrm{d}}{\mathrm{d}s} S(\rho_s || \pi_\beta(H_t))\Big|_{s=t}\right), \tag{4.54}$$

where it is important to keep in mind that, despite using the same notation, the ρ_τ in this expression and the one in Eq. (4.48) are different: whereas in Eq. (4.48) all the information about the dynamics is encoded in the state, in the equation above, due to the Markovianity of the evolution, ρ_τ is the effective state of the system which only retains informations that are local in time.

We can now analyse the two contributions to Eq. (4.54). As explained above, the first term in the equation only depends on the endpoints and it is a measure of the amount of the non-equilibrium free energy left in the state ρ_τ. On the other hand, since $\pi_\beta(H_t)$ is the fixed state of the dynamics at time t, it follows from the monotonicity of the relative entropy that the integrand in Eq. (4.54) (which we denote by $\langle \dot{\Sigma}_t \rangle$) is positive for Markovian evolutions (i.e., CP-divisible), and zero if and

only if the system is always at thermal equilibrium $\rho_t \equiv \pi_\beta(H_t)$. Hence, in the limit $\tau \to \infty$ one reaches the minimum of $\langle \dot{\Sigma}_t \rangle$.

This consideration suggests the following expansion. Denote by $\{\lambda_t^i\}$ the time dependent scalar parameters of the evolution (see, for example, the canonical decomposition of a driven Hamiltonian in Eq. (2.1)). In the limit of slow driving, one expects that $\dot{\lambda}_t^i \to 0$. Then, one can expand the average dissipation rate around this limit to obtain:

$$\langle \dot{\Sigma}_t \rangle = \dot{\lambda}_t^i \cancel{\partial_i \langle \dot{\Sigma}_t \rangle}\big|_{\{\dot{\lambda}_t\}\equiv 0} + \dot{\lambda}_t^i \left(\partial_i \partial_j \, \langle \dot{\Sigma}_t \rangle \big|_{\{\dot{\lambda}_t\}\equiv 0} \right) \dot{\lambda}_t^j + \mathcal{O}\left(||\dot{\lambda}||^3 \right) , \quad (4.55)$$

where we denote by $\partial_i = \frac{\partial}{\partial \lambda_t^i}$ and we implicitly sum over repeated indices. It should also be noticed that the first derivative cancels since we are expanding around a minimum. For the same reason, we know that the Hessian $(g_{i,j})_t = \partial_i \partial_j \, \langle \dot{\Sigma}_t \rangle \big|_{\{\dot{\lambda}_t\}\equiv 0}$ is a positive definite matrix. Hence, the average dissipation can be approximated for $\tau \to \infty$ by:

$$\langle \Sigma \rangle = \int_\gamma \mathrm{d}t \;\; \dot{\lambda}_t^i \, (g_{i,j})_t \dot{\lambda}_t^j \; + \mathcal{O}\left(\frac{1}{\tau^2} \right) , \quad (4.56)$$

where we used the notation γ to denote the trajectory of the driving in the parameter space.

It is interesting to study the properties of $(g_{i,j})_t$. First, notice that from standard arguments in linear response theory it follows that it should depend smoothly on the thermal state $\pi_\beta(H_t)$. Moreover, as it was argued above it is positive definite and symmetric, being the Hessian of a function around its minimum. These three properties (smoothness, positivity and symmetry) make $(g_{i,j})_t$ a metric on the space of parameters. Indeed, we can interpret Eq. (4.56) as the energy functional or the action of the curve γ with respect to the metric g [40]. This name comes from the formal analogy between Eq. (4.56) and the action of a system of free particles with mass tensor given by g.

The mathematical formalism that one introduces in this way, rather than being solely a theoretical framework in which to describe the same phenomenology, also has practical applications, and offers a set of powerful tools to give very general results about the entropy production. For example, it is a standard result in differential geometry (corresponding to locally applying the Cauchy-Schwarz inequality), that the energy functional of a curve γ is bounded by its length squared. This translates to the following bound on the average dissipation:

$$\langle \Sigma \rangle \geq \frac{1}{\tau} \, \ell_\gamma^2 \, , \quad (4.57)$$

where the length of γ is defined as:

$$\ell_\gamma = \int_\gamma \mathrm{d}t \ \sqrt{\dot{\lambda}_t^i \, (g_{i,j})_t \dot{\lambda}_t^j} \, . \tag{4.58}$$

Eq. (4.57) takes the name of "thermodynamic length inequality" [80] and shows that for slow driving systems the geodesic distance between two points naturally corresponds to the square of the minimal dissipation. Indeed, the inequality in Eq. (4.57) can be saturated if γ is a geodesic. Looking at what this means from the point of view of Cauchy-Schwarz, the equality in Eq. (4.57) corresponds to the request that $\sqrt{\dot{\lambda}_t^i \, (g_{i,j})_t \dot{\lambda}_t^j}$ is parallel to the constant function with respect to the L^2 scalar product. Since in this case being parallel corresponds to being proportional to each other, it turns out that optimal curves are characterised by having a constant entropy production rate $\langle \dot{\Sigma}_t \rangle$. This behaviour is exemplified in Fig. 4.2, where the Landauer erasure of a qubit is realised through a linear and a geodesic protocol. Especially in the last panel it becomes apparent that the geodesic drive distributes the entropy production more uniformly during the protocol.

Apart from giving insights on the nature of dissipation, introducing a metric structure can also have practical applications in the design of minimally dissipating protocols. Indeed, without needing to pass from the interpretation above, one can directly solve a system of differential equations, the geodesic equations, to obtain the optimal driving. These read:

$$\ddot{\lambda}_t^i + \Gamma_{j,k}^i\big|_{\lambda_t} \, \dot{\lambda}_t^j \, \dot{\lambda}_t^k = 0 \, , \tag{4.59}$$

where we again sum over repeated indices, and $\Gamma_{j,k}^i$ denote the Christoffel symbols, which are given by:

$$\Gamma_{j,k}^i|_{\lambda_t} = \frac{1}{2} g^{i,l} \left(\partial_j g_{l,k} + \partial_k g_{j,l} - \partial_l g_{j,k} \right)|_{\lambda_t} \, . \tag{4.60}$$

Here, $g^{i,l}$ is the inverse of the metric, and we use the shorthand notation $\partial_i g_{j,k}|_{\lambda_t} \equiv (\partial g_{j,k}/\partial \lambda_i)|_{\lambda=\lambda_t}$. Indeed, we obtained the optimal drive in Fig. 4.2 exactly through this method.

It should be noticed that the dissipation resulting from a geodesic drive can, in fact, behave significantly different from the one of suboptimal protocols. Consider again the Landauer erasure presented above, i.e., a qubit driven with the Hamiltonian $H_t = E(t) \, |1\rangle\langle 1|$. In Fig. 4.3 we

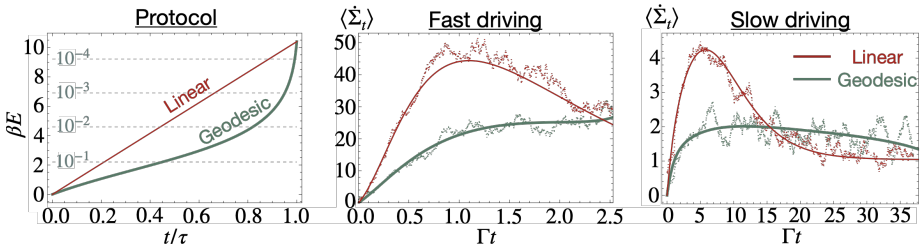


Figure 4.2: Comparison of a geodesic drive of a qubit with the corresponding linear protocol, where the Hamiltonian is varied as: $H_t = E(t)\,|1\rangle\langle 1|$. The first panel compares the two drivings. If one looks at the populations of the excited state for each energy (the numbers in grey), it is apparent that the geodesics is slower when the state is highly populated, and ramps up at the end of the protocol. Comparing this behaviour with the expression of the work in Eq. (4.49) gives an intuition of why this protocol is optimal. The last two panels show the entropy production rate for a fast and a slow realisation of the protocol. The solid lines are the theory prediction, while the dotted lines corresponds to actual experimental data from [3]. As it is apparent, the geodesic drive tends to a more uniform distribution of the entropy production. Interestingly, this effect appears already for fast driving speed.

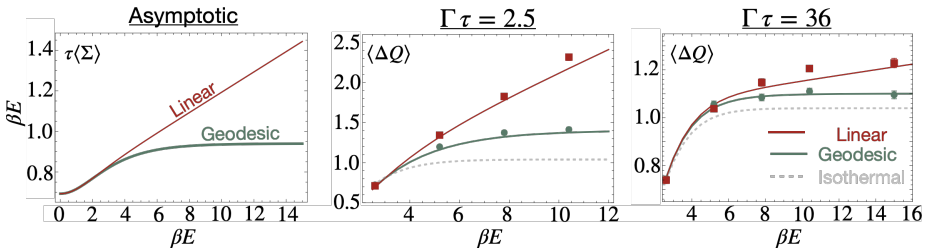


Figure 4.3: Comparison between the dissipation for a linear drive and a geodesic drive of $H_t = E(t)\,|1\rangle\langle 1|$, as a function of the final energy $E(\tau)$. The first panel shows the limit $\lim_{\tau\to\infty} \tau\langle\Sigma\rangle$, which singles out the slow driving contribution. The other two panels show the data for finite times (solid lines correspond to the theoretical predictions, while the points are measured in the experiment [3]). Notice that in the last two panels we plot the average heat and not the average dissipation. The grey line corresponds to the ideal case, $\langle\Delta Q\rangle = -\beta^{-1}\Delta S$.

plot the dissipation as a function of the final energy $E(\tau)$. In particular, in the first panel we consider the asymptotic behaviour of the $\tau\langle\Sigma\rangle$ in the limit $\tau \to \infty$. In this way, the only term surviving is the integral in Eq. (4.56). Then, it can be seen that as $E(\tau) \to \infty$ the dissipation associated with the linear drive diverges, while in the case of a geodesic drive it saturates to a constant. This stark difference can also be noticed when finite time effects are taken into account (interestingly, also for relatively fast drivings, as in the middle panel of Fig. 4.3).

4.3.1 Metric structure in different frameworks

Another strength of the geometric approach is its flexibility. Indeed, the metric g can be explicitly computed in many frameworks: as it was explained in [6] the same structural form can be obtained for a system undergoing quenches, for slowly driven Lindbladian dynamics, or even in the linear response regime of a unitary evolution. For example, by direct differentiation of Eq. (4.33) one can easily obtain:

$$\langle\Sigma\rangle = \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\beta^2}{2N}\int_0^1 \mathrm{d}s \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y\, \mathrm{cov}^y_{Ns}(\Delta\dot{H}_{Ns},\, \Delta\dot{H}_{Ns})\Big|_{\lambda=0} = \tag{4.61}$$

$$= \frac{\beta^2}{2N}\int_0^1 \mathrm{d}s \int_0^1 \mathrm{d}y\, \mathrm{cov}^y_{Ns}(\Delta\dot{H}_{Ns},\, \Delta\dot{H}_{Ns}) = \tag{4.62}$$

$$= \frac{\beta^2}{2N}\int_0^1 \mathrm{d}s\, \mathrm{Tr}\left[\Delta\dot{H}_{Ns}\, \mathbb{J}_{t,L}[\Delta\dot{H}_{Ns}]\right], \tag{4.63}$$

where in the last line we used the definition of $\mathbb{J}_{t,L}$ in terms of the y-covariance. This metric is well known in statistical mechanics and takes the name of Kubo-Mori-Bogoliubov (KMB) inner product (see Sec. 1.4.6). Moreover, thanks to the relation in Eq. (1.215), one can rewrite the average dissipation for quenched systems in terms of the second derivative of the free energy, i.e.:

$$\langle\Sigma\rangle = \frac{\beta^2}{2N}\int_0^1 \mathrm{d}s\, \frac{\partial^2}{\partial\varepsilon_1\partial\varepsilon_2}\, \log \mathcal{Z}_{H_{Ns}+\varepsilon_1\Delta\dot{H}_{Ns}+\varepsilon_2\Delta\dot{H}_{Ns}}\Big|_{\varepsilon_1,\varepsilon_2=0} = \tag{4.64}$$

$$= -\frac{\beta^3}{2N}\int_0^1 \mathrm{d}s\, \frac{\partial^2}{\partial\varepsilon_1\partial\varepsilon_2}\, F(H_{Ns}+\varepsilon_1\Delta\dot{H}_{Ns}+\varepsilon_2\Delta\dot{H}_{Ns})\Big|_{\varepsilon_1,\varepsilon_2=0}. \tag{4.65}$$

This equality shows how dissipation in the linear regime is related to the susceptibility of the free energy along the trajectory (a fact that

was already noticed in the classical scenario in [80–84]). This behaviour constitutes the backbone of average entropy production, and any modifications only account for the different equilibration timescales present in the system. Notice in fact that for the quench model the thermalisation steps imposes a unique timescale for all the degree of freedom of the system, which simplifies the dynamics and allows for the derivation of cleaner results.

In order to make the discussion above more precise, consider now the average dissipation associated to a Lindbladian dynamics. Differentiating Eq. (4.40), we obtain:

$$\langle\Sigma\rangle = -\frac{\mathrm{d}}{\mathrm{d}\lambda}\beta^2 \int_0^\tau \mathrm{d}t \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y\, \mathrm{cov}_t^y(\Delta\dot{H}_t,\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])])\Big|_{\lambda=0} = \tag{4.66}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \int_0^1 \mathrm{d}y\, \mathrm{cov}_t^y(\Delta\dot{H}_t,\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]) = \tag{4.67}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\Delta\dot{H}_t\, \mathbb{J}_{t,L}[\mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]]\right] = \tag{4.68}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\Delta\dot{H}_t\, \mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])\right] , \tag{4.69}$$

where in Eq. (4.68) we used the expression of $\mathbb{J}_{t,L}$ in terms of the y-covariance, and in the last step we simplified the two KMB inner products. We can highlight the fact that the trace in Eq. (4.69) is real by rewriting it as:

$$\langle\Sigma\rangle = -\beta^2 \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\mathbb{J}_{t,L}[(\mathcal{L}_t^+)^\dagger(\Delta\dot{H}_t)]\, \Delta\dot{H}_t\right] = \tag{4.70}$$

$$= -\frac{\beta^2}{2} \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\Delta\dot{H}_t\, \left(\mathcal{L}_t^+ + \mathbb{J}_{t,L}(\mathcal{L}_t^+)^\dagger\mathbb{J}_{t,L}^{-1}\right)(\mathbb{J}_{t,L}[\Delta\dot{H}_t])\right] , \tag{4.71}$$

where in the first line we moved the two operators from the right to the left, and in the second step we used the cyclicity of the trace to add Eq. (4.69) and Eq. (4.70). It should be noticed that if $\mathcal{L}$ is detailed balance, then the two contributions in the curved brackets coincide and one obtains again the original expression (i.e., since $\mathcal{L}$ is self-adjoint with respect to all the Fisher metrics, one does not need to manipulate Eq. (4.69) to make it evidently real).

Suppose now that the original Lindbladian has right eigenoperators given by $\{X_t^i\}$ and left eigenoperators given by $\{Y_t^i\}$, meaning that:

$$\mathcal{L}_t(X_t^i) = -\frac{1}{\tau_t^i}\, X_t^i\,; \qquad\qquad \mathbb{J}_{t,L} \circ \mathcal{L}_t^\dagger \circ \mathbb{J}_{t,L}^{-1}(Y_t^i) = -\frac{1}{\tilde{\tau}_t^i}\, Y_t^i\,. \tag{4.72}$$

It should be noticed that whereas the definition of the right eigenoperator is unambiguous, in the left eigenoperators one also has the freedom to choose the scalar product with respect to which to define them (in this case we used the Fisher metric $\mathbb{J}_{t,L}^{-1}$). It should be noticed, though, that both $\{X_t^i\}$ and $\{Y_t^i\}$ can be chosen to be Hermitian operators, since $\mathcal{L}_t$ is Hermitian preserving. Hence, without loss of generality we assume self-adjointness of both left and right eigenoperators.

By assumption $\pi_\beta(H_t)$ is always a right eigenoperator with eigenvalue zero, i.e., $\mathcal{L}_t(\pi_\beta(H_t)) = 0$. Moreover, it is also a left eigenoperator, since one has that:

$$\mathbb{J}_{t,L} \circ \mathcal{L}_t^\dagger \circ \mathbb{J}_{t,L}^{-1}(\pi_\beta(H_t)) = \mathbb{J}_{t,L} \circ \mathcal{L}_t^\dagger(\mathbb{1}) = 0\,, \tag{4.73}$$

where in the last step we used the fact that in order for the Lindbladian to generate a trace preserving evolution, it should hold that $\mathcal{L}_t^\dagger(\mathbb{1}) = 0$. This directly implies that:

$$\mathrm{Tr}\left[X_t^i\right] = \mathrm{Tr}\left[X_t^i\right] = -\tau_t^i\,\mathrm{Tr}\left[\mathcal{L}_t(X_t^i)\right] = -\tau_t^i\,\mathrm{Tr}\left[\mathcal{L}_t^\dagger(\mathbb{1})\, X_t^i\right] = 0\,; \tag{4.74}$$

$$\mathrm{Tr}\left[Y_t^i\right] = -\tilde{\tau}_t^i\,\mathrm{Tr}\left[\mathbb{J}_{t,L}\,\mathcal{L}_t^\dagger\,\mathbb{J}_{t,L}^{-1}(Y_t^i)\right] = -\tilde{\tau}_t^i\,\mathrm{Tr}\left[\mathbb{J}_{t,L}^{-1}(Y_t^i)\,\mathcal{L}_t(\mathbb{J}_{t,L}[\mathbb{1}])\right] = 0\,, \tag{4.75}$$

where in the last step one uses $\mathbb{J}_{t,L}[\mathbb{1}] = \pi_\beta(H_t)$. Moreover, it is also straightforward to verify that:

$$\mathrm{Tr}\left[X_t^i\,\mathbb{J}_{t,L}^{-1}[Y_t^j]\right] = -\tau_t^i\,\mathrm{Tr}\left[\mathcal{L}_t(X_t^i)\,\mathbb{J}_{t,L}^{-1}[Y_t^j]\right] \tag{4.76}$$

$$= -\tau_t^i\,\mathrm{Tr}\left[X_t^i\,\mathbb{J}_{t,L}^{-1}\left[\mathbb{J}_{t,L}(\mathcal{L}_t^\dagger)\mathbb{J}_{t,L}^{-1}[Y_t^j]\right]\right] = \tag{4.77}$$

$$= \frac{\tau_t^i}{\tilde{\tau}_t^j}\,\mathrm{Tr}\left[X_t^i\,\mathbb{J}_{t,L}^{-1}[Y_t^j]\right]\,. \tag{4.78}$$

Hence, in order to have consistency $\mathrm{Tr}\left[X_t^i\,\mathbb{J}_{t,L}^{-1}[Y_t^j]\right] = \delta_{i,j}$ should hold (where we impose without loss of generality that the scalar product is normalised) and $\tau_t^i = \tilde{\tau}_t^i$. Hence, there is an unambiguous definition for the eigenvalues.

The considerations above allow us to express the Drazin inverse directly in terms of the thermalisation timescales τ_t^i, i.e.,

$$\mathcal{L}_t^+(X_t^i) := -\int_0^\infty \mathrm{d}\nu\; e^{\nu\mathcal{L}_t}(X_t^i - \pi_\beta(H_t)\mathrm{Tr}\left[X_t^i\right]) = \tag{4.79}$$

$$= \left(-\int_0^\infty \mathrm{d}\nu\; e^{-\nu/\tau_t^i}\right) X_t^i = -\tau_t^i\, X_t^i\,. \tag{4.80}$$

Indeed, any pseudoinverse coincides with the actual inverse in the subspace in which the original map is full-rank. Moreover, one also has that:

$$\mathbb{J}_{t,L}(\mathcal{L}_t^+)^\dagger \mathbb{J}_{t,L}^{-1}(Y_t^i) := \tag{4.81}$$

$$= -\int_0^\infty \mathrm{d}\nu\; e^{\nu\mathbb{J}_{t,L}(\mathcal{L}_t^+)^\dagger\mathbb{J}_{t,L}^{-1}}(Y_t^i - \pi_\beta(H_t)\mathrm{Tr}\left[\mathbb{J}_{t,L}^{-1}[Y_t^i]\pi_\beta(H_t)\right]) = \tag{4.82}$$

$$= \left(-\int_0^\infty \mathrm{d}\nu\; e^{-\nu/\tau_t^i}\right) Y_t^i = -\tau_t^i\, Y_t^i\,. \tag{4.83}$$

We are now ready to connect the KMB inner product to the average dissipation during a slowly driven Lindbladian evolution. First, notice that $\mathbb{J}_{t,L}[\Delta\dot{H}_t]$ is a traceless operator, so it can be decomposed in terms of $\{X_t^i\}$ as:

$$\mathbb{J}_{t,L}[\Delta\dot{H}_t] = \sum_i \mathrm{Tr}\left[Y_t^i \mathbb{J}_{t,L}^{-1}[\mathbb{J}_{t,L}[\Delta\dot{H}_t]]\right]\; X_t^i = \tag{4.84}$$

$$= \sum_i \mathrm{Tr}\left[Y_t^i\, \Delta\dot{H}_t\right]\; X_t^i\,. \tag{4.85}$$

For what follows, it is also useful to introduce the notation $(\Delta\dot{H}_t)_i := \mathrm{Tr}\left[Y_t^i\, \Delta\dot{H}_t\right]$ and the symmetric matrix $(m_{i,j}^{KMB})_t := \mathrm{Tr}\left[X_t^i \mathbb{J}_{t,L}^{-1}[X_t^j]\right]$. Then, starting from Eq. (4.68) we have that:

$$\langle\Sigma\rangle = -\beta^2 \int_0^\tau \mathrm{d}t\; \mathrm{Tr}\left[\mathbb{J}_{t,L}[\Delta\dot{H}_t]\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]\right] = \tag{4.86}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t\; \sum_{i,j} \mathrm{Tr}\left[X_t^i\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(X_t^j)]\right] \mathrm{Tr}\left[Y_t^i \Delta\dot{H}_t\right] \mathrm{Tr}\left[Y_t^j \Delta\dot{H}_t\right] = \tag{4.87}$$

$$= \beta^2 \int_0^\tau \mathrm{d}t\; \sum_{i,j} \tau_t^j\, (m_{i,j}^{KMB})_t (\Delta\dot{H}_t)_i (\Delta\dot{H}_t)_j = \tag{4.88}$$

$$= \beta^2 \int_0^\tau \mathrm{d}t\; \sum_{i,j} \left(\frac{\tau_t^i + \tau_t^j}{2}\right) (m_{i,j}^{KMB})_t (\Delta\dot{H}_t)_i (\Delta\dot{H}_t)_j\,. \tag{4.89}$$

Notice that if all the thermalisation timescales were equal, one would obtain back an expression equivalent to the one for the quench scenario. Eq. (4.89), on the other hand, tells us that the average dissipation arises from the interplay between the susceptibility of the thermal state to external perturbations (encoded in $m^{KMB}_{i,j}$) and the thermalisation timescale in that direction. For this reason, it is conceptually justified to introduce the matrix $(T_{i,j})_t := \left(\frac{\tau^i_t + \tau^j_t}{2}\right)$, which encodes the information about the different equilibration velocity in each direction. Then, the metric takes the particularly simple form $(g_{i,j})_t = (T_{i,j})_t \circ (m^{KMB}_{i,j})_t$, where $\circ$ denotes the Hadamard product of matrices.

The same functional expression was found for classical stochastic processes [85]. Indeed, it appears to be quite general. For example, in [6] it was shown how one can translate results obtained in the context of linear response of an adiabatically driven unitary dynamics (see, e.g., [86]) in this language. In this context, the main object of study is the linear response of the expectation value $\mathrm{Tr}\left[X^i_t \pi_\beta(H_t)\right]$ to an external perturbation of the form $\sum_i \lambda^i_t X^i$. If the parameters are moved slowly, one can use the expression:

$$\mathrm{Tr}\left[X^i_t \rho_t\right] = \mathrm{Tr}\left[X^i_t \pi_\beta(H_t)\right] + \sum_j \chi^{ad}_t(X^i_t, X^j_t)\, \dot{\lambda}^j_t + \mathcal{O}\left(||\dot{\lambda}_t||^2\right) , \quad (4.90)$$

where χ^{ad}_t is the adiabatic response function given by:

$$\chi^{ad}_t(X^i_t, X^j_t) = -i \int_0^\infty \mathrm{d}\nu \, \left(\nu \,\mathrm{Tr}\left[\pi_\beta(H_t)\,[e^{iH\nu} X^i_t e^{-iH\nu}, X^j_t]\right]\right) . \quad (4.91)$$

Then, it was argued in [86] that this behaviour naturally leads to the definition of a thermodynamic metric, corresponding to the symmetrised adiabatic response function, i.e., $(g_{i,j})_t = \frac{1}{2}\left(\chi^{ad}_t(X^i_t, X^j_t) + \chi^{ad}_t(X^j_t, X^i_t)\right)$. Despite the apparent formal difference with the other metrics presented so far, it is not difficult to show that this expression can be recast in

terms of the KMB-inner product. Indeed, a two line computation gives:

$$\chi_t^{ad}(X_t^i, X_t^j) =$$

$$= -i \sum_{n,m} \int_0^\infty \mathrm{d}\nu \left(\nu \left(e^{i(E_t^{(n)} - E_t^{(m)})\nu}\right) \left(\frac{e^{-\beta E_t^{(n)}} - e^{-\beta E_t^{(m)}}}{\mathcal{Z}_{H_t}} \right) (X_t^i)_{n,m} (X_t^j)_{m,n} \right) = \tag{4.92}$$

$$= i \sum_{n,m} \frac{1}{\mathcal{Z}_{H_t}} \frac{e^{-\beta E_t^{(n)}} - e^{-\beta E_t^{(m)}}}{(E_t^{(n)} - E_t^{(m)})^2} (X_t^i)_{n,m} (X_t^j)_{m,n} = \tag{4.93}$$

$$= -\beta \int_0^\infty \mathrm{d}\nu \int_0^1 \mathrm{d}x \ \mathrm{Tr}\left[\pi_\beta(H_t)^{1-x} e^{iH_t\nu} X_t^i e^{-iH_t\nu} \pi_\beta(H_t)^x X_t^j \right] = \tag{4.94}$$

$$= -\beta \mathrm{Tr}\left[X_t^i \, \mathcal{U}_t^+ (\mathbb{J}_{t,L}[X_t^j]) \right] , \tag{4.95}$$

where in the first line we expressed everything in the eigenbasis of H_t and in the last step we defined the operator:

$$\mathcal{U}_t^+(A) := \int_0^\infty \mathrm{d}\nu \ \mathrm{Tr}_B \left[e^{-iH_t\nu} A \, e^{iH_t\nu} \right] , \tag{4.96}$$

tracing out the bath degrees of freedom. The expression for $\chi_t^{ad}(X_t^i, X_t^j)$ should be compared with Eq. (4.69), with which it has a striking similarity. Indeed, the eigenvalues of $\mathcal{U}_t^+$ encode the thermalisation timescales of the system, in the same way of $\mathcal{L}_t^+$ in the Lindbladian context. Hence, by simply repeating the same steps above, one can define the analogous of the matrix $(T_{i,j})_t$ for unitary adiabatic evolutions, so that even in this case the metric can be expressed as $(g_{i,j})_t = (T_{i,j})_t \circ (m_{i,j}^{KMB})_t$.

Hence, direct inspection of the metric structure tells us that in order to minimise the dissipation, one should vary only the parameters which at the same time thermalise fast (corresponding to $\tau_t^i \ll 1$) and that do not radically change the thermal state. The possibility of understanding the way in which the system dissipates simply by studying the metric will be further explored in the next section.

4.3.2 General principles from spectral analysis: coherences are detrimental

Apart from the possibility of automatically devising optimal protocols, the geometric approach is also useful when it comes to the interpretation

of the way in which entropy is dissipated. For example, in the caption of Fig. 4.2 it was discussed how one can already get some intuition on the shape of the geodesics by noticing that the metric decays with the population, so it will induce trajectories that are slow in the highly populated region, while accelerating exponentially as the population goes to zero.

A good example of a general principle which can be derived in this way is the one about coherences. Suppose that one wants to connect two Hamiltonian H_0 and H_τ such that $[H_0, H_\tau] = 0$. In the following we will show that under mild assumptions on the dynamics, any creation of coherence will be detrimental, i.e., optimal protocols only pass through diagonal Hamiltonians.

Consider, in fact, a Lindbladian that decouples the population dynamics from the coherences. Detailed balance implies this condition, but it is not necessary to enforce it. In particular, any normal Lindbladian satisfies this requirement, as one can verify from the condition [87]:

$$[\mathcal{U}, \mathcal{L}_{\mathcal{D}}] = 0\,. \tag{4.97}$$

Given a Hermitian operator A, there is a canonical decomposition $A = A^D + A^C$, where A^D only contains the diagonal elements of A (with respect to the corresponding thermal state), while A^C is implicitly defined as $A^C = A - A^D$. Then, under the assumption above, one has that $\mathcal{L}(A)^D = \mathcal{L}(A^D)$ and $\mathcal{L}(A)^C = \mathcal{L}(A^C)$. Moreover, it is straightforward to verify that the same happens for the superoperator $\mathbb{J}_{t,L}$. Hence, one can rewrite the entropy production in Eq. (4.69) as:

$$\langle \Sigma \rangle = -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta \dot{H}_t \, \mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t])\right] = \tag{4.98}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[(\Delta \dot{H}_t^D + \Delta \dot{H}_t^C)\, \left(\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t])^D + \mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t])^C\right)\right] = \tag{4.99}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta \dot{H}_t^D \, \mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t^D])\right] + \mathrm{Tr}\left[\Delta \dot{H}_t^C \, \mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t^C])\right] = \tag{4.100}$$

$$= \left\langle \Sigma^D \right\rangle + \left\langle \Sigma^C \right\rangle\,, \tag{4.101}$$

where in Eq. (4.100) we used the fact that the trace of the product of an off-diagonal operator and a diagonal one is zero. It should be noticed that both contributions to the entropy production satisfy $\left\langle \Sigma^{D/C} \right\rangle \geq 0$. This shows that the creation of coherence is always detrimental. In particular,

if both endpoints of the protocol are diagonal in the same basis, then the optimal path will also be purely diagonal, so that $\langle \Sigma^C \rangle \equiv 0$.

This concludes the discussion of the geometric picture arising from the average dissipation. As it was argued, it provides a flexible framework to automatically design minimally dissipating protocols and to get insights on the nature of dissipation. Interestingly, also quasistatic fluctuations give rise to a metric structure. This will be the subject of next section.

4.4 Fluctuations and quantum signatures

Moving on in the study of the statistics of work, the second natural object to be considered is the fluctuations around the average value. These are computed as:

$$\langle \sigma_w^2 \rangle = \langle w^2 \rangle - \langle w \rangle^2 \ . \tag{4.102}$$

Notice that, since the fluctuations are invariant under translation of their average value, the fluctuations in work are proportional to the ones in the entropy production:

$$\langle \sigma_\Sigma^2 \rangle = \langle \Sigma^2 \rangle - \langle \Sigma \rangle^2 = \beta^2 (\langle (w - \Delta F)^2 \rangle - \langle (w - \Delta F) \rangle^2) = \tag{4.103}$$

$$= \beta^2 (\langle w^2 \rangle - \cancel{2 \langle w \rangle \Delta F} + \cancel{\Delta F} - \langle w \rangle^2 + \cancel{2 \langle w \rangle \Delta F} - \cancel{\Delta F}) = \tag{4.104}$$

$$= \beta^2 \langle \sigma_w^2 \rangle \ . \tag{4.105}$$

Then, one can obtain their expression from the CGF as:

$$\beta^2 \langle \sigma_w^2 \rangle = \langle \sigma_\Sigma^2 \rangle = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} K^{\mathrm{diss}}(\lambda) \Big|_{\lambda=0} = \tag{4.106}$$

$$= \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} \log \left(1 + \int_0^\lambda \mathrm{d}x \ \mathrm{Tr} \left[\pi_\beta(H_\tau)^x (\log \pi_\beta(H_\tau) - \log \rho_\tau) \rho_\tau^{1-x} \right] \right) = \tag{4.107}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\mathrm{Tr} \left[\pi_\beta(H_\tau)^\lambda (\log \pi_\beta(H_\tau) - \log \rho_\tau) \rho_\tau^{1-\lambda} \right]}{1 + \int_0^\lambda \mathrm{d}x \ \mathrm{Tr} \left[\pi_\beta(H_\tau)^x (\log \pi_\beta(H_\tau) - \log \rho_\tau) \rho_\tau^{1-x} \right]} \Bigg|_{\lambda=0} = \tag{4.108}$$

$$= \mathrm{Tr} \left[\rho_\tau (\log \pi_\beta(H_\tau) - \log \rho_\tau)^2 \right] - \mathrm{Tr} \left[\rho_\tau (\log \pi_\beta(H_\tau) - \log \rho_\tau) \right]^2 = \tag{4.109}$$

$$= 2 \, H_V(\rho_\tau || \pi_\beta(H_\tau)) - S(\rho_\tau || \pi_\beta(H_\tau))^2 \, , \tag{4.110}$$

where we used the same integral expression from Eq. (4.47) and used the notation $H_V(\rho||\sigma)$ for the quantum information variance (see Sec. 1.4.7).

This quantity is obviously positive, as it is a contrast function. Moreover, the difference in Eq. (4.110) is also positive. We define this quantity relative entropy variance, which we denote by $V(\rho_\tau||\pi_\beta(H_\tau))$. On the one hand, the positivity of $V(\rho_\tau||\pi_\beta(H_\tau))$ directly follows from the positivity of $\langle\sigma_\Sigma^2\rangle$. At the same time, this can also be proven through Cauchy-Schwartz inequality, since for any ρ and σ one has:

$$S(\rho||\sigma)^2 = \mathrm{Tr}\left[\rho\,(\log\rho-\log\sigma)\mathbb{1}\right]^2 \leq \tag{4.111}$$

$$\leq \mathrm{Tr}\left[\rho\,(\log\rho-\log\sigma)^2\right]\mathrm{Tr}\left[\rho\,\mathbb{1}^2\right] = 2\,H_V(\rho||\sigma)\,, \tag{4.112}$$

which directly implies that $V(\rho||\sigma)\geq 0$. This gives us the extra information about what are the situations that saturate the bound, i.e., for which $V(\rho||\sigma)\equiv 0$. This happens when there exists a scalar λ such that $(\log\rho-\log\sigma)=\lambda\mathbb{1}$. It is straightforward to prove that this is possible only for $\lambda=0$ and $\rho=\sigma$. Hence, the relative entropy variance is positive definite and zero if and only if $\rho\equiv\sigma$.

It should be noticed that the relative entropy variance is also jointly convex, so one can repeat the same arguments that led to Eq. (1.37) to prove that it is also monotone. Hence, if the dynamics is Markovian fluctuations are monotonically increasing during the evolution. Indeed, differentiating directly Eq. (4.18) (notice that the second term in the integral was proven to be zero) one has:

$$\langle\sigma_\Sigma^2\rangle = \int_0^\tau \mathrm{d}t\ \frac{\mathrm{d}}{\mathrm{d}s}V(\rho_t||\pi_\beta(H_s))\bigg|_{s=t} = \tag{4.113}$$

$$= \int_0^\tau \mathrm{d}t\ \left(\frac{\mathrm{d}}{\mathrm{d}t}V(\rho_t||\pi_\beta(H_t)) - \frac{\mathrm{d}}{\mathrm{d}s}V(\rho_s||\pi_\beta(H_t))\bigg|_{s=t}\right) = \tag{4.114}$$

$$= V(\rho_\tau||\pi_\beta(H_\tau)) + \int_0^\tau \mathrm{d}t\ \left(-\frac{\mathrm{d}}{\mathrm{d}s}V(\rho_s||\pi_\beta(H_t))\bigg|_{s=t}\right), \tag{4.115}$$

where again it should be noticed that the meaning of ρ_τ in the last expression is not the same as in Eq. (4.110), as in this case the state only contains information that are local in time, whereas in Eq. (4.110) ρ_τ encodes the whole dynamics.

We denote the integrand in Eq. (4.115) by $\langle(\dot\sigma_\Sigma^2)_t\rangle$. Thanks to the monotonicity of the relative entropy variance, $\langle(\dot\sigma_\Sigma^2)_t\rangle\geq 0$, with equality if and only if $\rho_t\equiv\pi_\beta(H_t)$ for all times. This condition is satisfied only in the limit in which the driving is realised in infinite time, so we can carry out an expansion around $\dot\lambda_t^i\approx 0$ completely analogous to the one

in Eq. (4.55):

$$\langle(\dot{\sigma}^2_\Sigma)_t\rangle = \dot{\lambda}^i_t \cancel{\partial_i \langle(\dot{\sigma}^2_\Sigma)_t\rangle}\big|_{\{\dot{\lambda}_t\}\equiv 0} + \dot{\lambda}^i_t \left(\partial_i\partial_j \langle(\dot{\sigma}^2_\Sigma)_t\rangle\big|_{\{\dot{\lambda}_t\}\equiv 0}\right) \dot{\lambda}^j_t + \mathcal{O}\left(||\dot{\lambda}||^3\right), \tag{4.116}$$

where we sum over the repeated indices. Since we are expanding around a minimum, the first derivative cancels, while the Hessian matrix $(g^\sigma_{i,j})_t := \partial_i\partial_j \langle(\dot{\sigma}^2_\Sigma)_t\rangle\big|_{\{\dot{\lambda}_t\}\equiv 0}$ is positive definite.

Hence, the local behaviour of the fluctuations in the entropy production can also be described geometrically. Interestingly, whereas the metric arising from the average dissipation was well known at least from the '80s, the existence of a metric associated to the fluctuations was discovered only in [8]. The reason for this delay can be best understood directly from the CGF. Indeed, if one can assume time translation covariance of the Lindbladian (i.e., that the evolution of the population completely decouples from the one of the coherences), Eq. (4.40) can be rewritten as:

$$K^{\rm diss}(\lambda) = K^{\rm diss}_D(\lambda) + K^{\rm diss}_C(\lambda)\,, \tag{4.117}$$

where we isolated the diagonal and coherent contribution. The first term is given by:

$$K^{\rm diss}_D(\lambda) = \beta^2 \int_0^\tau {\rm d}t \int_0^\lambda {\rm d}x \int_x^{1-x} {\rm d}y\, {\rm cov}^y_t(\Delta\dot{H}^D_t,\, \mathbb{J}^{-1}_{t,L}[\mathcal{L}^+_t(\mathbb{J}_{t,L}[\Delta\dot{H}^D_t])]) = \tag{4.118}$$

$$= \beta^2 \int_0^\tau {\rm d}t \int_0^\lambda {\rm d}x \int_x^{1-x} {\rm d}y\, {\rm Tr}\left[\pi_\beta(H_t)\,\Delta\dot{H}^D_t\,\mathcal{L}^+_t(\Delta\dot{H}^D_t)\right] = \tag{4.119}$$

$$= \beta^2\,(\lambda-\lambda^2) \int_0^\tau {\rm d}t\ {\rm Tr}\left[\pi_\beta(H_t)\,\Delta\dot{H}^D_t\,\mathcal{L}^+_t(\Delta\dot{H}^D_t)\right], \tag{4.120}$$

where in Eq. (4.119) we used the fact that everything commutes, and in the last step we explicitly carried out the two integrations, since there is no dependence on y in the integrand. The coherent contribution on the other hand is simply given by:

$$K^{\rm diss}_C(\lambda) = \beta^2 \int_0^\tau {\rm d}t \int_0^\lambda {\rm d}x \int_x^{1-x} {\rm d}y\, {\rm cov}^y_t(\Delta\dot{H}^C_t,\, \mathbb{J}^{-1}_{t,L}[\mathcal{L}^+_t(\mathbb{J}_{t,L}[\Delta\dot{H}^C_t])])\,. \tag{4.121}$$

Suppose now that the evolution is purely classical, which implies that $K^{\rm diss}(\lambda) \equiv K_D^{\rm diss}(\lambda)$. Then, it is straightforward to verify that:

$$\langle \Sigma \rangle = -\frac{\mathrm{d}}{\mathrm{d}\lambda} K_D^{\rm diss}(\lambda)\Big|_{\lambda=0} = -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\pi_\beta(H_t)\, \Delta\dot{H}_t^D\, \mathcal{L}_t^+(\Delta\dot{H}_t^D)\right] ; \tag{4.122}$$

$$\left\langle \sigma_\Sigma^2 \right\rangle = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} K_D^{\rm diss}(\lambda)\Big|_{\lambda=0} = -2\,\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\pi_\beta(H_t)\, \Delta\dot{H}_t^D\, \mathcal{L}_t^+(\Delta\dot{H}_t^D)\right] , \tag{4.123}$$

which implies the following fluctuation-dissipation relation (FDR) for classical systems:

$$\frac{1}{2}\left\langle \sigma_\Sigma^2 \right\rangle = \langle \Sigma \rangle \ . \tag{4.124}$$

Hence, whenever no coherence is created along the protocol the metric for the average dissipations and fluctuations actually coincide, which explains the delay in its formulation for classical systems. In the general case, on the other hand, the metric takes the form:

$$\left\langle \sigma_\Sigma^2 \right\rangle = \beta^2 \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} \int_0^\tau \mathrm{d}t \int_0^\lambda \mathrm{d}x \int_x^{1-x} \mathrm{d}y \ \mathrm{cov}_t^y(\Delta\dot{H}_t,\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]) = \tag{4.125}$$

$$= \beta^2 \frac{\mathrm{d}}{\mathrm{d}\lambda} \int_0^\tau \mathrm{d}t \int_\lambda^{1-\lambda} \mathrm{d}y \ \mathrm{cov}_t^y(\Delta\dot{H}_t,\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]) = \tag{4.126}$$

$$= -2\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta\dot{H}_t\, \mathbb{J}_{t,B}[\mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]]\right] , \tag{4.127}$$

where in the last step we use the notation $\mathbb{J}_{t,B} := \mathbb{J}_B\big|_{\pi_\beta(H_t)}$, corresponding to the Bures metric defined in Eq. (1.128). If the Lindbladian is detailed balance, the equation above further simplifies to:

$$\left\langle \sigma_\Sigma^2 \right\rangle = -2\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta\dot{H}_t\, \mathbb{J}_{t,B}[(\mathcal{L}_t^+)^\dagger(\Delta\dot{H}_t)]\right] = \tag{4.128}$$

$$= -2\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta\dot{H}_t\, (\mathcal{L}_t^+)(\mathbb{J}_{t,B}[\Delta\dot{H}_t])\right] , \tag{4.129}$$

which is the expression originally obtained in [8]. In this case, one can give a particularly simple form to the quantum correction to the FDR in

Eq. (4.124):

$$\mathcal{Q} := \frac{1}{2}\left\langle \sigma_\Sigma^2 \right\rangle - \langle \Sigma \rangle = \tag{4.130}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\Delta\dot{H}_t \, (\mathcal{L}_t^+)((\mathbb{J}_{t,B} - \mathbb{J}_{t,L})[\Delta\dot{H}_t])\right] , \tag{4.131}$$

where we used the expression for the average dissipation from Eq. (4.69). As it was pointed out in Sec. 1.3 (see Eq. (1.75)) the ordering of the Fisher information scalar products $\mathbb{J}_f$ follows the one of the corresponding defining functions, which implies that $(\mathbb{J}_{t,B} - \mathbb{J}_{t,L})$ is a positive superoperator. Moreover, since $\mathcal{L}$ is detailed balance, its eigenoperators are orthonormal with respect to both $\mathbb{J}_{t,B}$ and $\mathbb{J}_{t,L}$, and all its eigenvalues are negative. Putting these two remarks together, we obtain that the quantum correction is non-negative, and zero if and only if $\dot{H}_t$ commutes with H_t at all time, i.e.,

$$\dot{\mathcal{Q}}_t \geq 0 \, , \tag{4.132}$$

where $\dot{\mathcal{Q}}_t$ denotes the integrand in Eq. (4.131). In a previous work (see [8]), it was claimed that the positivity of Eq. (4.132) could be obtained only in the presence of detailed balance dynamics. Here we show that this is an unnecessary constraint. Indeed, since $\mathbb{J}_{t,B}[\Delta\dot{H}_t]$ is traceless, one can give an expansion of the fluctuations akin to the one in Eq. (4.89), which reads:

$$\frac{1}{2}\left\langle \sigma_\Sigma^2 \right\rangle = -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\mathbb{J}_{t,B}[\Delta\dot{H}_t] \, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])]\right] = \tag{4.133}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \ \sum_{i,j} \mathrm{Tr}\left[X_t^i \, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(X_t^j)]\right] \mathrm{Tr}\left[Y_t^i \mathbb{J}_{t,L}^{-1}\mathbb{J}_{t,B}[\Delta\dot{H}_t]\right] \mathrm{Tr}\left[Y_t^j \Delta\dot{H}_t\right] = \tag{4.134}$$

$$= \beta^2 \int_0^\tau \mathrm{d}t \ \sum_{i,j} \tau_t^j \, (m_{i,j}^{KMB})_t (\mathbb{J}_{t,B}[\Delta\dot{H}_t])_i (\Delta\dot{H}_t)_j \, , \tag{4.135}$$

where we used the same notation as in Sec. 4.3.1. Since the eigenvalues of $\mathbb{J}_{t,L}^{-1}\mathbb{J}_{t,B}$ are all larger or equal to one (with equality only on the commutant) we have the lower bound:

$$(\mathbb{J}_{t,B}[\Delta\dot{H}_t])_i \geq (\Delta\dot{H}_t)_i \, . \tag{4.136}$$

Then, comparing the integrand in Eq. (4.135) with the one in Eq. (4.88) we have the proof that $\dot{\mathcal{Q}}_t \geq 0$ in general. Moreover, once more we

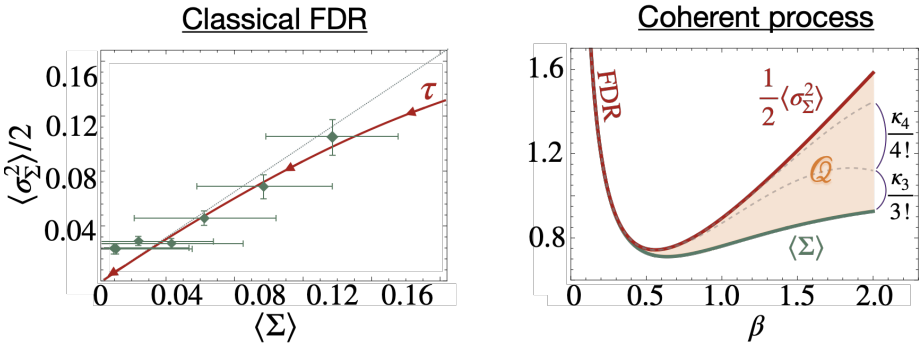


Figure 4.4: On the left, we show how an incoherently driven system approaches the fluctuation-dissipation relation in Eq. (4.124) as the duration of the protocol τ is increased. The Hamiltonian is driven as $H_t = E(t)\,|1\rangle\langle 1|$. The red continuous line are the theoretical predictions, while the green points correspond to experimental data from [5]. On the right, we show the dissipation and fluctuations for a protocol of the form $H_t = t\hat{\sigma}_x + (1-t)\hat{\sigma}_z$. We see that for $\beta \to 0$ the FDR are effectively satisfied, while in the opposite limit the quantum correction $\mathcal{Q}$ grows more and more.

also obtain that equality can only be obtained for commuting protocols. This behaviour is exemplified in Fig. 4.4, where we show the FDR for an incoherently driven system, and its breakdown in the case $[H_t, \dot{H}_t] \neq 0$. Interestingly, from the right panel one can also see that as $\beta \to 0$ (i.e., in the high temperature limit) one goes back to the classical scenario.

Interestingly, the presence of a non-zero $\mathcal{Q}$ is connected to the appearance of higher order cumulants in the distribution of work. In fact, it should be noticed that the CGF can be expressed as:

$$K^{\rm diss}(\lambda) = \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!}\kappa_n\,, \tag{4.137}$$

where $\kappa_1 \equiv \langle\Sigma\rangle$ and $\kappa_2 \equiv \langle\sigma_\Sigma^2\rangle$. Moreover, thanks to Jarzynski equality $K^{\rm diss}(1) = 0$ (see Eq. (3.56)). Hence, setting λ to one in Eq. (4.137), we obtain that:

$$-\langle\Sigma\rangle + \frac{\langle\sigma_\Sigma^2\rangle}{2} + \sum_{n=3}^{\infty}\frac{(-1)^n}{n!}\kappa_n = 0 \implies \mathcal{Q} = -\sum_{n=3}^{\infty}\frac{(-1)^n}{n!}\kappa_n\,. \tag{4.138}$$

This behaviour is exemplified in Fig. 4.4, where we show how $\mathcal{Q}$ can be divided in terms of successive cumulants. Moreover, since higher order

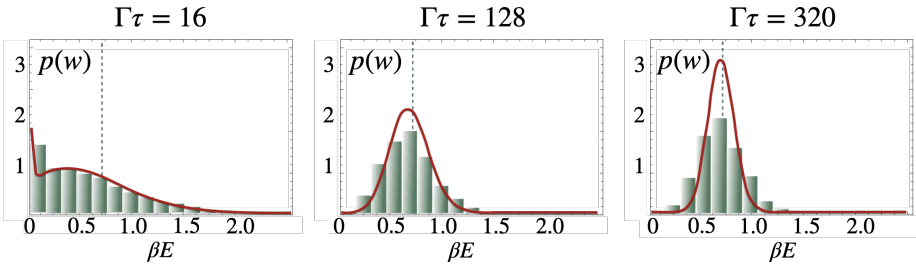


Figure 4.5: Behaviour of the work distribution of an incoherently driven system as the duration of the protocol increases. The histogram represents the experimental data from the same experiment of Fig. 4.4, while the continuous lines are the theoretical prediction [5]. Finally, the dotted line corresponds to the ideal work extraction, i.e., to ΔF.

cumulants are positive in the quasistatic regime (see Sec. 4.5), this also means that the FDR directly implies the fact that all other cumulants are zero. This is the defining property of a Gaussian. Indeed, the observation of FDR in the slow driving regime witnesses the Gaussianity in the distribution of the work, behaviour presented in Fig. 4.5 for an incoherently driven quantum dot. This connection was already noticed by Jarzynski in its original paper [69], where he derived the FDR from arguments justifying the Gaussianity of work distribution of slowly driven classical systems by appealing to the central limit theorem.

His reasoning can indeed be made precise in the quench model (see Fig. 4.6). In this context, increasing the number of steps N makes the work scale as $1/N$. Moreover, the thermalisation between each of the steps completely erases any memory of the past, so that the total work is given by a sum of independent random variables scaling as $1/N$. These are the premises of the central limit theorem, which explains why for diagonal systems the work distribution approaches a Gaussian. This behaviour can also be expected in general, as one can verify from inspection of $K_D^{\text{diss}}(\lambda)$ in Eq. (4.120). Indeed, a quadratic CGF implies that the corresponding probability distribution is Gaussian.

It should be noticed that Eq. (4.138) also tells us that any deviation from a Gaussian behaviour in the slow driving regime is a witness of the creation of coherence. Indeed, the difference in the statistical behaviour can be quite stark. Consider, for example, a system in which the driving does not affect the energy spacing of the spectrum, but only creates coherences between different energy levels (again for clarity of the ex-

planation, we consider the quench model of the dynamics). This kind of protocols are called purely coherent and are depicted in the bottom panel of Fig. 4.6. In this case, increasing the number of steps only affects the transition probability, while the possible work output remains unvaried. Hence, in the limit of $N \gg 1$, the final distribution will be a comb of δ-functions, whereas one would have expected a continuous behaviour.

Wrapping up, in this section we showed how the expansion of the fluctuations of entropy production in the slow driving regime gives rise to a metric that is equivalent to the one for the dissipation in the classical regime, but which becomes different whenever coherence is created along the protocol. Indeed, one can prove fluctuation-dissipation relations that hold for diagonally driven systems, but which breakdown whenever $[H_t, \dot{H}_t] \neq 0$ for any time of the protocol. In turn, the FDR are equivalent to the Gaussianity of the distribution, and any deviation from this behaviour also witnesses the creation of coherence.

It should be noticed that these effects make the phenomenology of quantum thermodynamics systems much richer than their classical counterpart. For one thing, whereas for incoherent protocols one can minimise the dissipation and the fluctuations at the same time, this is no longer the case when one has a non-commutative driving. Indeed, whereas in the classical regime a single metric is sufficient to completely characterise (and optimise) thermodynamic protocols close to equilibrium, whenever coherence is present one has a whole family of different metrics, corresponding to different Fisher informations. Starting from Eq. (2.32) we can in fact give the following quasistatic expansion:

$$\left\langle g(e^{-\Sigma})\right\rangle = \int \mathrm{d}w\; p(w)\, g(e^{-\beta(w-\Delta F)}) = \tag{4.139}$$

$$= H_g(U_\tau \pi_\beta(H_0) U_\tau^\dagger \,||\, \pi_\beta(H_\tau)) = \tag{4.140}$$

$$= \int_0^\tau \mathrm{d}t\;\; \frac{\mathrm{d}}{\mathrm{d}s}\, H_g(\rho_t||\pi_\beta(H_s))\Big|_{s=t} = \tag{4.141}$$

$$= \int_0^\tau \mathrm{d}t\;\; \frac{\mathrm{d}}{\mathrm{d}s}\, H_g(\pi_\beta(H_t) + \delta\rho_t||\pi_\beta(H_s))\Big|_{s=t} = \tag{4.142}$$

$$= \beta \int_0^\tau \mathrm{d}t\;\; \mathrm{Tr}\left[\delta\rho_t\, \mathbb{J}_{t,f}^{-1}[\mathbb{J}_{t,L}[\Delta\dot{H}_t]]\right], \tag{4.143}$$

where we have reproduced the standard procedure presented above to obtain a metric for $\rho_t \approx \pi_\beta(H_t) + \delta\rho_t$ and we have used the shorthand notation $\mathbb{J}_{t,f} := \mathbb{J}_f\big|_{\pi_\beta(H_t)}$. Moreover, for Lindbladian evolutions we can

further simplify Eq. (4.143) to:

$$\left\langle g(e^{-\Sigma})\right\rangle = -\beta^2 \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t])\,\mathbb{J}_{t,f}^{-1}[\mathbb{J}_{t,L}[\Delta\dot{H}_t]]\right] . \quad (4.144)$$

Interestingly, all the functionals above collapse into the same value for classically driven systems. On the other hand, since we know that $g(x) = (x-1)^2/(x+1)$ is maximal, one can define a whole family of positive quantum signatures, given by:

$$\mathcal{Q}_g := \left\langle \frac{(1-e^{-\Sigma})^2}{1+e^{-\Sigma}} \right\rangle - \left\langle g(e^{-\Sigma})\right\rangle = \quad (4.145)$$

$$= \left\langle 2\, e^{-\frac{\Sigma}{2}} \sinh\left(\frac{\Sigma}{2}\right) \tanh\left(\frac{\Sigma}{2}\right)\right\rangle - \left\langle g(e^{-\Sigma})\right\rangle , \quad (4.146)$$

which witnesses the creation of coherences during the protocol.

4.5 Higher order cumulants

After discussing the first two cumulants, we can also briefly analyse how higher terms behave. First, it should be noticed that since we have to take at least three derivatives, there is no contribution coming from $K_D^{\mathrm{diss}}(\lambda)$, as the latter has only a quadratic dependence on λ. As it was discussed above, this already shows that higher order cumulants can be non-zero if and only if the driving is non-commuting for some time. Hence, one can restrict the attention to the differentiation of:

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} K_C^{\mathrm{diss}}(\lambda) = -2\beta^2 \,\mathrm{Re} \int_0^\tau \mathrm{d}t \ \mathrm{cov}_t^\lambda(\Delta\dot{H}_t^C,\, \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t^C])]) = \quad (4.147)$$

$$= -2\beta^2 \,\mathrm{Re} \int_0^\tau \mathrm{d}t \ \mathrm{Tr}\left[\pi_t^\lambda \Delta\dot{H}_t^C\, \pi_t^{1-\lambda} \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta\dot{H}_t^C])]\right] , \quad (4.148)$$

where in order to keep the notation compact we used the abbreviation $\pi_t := \pi_\beta(H_t)$. It is useful to investigate how the integrand alone behaves under differentiation. To this end, consider two generic operators A and B, and notice that the following two relations hold:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathrm{Tr}\left[\pi_t^\lambda A\, \pi_t^{1-\lambda} B\right] = -\beta\,\mathrm{Tr}\left[\pi_t^\lambda [H,A]\, \pi_t^{1-\lambda} B\right] ; \quad (4.149)$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathrm{Tr}\left[\pi_t^\lambda A\, \pi_t^{1-\lambda} B\right] = -\beta^2\,\mathrm{Tr}\left[\pi_t^\lambda [H,A]\, \pi_t^{1-\lambda} [H,B]\right] , \quad (4.150)$$

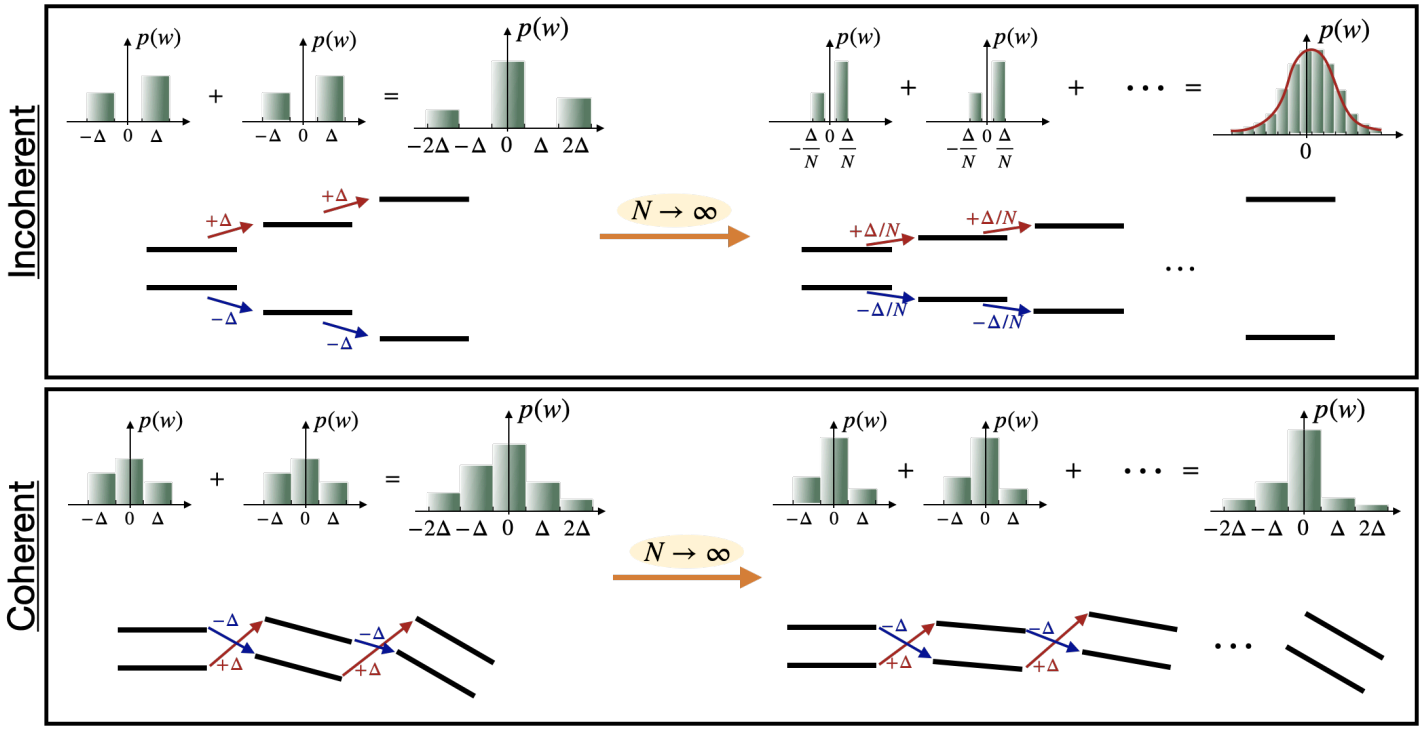


Figure 4.6: For diagonally driven systems, increasing the number of steps in which a protocol is realised makes the work output at each step scale as $1/N$. This allows for the application of the central limit theorem. In the opposite scenario, when purely coherent protocols are considered, increasing the number of steps only affects the probability of having a transition, but not the energy spacing. This leads to a discrete distribution for all N.

as each differentiation gives a term of the form $\pm \log \pi_t$, and it can be easily verified that the terms coming from the partition function do not contribute to the final result. Hence, we can rewrite all the cumulants in terms of the two inductively defined operators $C_t^{(n)}$ and $D_t^{(n)}$, where:

$$C_t^{(1)} = \Delta \dot{H}_t^C \,; \qquad C_t^{(n)} = [H,\, C_t^{(n-1)}] \,; \tag{4.151}$$

$$D_t^{(1)} = \mathbb{J}_{t,L}^{-1}[\mathcal{L}_t^+(\mathbb{J}_{t,L}[\Delta \dot{H}_t^C]) \,; \qquad D_t^{(n)} = [H,\, D_t^{(n-1)}] \,, \tag{4.152}$$

and each of the cumulants of order three or higher is given by:

$$\kappa_{2n+1} = -2\beta^{2n+1} \operatorname{Re} \int_0^\tau \mathrm{d}t \operatorname{Tr} \left[\pi_t \, D_t^{(n)} C_t^{(n+1)} \right] \,; \tag{4.153}$$

$$\kappa_{2n+2} = -2\beta^{2n+2} \operatorname{Re} \int_0^\tau \mathrm{d}t \operatorname{Tr} \left[\pi_t \, D_t^{(n+1)} C_t^{(n+1)} \right] \,. \tag{4.154}$$

In order to prove the positivity of all κ_n it is useful to expand Eq. (4.148) in terms of the eigenoperators of the Lindbladian, in the same way as we did in the previous sections. Then, one obtains:

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} K_C^{\mathrm{diss}}(\lambda) = \beta^2 \int_0^\tau \mathrm{d}t \, \sum_{i,j} \tau_t^j \, (m_{i,j}^{KMB})_t (\mathbb{J}_{t,L}^{-1}[\Delta \dot{H}_t^C]_\lambda)_i (\Delta \dot{H}_t^C)_j \,, \tag{4.155}$$

where we denoted by $(\mathbb{J}_{t,L}^{-1}[\Delta \dot{H}_t^C]_\lambda)_i$ the following two traces:

$$\operatorname{Tr} \left[Y_t^i \, \mathbb{J}_{t,L}^{-1}[\pi^\lambda \Delta \dot{H}_t^C \pi^{1-\lambda}] \right] + \operatorname{Tr} \left[Y_t^i \, \mathbb{J}_{t,L}^{-1}[\pi^{1-\lambda} \Delta \dot{H}_t^C \pi^\lambda] \right] \,. \tag{4.156}$$

This decomposition is particularly useful because it reduces the whole dependency on λ to the expression above. We can now show that for each n there exists a positive constant c_n such that:

$$(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}\lambda^n} (\mathbb{J}_{t,L}^{-1}[\Delta \dot{H}_t^C]_\lambda)_i \Big|_{\lambda=0} \geq c_n \, (\Delta \dot{H}_t^C)_i \,. \tag{4.157}$$

Indeed, focusing on the λ dependent part, we can give the following coordinate expression:

$$\pi^\lambda \, |i\rangle\langle j| \, \pi^{1-\lambda} + \pi^{1-\lambda} \, |i\rangle\langle j| \, \pi^\lambda = \tag{4.158}$$

$$= \left((\pi_t)_i^\lambda (\pi_t)_j^{1-\lambda} + (\pi_t)_i^\lambda (\pi_t)_j^{1-\lambda} \right) \, |i\rangle\langle j| \tag{4.159}$$

where $\{|i\rangle\}$ is the eigenbasis of π_t, and we assume $i \neq j$. It should be noticed that this eigeonperator is diagonal in the same basis as $\mathbb{J}_{t,L}$, so their spectrum simply multiplies. Let us now restrict to the study of the behaviour of this expression under derivation. To this end, it is useful to distinguish between even and odd derivatives, which give:

$$\begin{aligned}
&\frac{\mathrm{d}^{2n}}{\mathrm{d}\lambda^{2n}} \left(\pi^{\lambda} \ |i\rangle\langle j| \ \pi^{1-\lambda} + \pi^{1-\lambda} \ |i\rangle\langle j| \ \pi^{\lambda}\right)\Big|_{\lambda=0} = \\
&\qquad = ((\pi_t)_i + (\pi_t)_j) \left(\log \frac{(\pi_t)_i}{(\pi_t)_j}\right)^{2n} |i\rangle\langle j| \ ; \qquad (4.160) \\
&-\frac{\mathrm{d}^{2n+1}}{\mathrm{d}\lambda^{2n+1}} \left(\pi^{\lambda} \ |i\rangle\langle j| \ \pi^{1-\lambda} + \pi^{1-\lambda} \ |i\rangle\langle j| \ \pi^{\lambda}\right)\Big|_{\lambda=0} = \\
&\qquad = ((\pi_t)_i - (\pi_t)_j) \left(\log \frac{(\pi_t)_i}{(\pi_t)_j}\right)^{2n+1} |i\rangle\langle j| \ . \qquad (4.161)
\end{aligned}$$

It is not hard to see that in both cases the spectrum is positive. Indeed, for even terms, both the sum in the first parenthesis and the logarithm part are positive, so the result trivially follows. On the other hand, for odd terms, one should notice that both components have the same sign. Indeed, if $(\pi_t)_i > (\pi_t)_j$, then clearly $(\pi_t)_i - (\pi_t)_j > 0$ and $\log \frac{(\pi_t)_i}{(\pi_t)_j} > 0$, so the result is positive. In the opposite case, namely $(\pi_t)_i < (\pi_t)_j$, both inequalities flip, so their product is again positive.

We are now ready to prove the claim. First notice that both $\mathbb{J}_{t,L}$ and the derivatives of the operator in Eq. (4.158) have positive spectrum, which is strictly non-zero on the subspace of non-commuting operators. Moreover, the differentiation does not change the eigenbasis presented in Eq. (4.158), so the total spectrum is just the product of the two. Hence, by setting c_n to the smallest term in this product, Eq. (4.157) is always satisfied. But then, the positivity of higher cumulants follows from the same proof we used for the fluctuations, see Eq. (4.136). This concludes the proof of the positivity of all the cumulants of the dissipation, i.e.,

$$\forall n \qquad \kappa_n > 0 \, . \qquad (4.162)$$

It should be noticed though, that differently from what happened for the average dissipation and the fluctuations (and for the functionals in Eq. (4.139)) higher order cumulants do not subsume a metric structure, as they are not positive definite. Still, one can use Eq. (4.162) to infer some information about the shape of the distribution: indeed, the skewness

γ_1 and the excess kurtosis γ_2 are connected to the cumulants by the relations:

$$\gamma_1 = \frac{\kappa_3}{(\kappa_2)^{3/2}}, \tag{4.163}$$

$$\gamma_2 = \frac{\kappa_4}{(\kappa_2)^2}. \tag{4.164}$$

The positivity of κ_3 and κ_4 then means that the probability distribution has a fat tail on the right of the average $\langle\Sigma\rangle$, i.e., that compared to a normal distribution, values of the dissipation which are larger than the average by five or more standard deviations are more likely to occur due to quantum fluctuations.

Finally, it is interesting to point out the dependence of the higher cumulants on temperature. As it apparent from Eq. (4.147), this is directly connected to the behaviour of the y-covariance: indeed, if the variation of the parameter y does not change the corresponding covariance, then any differentiation of Eq. (4.147) will result in cumulants that are identically zero. This is exactly what happens in the high-temperature regime, as it is exemplified in Fig. 4.7 for a qubit. In there, we plot the y-covariance both for $[H_t, \dot{H}_t] = 0$ (incoherent driving) and $[H_t, \dot{H}_t] \neq 0$ (coherent driving). Whereas the two become more and more different as β increases, for $\beta \to 0$ one can see that even for the coherent protocol one finds the same flat behaviour obtained for the classical driving, which in turns witnesses the cancellation of all higher order cumulants. In this way, we can see how statistical features of the distribution of dissipation can be deduced from how the y-covariance depends on the parameter y.

4.6 Statistical properties close to equilibrium

Entropy production out of equilibrium can become incredibly complicated to characterise and is, in general, quite system dependent. Apart from the remarkable results given by the fluctuation theorems (see Sec. 3.2) little is known, and one would lean to expect that few universal results can be found.

The intermediate situation presented so far, the one of slow protocols, allows to make quite general statements about the statistics of the dissipation. One could prove the existence of FDR for classical systems, whose breakdown witnesses the creation of coherences along the driving. It was further shown that the distribution for an incoherently driven system is always a Gaussian for slow enough protocols, and that, thanks to

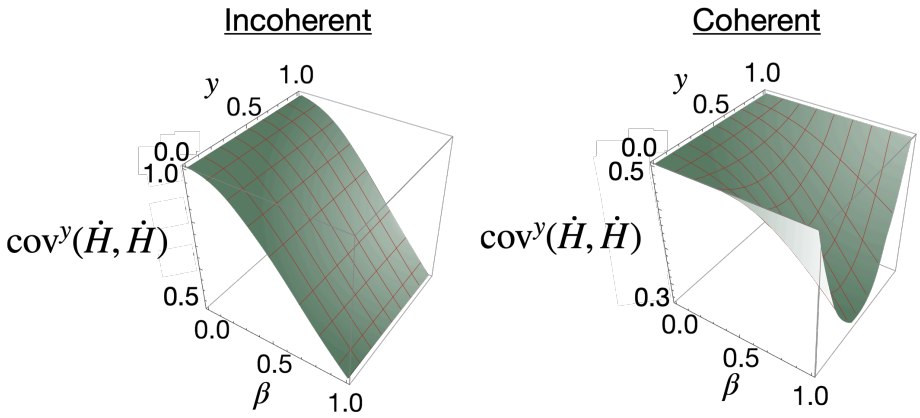


Figure 4.7: Depiction of the y-covariance for a qubit in the thermal state corresponding to $H = \sigma_z$ for the two drivings and $\dot{H} = \sigma_z$, for the left panel (incoherent driving), and $\dot{H} = \sigma_x$ for the right one (coherent driving). In the first case, we can see that there is no dependency on the parameter y, so any differentiation of the CGF higher than the second will give zero. On the other hand, for the coherent driving at finite temperature, one can see that the corresponding surface becomes quite curved. In the high temperature limit though, i.e., as $\beta \to 0$, this feature washes out.

the FDR, it is sufficient to know the average value of the dissipation in order to completely characterise the corresponding probability distribution. Again, the breakdown of this behaviour are neatly related to the creation of coherence. In this case, one can further prove that all the cumulants are non-negative, and in particular positive for non-commuting protocols. Translating this last result in terms of skewness and excess kurtosis tells us that the probability distribution has a fat tail on the right of the average value.

The possibility of giving such a complete characterisation of the entropy production is connected to two mechanisms that highly simplify the physics of the system. Indeed, we will show that the entropy production becomes completely time symmetric close to equilibrium, which allows for a stronger version of the Crooks relations to be derived, the so called Evans-Searles theorem. This result highly constrains the corresponding distribution. Before that, we will also show that, differently from the general case, the dissipation of thermodynamic resources connected to athermality and asymmetry decouples close to equilibrium. This allows to represent the probability distribution of slowly driven protocols as the sum of two independent random variables, each associated to a different type of dissipation. These two phenomena are peculiar to the slow driving regime and allow for a much simpler treatment of the entropy production.

4.6.1 Separation of different channels of entropy production

At the inception of thermodynamics there is the effort to characterise the capability of an agent to perform useful tasks under the constraint that some of the energy which can be accessed is not completely usable. The same framework appears in modern resource theory [88], making thermodynamics one of the most natural examples for this abstract approach. Indeed, among the firsts to set thermodynamics in this language we can already find the work by Lieb and Yngvason in 1999 [89], where the uniqueness of the entropy functional was proved on isothermal slices.

Out of the equilibrium, it was shown that a single law is not sufficient to characterise all thermodynamic processes, not even when all the states are diagonal. Indeed, a necessary condition ensuring the existence of a thermodynamic transformation from a state ρ to a new state σ is that

the following condition is satisfied for all Rényi divergences [90]:

$$S_\lambda(\varrho||\pi_\beta(H)) \geq S_\lambda(\sigma||\pi_\beta(H)) \quad (\forall\lambda \geq 0)\,, \tag{4.165}$$

where $\pi_\beta(H)$ is the reference thermal state, and the Hamiltonian is fixed in time. Since the Rényi divergences are positive definite, we can interpret each of them as a different quantifier of a resource referred to as *athermality*, which degrades over time during thermodynamic processes. Indeed, we know that as time passes every thermodynamic transformation brings any state to thermal equilibrium, for which the divergences in Eq. (4.165) go to zero, signalling a complete depletion of all the resources. In this context, the thermal state is the most passive of the theory, as no transition $\pi_\beta(H) \to \rho$ is allowed thermodynamically.

If the state contains off-diagonal terms in the energy eigenbasis an additional family of constraints have to be satisfied by all thermal transitions $\rho \to \sigma$ [91, 92]:

$$S_\lambda(\varrho||\mathcal{D}_H(\varrho)) \geq S_\lambda(\sigma||\mathcal{D}_H(\sigma)) \quad (\forall\lambda \geq 0)\,, \tag{4.166}$$

where $\mathcal{D}_H$ is the dephasing operator in the H eigenbasis. This set of constraints quantifies the amount of coherence in each state and the corresponding resource is called *asymmetry*, and it is connected with the breakdown of the time translation invariance of the state.

Interestingly, the expressions in Eq. (4.165) and Eq. (4.166) closely resemble the ones obtained in Eq. (4.16) for the dissipative part of the CGF and with Eq. (4.48) for its average version, despite the different frameworks in which the two were derived (for one thing, in the latter case the Hamiltonian varies in time). Despite this similarity, though, it is in general impossible to divide the effects coming from the dissipation of athermality from the one of asymmetry in the statistics of work. A notable exception in this context is given by average quantities, that, due to their connection to the relative entropy, are additive in the two contributions [91].

When one restricts the attention to close-to-equilibrium protocols the same kind of simplification appears at all the level of statistics. In order to exemplify this effect, we focus here on the quench model, but it should be noticed that the same conclusions also hold for any normal Lindbladian, since the condition in Eq. (4.97) makes the dynamics of the diagonal terms completely independent from the coherence part. In this context, we can intuitively understand how the creation and depletion of resources take place during each quench: first, since the system starts in a thermal

state there is no athermality in it, nor asymmetry, as it is diagonal. All the resources are introduced by quenching the Hamiltonian, which by providing work brings the state out of equilibrium and, at the same time, can in principle break its time symmetry invariance by introducing off-diagonal terms. Hence, part of the work is converted in athermality, part in asymmetry. Right after the quench a perfectly thermalising operation is applied, which dissipates both resources, bringing the system back to a symmetric equilibrium state.

If the system were always at equilibrium, from the result of Lieb and Yngvason one would expect the existence of a single resource quantifier to certify possible transitions. Indeed, in the case of isothermal transformations a transition $\rho \to \sigma$ is possible only if $w = \Delta F_{\rho\to\sigma}$ [89]. When moving out of equilibrium though, we know that one also need to take into account the effects of the stochastic variable Σ, the entropy production, which justifies the need for the many constraints which have to be satisfied in Eq. (4.165) and Eq. (4.166). Interestingly, though, in the quasistatic regime the cumulant generating function for the entropy production decouples as (see Eq. (4.117))

$$K^{\mathrm{diss}}(\lambda) = K_D^{\mathrm{diss}}(\lambda) + K_C^{\mathrm{diss}}(\lambda)\,, \tag{4.167}$$

signalling the existence of two independent stochastic variables Σ_{D} and Σ_{C}, corresponding respectively to the dissipation of athermality and to the one of asymmetry, as the first comes from the diagonal part of the driving while the second follows from the off-diagonal contributions. This intuitive separation can be made precise by presenting two independent protocols each reproducing one part of $K^{\mathrm{diss}}(\lambda)$. To this end, it is useful to rewrite explicitly the two contributions in the quench model, which are given by:

$$K_D^{\mathrm{diss}}(\lambda) = \frac{\beta^2(\lambda^2-\lambda)}{2N^2}\int_0^N \mathrm{d}t\ \mathrm{Var}_t(\Delta\dot{H}_t^D)\,, \tag{4.168}$$

$$K_C^{\mathrm{diss}}(\lambda) = -\frac{\beta^2}{2N^2}\int_0^N \mathrm{d}t\int_0^\lambda \mathrm{d}x\int_x^{1-x}\mathrm{d}y\ \mathrm{cov}_t^y(\Delta\dot{H}_t^C,\,\Delta\dot{H}_t^C)\,, \tag{4.169}$$

where $\Delta\dot{H}_t^{D/C}$ are the same as the one used in Eq. (4.117), and we introduced the variance $\mathrm{Var}_t(A) := \mathrm{Tr}\left[A^2, \pi_\beta(H_t)\right] - \mathrm{Tr}\left[A\,\pi_\beta(H_t)\right]^2$.

Consider now a new quench protocol for which each step is divided in two parts: (a) a change of the Hamiltonian which affects only the diagonal $H_i \to (H_i + \mathcal{D}_i(H_{i+1} - H_i))$, followed by complete thermalisation (where

$\{H_i\}$ are the one of the driving, and $\mathcal{D}_i := \mathcal{D}_{H_i}$); (b) a rotation of the energy basis $(H_i + \mathcal{D}_i(H_{i+1} - H_i)) \to H_{i+1}$, again followed by complete thermalisation. The corresponding cumulant generating function takes the form:

$$K^{\text{diss}}_{\text{sim}}(\lambda) = K^{\text{diss}}_A(\lambda) + K^{\text{diss}}_B(\lambda)\,, \tag{4.170}$$

where $K^{\text{diss}}_{A/B}(\lambda)$ correspond to the two parts of the protocol, and their independence is a consequence of the thermalisation step that completely decouples one process from the other. Writing down the explicit form of the two terms in Eq. (4.170):

$$K^{\text{diss}}_A(\lambda) = (\lambda - 1)\sum_{i=1}^{N} S_\lambda(\pi_\beta(H_i + \mathcal{D}_i(H_{i+1} - H_i))||\pi_\beta(H_i))\,, \tag{4.171}$$

$$K^{\text{diss}}_B(\lambda) = (\lambda - 1)\sum_{i=1}^{N} S_\lambda(\pi_\beta(H_{i+1})||\pi_\beta(H_i + \mathcal{D}_i(H_{i+1} - H_i)))\,, \tag{4.172}$$

we see that in general the simulated process differs from the original one, as $K^{\text{diss}}(\lambda) \neq K^{\text{diss}}_{\text{sim}}(\lambda)$. In the quasistatic limit though things highly simplify. First, it should be noticed that in the limit $N \to \infty$, standard arguments from perturbation theory show that H_{i+1} has the same spectrum as $(H_i + \mathcal{D}_i(H_{i+1} - H_i))$. If β is small enough (or N large enough), the same also holds for the corresponding thermal states. Hence, expanding the first term one obtains:

$$K^{\text{diss}}_A(\lambda) = \frac{\beta^2(\lambda^2 - \lambda)}{2N^2}\int_0^N \mathrm{d}t\ \mathrm{Var}_t(\Delta\mathcal{D}_t(\dot{H}_t)) = K^{\text{diss}}_D(\lambda)\,, \tag{4.173}$$

since $\mathcal{D}_t(\dot{H}_t) = \dot{H}^D_t$.

If we now focus on the second term, $K^{\text{diss}}_B(\lambda)$, it should be noticed that the variation in the Hamiltonian is given by:

$$H_{i+1} - (H_i + \mathcal{D}_i(H_{i+1} - H_i)) = H_{i+1} - \mathcal{D}_i(H_{i+1}) = H^C_{i+1}\,. \tag{4.174}$$

Hence, in the same limit of small β, or large N one can express the second part of the CGF as:

$$K^{\text{diss}}_B(\lambda) = -\frac{\beta^2}{2N^2}\int_0^N \mathrm{d}t\int_0^\lambda \mathrm{d}x\int_x^{1-x} \mathrm{d}y\ \mathrm{cov}^y_t(\Delta\dot{H}^C_t, \Delta\dot{H}^C_t) = K^{\text{diss}}_C(\lambda)\,. \tag{4.175}$$

The computations just completed show that one can indeed simulate the action of quasistatic protocols by splitting it in two independent processes, one in which only athermality is consumed, and the other in which only asymmetry is depleted, as represented in Fig. 4.8. Indeed, by looking at Eq. (4.171) and Eq. (4.172) one can notice that the first CGF mirrors the form of the constraints on athermality imposed by the second laws in Eq. (4.165) (as all the states are diagonal), while the latter reproduces the structure of Eq. (4.166), in which one compares a state with its own dephased version. It should be noticed, though, as it was correctly pointed out in [93], that due to the dependence of the thermal state on the temperature, in order to ensure the correctness of the calculations above, one needs to take N exponentially large in β, limiting the practical use of the split presented. Still, Eq. (4.173) and Eq. (4.174) should be understood as a proof of principle of the separation of the entropy production close to equilibrium into two independent channels, one for each kind of resource.

In this context, it is interesting to notice that the family of second laws in Eq. (4.165) collapse into a single constraint, since $\mathrm{cov}_t^y(\Delta\dot{H}_t^D,\, \Delta\dot{H}_t^D) = \mathrm{Var}_t(\Delta\mathcal{D}_t(\dot{H}_t))$. Indeed, close to equilibrium, in the absence of coherences, one can completely characterise the entropy production Σ^D in terms of its average value alone, so it is clear that one can look just at one quantity (other than ΔF) to assess whether a transition is thermodynamically possible or not.

Another interesting aspect of the splitting in Eq. (4.167) is that one can rewrite the coherent contribution as:

$$K_C^{\mathrm{diss}}(\lambda) = \frac{\beta^2(\lambda^2-\lambda)}{2N^2}\int_0^N \mathrm{d}t\ \mathrm{Var}_t(\Delta\dot{H}_t^C) + \tag{4.176}$$

$$+\frac{\beta^2}{2N^2}\int_0^N \mathrm{d}t\int_0^\lambda \mathrm{d}x\int_x^{1-x}\mathrm{d}y\ I_t^y(\pi_\beta(H_t),\,\Delta\dot{H}_t^C)\,, \tag{4.177}$$

where in the second line we used the definition of the Wigner-Yanase-Dyson skew information (see Eq. (1.182)). This quantifies the amount of uncertainty about the coherent driving as measured in the thermal state, and it is widely used in the theory of asymmetry [94]. The expression just presented does not depend on the splitting used in Eq. (4.170), showing a general connection between the coherent part of the driving and resource theoretic quantities. It should also be noticed that the same splitting is also possible for the Lindbladian model, by decomposing $\Delta\dot{H}_t^C$ in terms of the corresponding right eigenoperators.

The splitting into diagonal and coherent contribution is present for

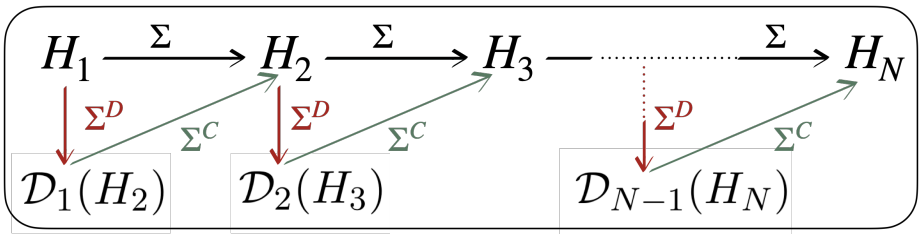


Figure 4.8: In the quasistatic regime the entropy production Σ splits in two additive contributions, one which accounts only for the dissipation associated with the athermality created at each step (Σ^D, red line), and a part coming solely from a change in the energy basis (Σ^C, green line).

all cumulants, i.e., $\kappa_n = \kappa_n^D + \kappa_n^C$. In particular, since for $n > 2$, $\kappa_n^D = 0$, and the first two cumulants are positive, we have that:

$$\kappa_n \geq \kappa_n^D \,, \tag{4.178}$$

confirming the intuition that the additional channel of entropy production provided by the dissipation of asymmetry worsen the quality of the work extraction. Hence, close to equilibrium, coherences are not only detrimental to the average work extraction but at all levels of statistics. While, as it was pointed out in the discussion above, the result for the average is not new, as it derives from the additivity of the relative entropy, the one for higher cumulants follows exclusively from the independence in the dissipation of athermality and asymmetry for slow driving protocols.

Additionally, the two terms in Eq. (4.167) independently satisfy the Jarzynski equality, as it can be explicitly verified by evaluating Eq. (4.168) and Eq. (4.169) for $\lambda = 1$. This means that both $K_D^{\text{diss}}(\lambda)$ and $K_C^{\text{diss}}(\lambda)$ can indeed be considered as arising from two independent thermal processes. The probability distribution of the total dissipated work will then be given by the convolution between the Gaussian coming from the dissipation of athermality resources with the probability distribution coming from the degradation of asymmetry. This observation also implies that the two paradigmatic regimes studied in section 4.4 (see Fig. 4.6) can be considered to be the cornerstone of any quasistatic thermodynamic process.

Finally, one last interesting effect is the fact that the modified FDR are also satisfied independently by the two different channels of entropy

production. We can then write the inequality:

$$\frac{\langle \Sigma^D \rangle}{\sigma^2_{\Sigma^D}} = \frac{1}{2} \geq \frac{\langle \Sigma^C \rangle}{\sigma^2_{\Sigma^C}} = \frac{1}{2} - \frac{\mathcal{Q}^C}{\sigma^2_{\Sigma^C}} \,. \tag{4.179}$$

This shows that for a fixed amount of fluctuations, coherent processes always dissipate less. It should be noticed though that the presence of a negative correction implies that higher cumulants are non-zero, as it was shown in Eq. (4.138), signalling a tendency of the system to fluctuate above $\langle \Sigma \rangle$. Comparing this behaviour with the one happening for diagonal driving, it can be noticed how the entropy production associated with the degradation of asymmetry is inherently different than the one associated to the dissipation of athermality, having bigger fluctuations, arising partly from the thermal disorder, partly from the genuinely quantum uncertainty in the state. Indeed, this behaviour was already noticed in [95].

4.6.2 Time reversal symmetry

As it was pointed out in Sec. 3.2.3, one can derive the Crooks relations for the entropy production:

$$\frac{p^F(\Sigma)}{p^R(-\Sigma)} = e^{\Sigma} \tag{4.180}$$

from the equality $K_R^{\text{diss}}(\lambda) = K_F^{\text{diss}}(1-\lambda)$ (see Eq. (3.63)). Interestingly enough, in the slow driving regime one can prove that $K_F^{\text{diss}}(\lambda) = K_F^{\text{diss}}(1-\lambda)$, which implies that Eq. (4.180) holds without having to revert the protocol. This can be verified from Eq. (4.27), by noticing that the extremes of integration are invariant under the change of variable $\lambda \to 1-\lambda$. A more elegant proof can be given as follows: in the TPM scheme the cumulant generating function associated to the reverse protocol can be written as:

$$K_R^{\text{diss}}(\lambda) = (\lambda - 1)\, S_\lambda(\pi_\beta(H_0) || U_\tau^\dagger \pi_\beta(H_\tau) U_\tau) = \tag{4.181}$$

$$= (\lambda - 1)\, S_\lambda(U_\tau \pi_\beta(H_0) U_\tau^\dagger || \pi_\beta(H_\tau)) = \tag{4.182}$$

$$= (\lambda - 1)\, S_\lambda(\rho_\tau || \pi_\beta(H_\tau)) \,, \tag{4.183}$$

i.e., by inverting the arguments of the Rényi entropy in Eq. (4.16) for the forward driving. Hence, in order to obtain the slow driving expansion for the reverse protocol, one can simply carry out the same expansion

presented in Sec. 4.2 but with the arguments inverted. As it was pointed out in Sec. 1.2.1, though, all contrast functions become symmetric in their arguments for close-by states, so one can deduce that $K_F^{\rm diss}(\lambda) = K_R^{\rm diss}(\lambda)$. This means that close to equilibrium the entropy production becomes completely time symmetric. This should not be surprising as we are expanding the entropy production around its minimum, so that the corresponding CGF can be expressed in terms of quadratic functions, which do not distinguish between the driving $\dot{H}_t$ and its reverse $\dot{H}_t^{\rm rev} = -\dot{H}_{1-t}$.

Going back to Eq. (4.180) we can then obtain the relation:

$$\frac{p(\Sigma)}{p(-\Sigma)} = e^{\Sigma}\,, \tag{4.184}$$

which goes under the name of Evans-Searles fluctuation theorem [96]. It places a considerable constraint on the fluctuations in entropy production, with negative values exponentially suppressed. Moreover, it should be noticed that by splitting the entropy production in diagonal and coherent contribution $\Sigma = \Sigma^D + \Sigma^C$ one obtains the two independent Evans-Searles fluctuation theorems:

$$\frac{p(\Sigma_x)}{p(-\Sigma_x)} = e^{\Sigma_x}, \quad x = \{D, C\}\,, \tag{4.185}$$

which further constrain the distribution of entropy production. This discussion, together with the one from the previous section, show that the effects of the time symmetry, together with the separation of different channels of dissipation are at the origin of the simplicity with which we can give so general results close to equilibrium.

4.7 Engine optimisation

The maximum efficiency of ideal engines is given by the Carnot formula in Eq. (4.12). This result is universal, in the sense that it does not depend on any detail of the particular system in exam, but only on the temperatures of the two baths used to perform the cycle. It should be noticed, though, that the power output of such engines is zero, since this is defined as:

$$P := -\frac{\langle w \rangle}{\tau_{\rm tot}}\,, \tag{4.186}$$

and for ideal transformations one has $\tau_{\rm tot} = \infty$.

When moving to the slow driving regime one loses some part of the universality, but it is anyway possible to put forward a set of very general principles that optimal engines should satisfy. In particular, in this context we refer to optimality of a cycle as the property of maximising the power output for a fixed total time $\tau_{\rm tot}$.

Let us then consider the case of a cycle performed between n different baths, specified by their inverse temperature $\{\beta_i\}$. Each isothermal is performed in time τ_i, and the total time of the transformation is given by $\tau_{\rm tot} := \sum_i \tau_i$. Finally, we assume that different isothermals are connected through adiabatic jumps, in analogy with what happened for the Carnot engine in Fig. 4.1. Thanks to the first law, after a full cycle one has $\langle w \rangle = \sum_i Q_i$, where Q_i denote the heat exchanged with the i-th bath. Then, the power output can be expressed as:

$$P = -\frac{\langle w \rangle}{\tau_{\rm tot}} = -\frac{\sum_i Q_i}{\tau_{\rm tot}} = \frac{\sum_i \beta_i^{-1}(\Delta S_i - \langle \Sigma_i \rangle)}{\sum \tau_i} \,. \tag{4.187}$$

Since the power output is a positive quantity, it is apparent that, in order to maximise it, one needs to minimise the negative contributions coming from $\sum \langle \Sigma_i \rangle$. Moreover, it should be noticed that this expression of the power is general, so the same reasoning applies also outside of the slow driving regime. For this reason, we can express the 0-th principle of engines optimisation at finite time as:

0. Maximising the power at fixed total time is equivalent to minimising the dissipation.

Now, the particularity of the slow driving regime is that $\langle \Sigma_i \rangle = \mathcal{O}\left(\tau_i^{-1}\right)$. This means that in this context $(\tau_i \langle \Sigma_i \rangle)$ is a quantity that does not depend on the duration of the protocol. Moreover, it should be noticed that one can bound the total dissipation through a Cauchy-Schwartz inequality, given by:

$$\left(\sum_i \beta_i^{-1} \langle \Sigma_i \rangle\right)\left(\sum_j \tau_j\right) \geq \left(\sum_i \sqrt{\beta_i^{-1} \tau_i \langle \Sigma_i \rangle}\right)^2 . \tag{4.188}$$

The lower bound is saturated for $\{\beta_i^{-1} \langle \Sigma_i \rangle\}$ and $\{\tau_i\}$ being parallel, corresponding to the requirement $\beta_i^{-1} \langle \Sigma_i \rangle / \tau_i = const.$ This can also be expressed by saying that the dissipation rate is constant across different isothermal transformations, leading to the principle:

1. The minimum total dissipation in the slow driving regime is obtained by keeping the ratio $\beta_i^{-1} \langle \Sigma_i \rangle / \tau_i$ constant for all isothermal drives. This leads to the equality:

$$\sum_i \beta_i^{-1} \langle \Sigma_i \rangle = \frac{1}{\tau_{\text{tot}}} \left(\sum_i \sqrt{\beta_i^{-1} \tau_i \langle \Sigma_i \rangle} \right)^2 . \tag{4.189}$$

It should be noticed again that thanks to the slow driving assumption $(\tau_i \langle \Sigma_i \rangle)$ is a constant function of τ_i, so the lower bound in Eq. (4.188) does not depend on time.

Using Eq. (4.189) one can substitute the total time appearing in the definition of the power as:

$$P = \frac{\left(\sum_i \ \beta_i^{-1} \Delta S_i\right) \left(\sum_i \ \beta_i^{-1} \langle \Sigma_i \rangle\right) - \left(\sum_i \ \beta_i^{-1} \langle \Sigma_i \rangle\right)^2}{\left(\sum_i \ \sqrt{\beta_i^{-1} \tau_i \langle \Sigma_i \rangle}\right)^2} . \tag{4.190}$$

We can carry out the maximisation of the numerator and the minimisation of the denominator independently. Then, starting from the latter we know that the minimum dissipation along each cycle is obtained following geodesics paths. Following the discussion around Eq. (4.57) this implies that the entropy production rate $\langle \dot{\Sigma}_t \rangle$ is constant along each of the trajectories. Moreover, since $(\tau_i \langle \Sigma_i \rangle)$ is also constant, this implies that the entropy production rate is always the same for all the transformations. We can summarise this discussion in the two principles:

2. The isothermal transformations of optimal engines in the slow driving regime follow the corresponding geodesic drives;

3. The entropy production rate $\langle \dot{\Sigma}_t \rangle$ of optimal engines in the slow driving regime is constant along the whole cycle.

A simple corollary of the principles above is that in optimal engines the creation of coherence is avoided, as in the slow driving regime coherences are always detrimental. Hence, we can state the following principle:

4. Optimal engines in the slow driving regime avoid the creation of coherences, meaning that the Hamiltonians $\{H_t\}$ are all diagonal in the same basis.

We can now pass to the optimisation of the numerator. In this case, by inspection of Eq. (4.190) it is clear that this is obtained for $\sum_i \beta_i^{-1} \langle \Sigma_i \rangle = \frac{1}{2} \left(\sum_i \beta_i^{-1} \Delta S_i \right)$, which gives:

$$P_{\max} = \left(\frac{\left(\sum_i \beta_i^{-1} \Delta S_i \right)}{2 \left(\sum_i \sqrt{\beta_i^{-1} \tau_i \langle \Sigma_i \rangle} \right)} \right)^2 = \frac{\sum_i \beta_i^{-1} \Delta S_i}{2 \tau_{\text{tot}}} , \tag{4.191}$$

where in the last step we used Eq. (4.189) to substitute back τ_{tot}. It should be noticed that the expression in the numerator corresponds to the work output for an ideal cycle. Hence, we obtain that in the slow driving regime the maximum power of an engine is given by half the maximum work output along that cycle divided by the total time. A similar result was discussed in [97] for engines that can be optimised by keeping the entropy production rate constant. Still, here we show that this assumption applies to any engine operated close to equilibrium.

At this point it is useful to recall the thermodynamic length inequality in Eq. (4.57), given by:

$$\langle \Sigma_i \rangle \geq \frac{1}{\tau_i} \ell_{\gamma_i}^2 , \tag{4.192}$$

which is saturated along geodesics. Then, we can rewrite the maximum power in Eq. (4.191) as:

$$P_{\max} = \frac{\left(\sum_i \beta_i^{-1} \Delta S_i \right)^2}{2 \left(\sum_i \sqrt{\beta_i^{-1}} \ell_{\gamma_i} \right)^2} . \tag{4.193}$$

This quantity only depends on the endpoints chosen for the transformations. In [6] it was shown that the optimal power is always obtained by restricting the number of baths to two, i.e., for a Carnot like cycle. This should not be surprising, as the same behaviour also holds for ideal engines. Hence, in the following, without loss of generality, we will assume that the cycle operates between two baths only.

Moreover, we further assume that the thermodynamic length satisfies a relation of the form $\sqrt{\beta_c^{-1}} F(\beta_c) \ell_{\gamma_c} = \sqrt{\beta_h^{-1}} F(\beta_h) \ell_{\gamma_h}$ for some function F. This assumption is well justified when all the baths have the same spectral density [6], so this does not restrict too much the validity of the

discussion below. Then, simple manipulations of Eq. (4.193) leads to:

$$P_{\max}^{\text{Carnot}} = \frac{\beta_h^{-1}}{4} \frac{\Delta S_h^2}{\ell_{\gamma_h}^2} \left(\frac{(\beta_c - \beta_h) F(\beta_c)}{(F(\beta_c) + F(\beta_h)) \beta_c} \right)^2 \tag{4.194}$$

where we implicitly used the relation $\Delta S_c = -\Delta S_h$ that holds for cycles. Hence, in order to complete the optimisation, we should maximise the ratio $\Delta S_h^2 / \ell_{\gamma_h}^2$. Interestingly, even this problem can be solved by resorting to the Cauchy-Schwarz inequality [98]. To this end, it is useful to express the difference of entropy between the endpoints in terms of the Kubo-Mori-Bogoliubov metric:

$$\Delta S = S(\pi_\beta(H_\tau)) - S(\pi_\beta(H_0)) = \int_0^\tau \mathrm{d}t \ \frac{\mathrm{d}}{\mathrm{d}t} S(\pi_\beta(H_t)) = \tag{4.195}$$

$$= \beta \int_0^\tau \mathrm{d}t \ \frac{\mathrm{d}}{\mathrm{d}t} \operatorname{Tr}\left[\frac{e^{-\beta H_t}}{\mathcal{Z}_t} (H_t - \beta^{-1} \mathcal{Z}_t) \right] = \tag{4.196}$$

$$= -\beta^2 \int_0^\tau \mathrm{d}t \ \operatorname{Tr}\left[\Delta H_t \, \mathbb{J}_{t,L}[\Delta \dot{H}_t] \right] . \tag{4.197}$$

It should be noticed that if there exists a single equilibration timescale in the system one can rewrite the thermodynamics metric as $(g_{i,j})_t = \tau^{\text{eq}} (m_{i,j}^{KMB})_t$ (see, for example, Eq. (4.89))[1]. Then, it follows that:

$$\frac{\Delta S_h^2}{\ell_{\gamma_h}^2} = \frac{\Delta S_h^2}{\tau_h \langle \Sigma \rangle} = \frac{\beta_h^2 \left(\int_0^{\tau_h} \mathrm{d}t \ \sum_{i,j} (m_{i,j}^{KMB})_t (\Delta H_t)_i (\Delta \dot{H}_t)_j \right)^2}{\tau_h \, \tau^{\text{eq}} \int_0^{\tau_h} \mathrm{d}t \ \sum_{i,j} (m_{i,j}^{KMB})_t (\Delta \dot{H}_t)_i (\Delta \dot{H}_t)_j} \leq \tag{4.198}$$

$$\leq \frac{\beta_h^2}{\tau_h \, \tau^{\text{eq}}} \int_0^{\tau_h} \mathrm{d}t \ \sum_{i,j} (m_{i,j}^{KMB})_t (\Delta H_t)_i (\Delta H_t)_j = \tag{4.199}$$

$$= \frac{\beta_h^2}{\tau_h \, \tau^{\text{eq}}} \int_0^{\tau_h} \mathrm{d}t \ \operatorname{Tr}\left[\Delta H_t \, \mathbb{J}_{t,L}[\Delta H_t] \right] = \frac{1}{\tau_h \, \tau^{\text{eq}}} \int_0^{\tau_h} \mathrm{d}t \ C_{H_t}(\beta_h) \, , \tag{4.200}$$

where we used the fact that for geodesics the equality $\ell_{\gamma_h}^2 = (\tau_h \langle \Sigma \rangle)$ holds (see Eq. (4.57)), together with the definition of heat capacity in Eq. (1.223). Hence, the inequality just found tells us that:

$$\frac{\Delta S_h^2}{\ell_{\gamma_h}^2} \leq \frac{1}{\tau_h \, \tau^{\text{eq}}} \int_0^{\tau_h} \mathrm{d}t \ C_{H_t}(\beta_h) \, , \tag{4.201}$$

[1] If this is not the case, one can always use the bound $(g_{i,j})_t \geq \tau_{\min}^{\text{eq}} (m_{i,j}^{KMB})_t$, where $\tau_{\min}^{\text{eq}}$ is the smallest equilibration time.

which can be further bounded as:

$$\frac{\Delta S_h^2}{\ell^2_{\gamma_h}} \leq \frac{1}{\tau_h\, \tau^{\rm eq}} \int_0^{\tau_h} {\rm d}t\ C_{H_t}(\beta_h) \leq \frac{1}{\tau^{\rm eq}} \max_{\gamma_h} C_{H_t}(\beta_h)\,, \tag{4.202}$$

where we simply upper bounded the integrand by its maximal value. The first inequality is saturated if the integrand in Eq. (4.197) is constant, i.e., if there is a uniform creation of entropy along the trajectory. The second inequality can be saturated by doing infinitesimal cycles around the maximum of γ_h. This leads to the principle:

5. Optimal engines in the slow driving regime with a single thermalisation timescale operate in infinitesimal cycles.

Moreover, the best position in the parameter space where to perform the infinitesimal cycle is exactly the one that maximises the heat capacity:

6. The infinitesimal cycles of optimal engines in the slow driving regime with a single thermalisation timescale are peaked around the maximum of the heat capacity in the Hamiltonian space.

The discussion presented in this section gives a complete optimisation of the power output of engines close to equilibrium: principle 1 tells us how to choose the different times $\{\tau_i\}$, namely in such a way that the entropy production is constant for each isothermal; then, principle 2, 3 and 4 specify the driving, constraining it to be geodesic and without coherences; finally, principle 5 and 6 tell us how to choose the driving among all the possible geodesics, showing that infinitesimal cycles around the maximum heat capacity are optimal. In this way one arrives to the expression for the maximum power:

$$P^{\rm Carnot}_{\rm max} = \frac{\beta_h^{-1}}{4\tau^{\rm eq}} \left(\max_H C_H(\beta_h)\right) \left(\frac{(\beta_c-\beta_h)\, F(\beta_c)}{(F(\beta_c)+F(\beta_h))\,\beta_c}\right)^2 . \tag{4.203}$$

This equation shows that there are no other free parameters to optimise, giving a universal expression of the maximum power depending only on the temperatures of the baths, together with the system dependent quantities $\tau^{\rm eq}$, $C_H(\beta)$ and $F(\beta)$.

4.8 Signature of contextuality in the linear response regime

All the results presented so far are derived in the TPM scheme, so it is hard to imagine that any genuinely quantum effects could arise in

this context (see Thm. 16). Emblematic of the problem is the quench model in Eq. (4.33), in which the state before each quench is completely thermal, i.e., with no coherence, so that the TPM and WWS schemes coincide. Hence, one could expect that even by studying the statistics of work defined in a more general scheme could not result in any genuinely quantum effects, as the property of being close to equilibrium is enough to wash out any contextuality in the process.

Interestingly, this is not the case: one can find contextual advantages even in the linear response regime [99]. In particular, consider an engine defined as follows: first, the system is coupled at the same time to a hot and a cold bath in order to generate a non-equilibrium steady state ρ; at this point the system is driven unitarily in time τ through the transformation generated by the Hamiltonian:

$$H(t) = H_0 + g\,V(t)\,; \qquad\qquad V(0) = V(\tau) = 0\,. \tag{4.204}$$

Notice that the second condition allows us to refer to this as a cyclic transformation. Moreover, it should be noticed that the linear response is defined by the regime in which $(g\tau) \ll 1$, i.e., for very short duration of the pulses, or for small interactions strength. In the following we will consider the second case. This is an implicit assumption in the physical derivation of the Lindbladian equation, so the subject of this section is as general as the results presented in the TPM scheme. Then, the average change of energy during one cycle is given by:

$$\left\langle w^Q \right\rangle = \mathrm{Tr}\left[\rho\, U_\tau^\dagger\, H_0\, U_\tau\right] - \mathrm{Tr}\left[\rho\, H_0\right] = \tag{4.205}$$

$$= -i \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\rho\, U_t^\dagger [H_0, H(t)] U_t\right]\,. \tag{4.206}$$

At this point it is useful to define the Hamiltonian in the interaction picture, i.e., $H_I(t) = e^{iH_0 t}\, H(t)\, e^{-iH_0 t} = H_0 + g\, V_I(t)$. Indeed, by expanding Eq. (4.206) at first order in powers of g one obtains:

$$\left\langle w^Q \right\rangle = -ig \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\rho\, e^{iH_0 t}\, [H_0, V(t)] e^{-iH_0 t}\right] + \mathcal{O}\left(g^2\right) = \tag{4.207}$$

$$= -ig \int_0^\tau \mathrm{d}t\ \mathrm{Tr}\left[\rho\, [H_I(0), H_I(t)]\right] + \mathcal{O}\left(g^2\right)\,, \tag{4.208}$$

where in the last step we highlighted the relation between $\left\langle w^Q \right\rangle$ and the two-points correlation function. Expressing Eq. (4.208) in coordinates

one also gets:

$$\left\langle w^{Q}\right\rangle = -2g\tau \sum_{i} E_0^{(i)} \operatorname{Im}\operatorname{Tr}\left[\rho\,\Pi_0^{(i)}\left(\frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\ V_I(t)\right)\right] + \mathcal{O}\left(g^2\right) = \tag{4.209}$$

$$= -2g\tau \sum_{i} E_0^{(i)} \operatorname{Tr}\left[\rho\,\Pi_0^{(i)}\right] \operatorname{Im}\left[\frac{\operatorname{Tr}\left[\rho\,\Pi_0^{(i)} X\right]}{\operatorname{Tr}\left[\rho\,\Pi_0^{(i)}\right]}\right] + \mathcal{O}\left(g^2\right)\,, \tag{4.210}$$

where we implicitly defined the operator $X := \frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\ V_I(t)$. It should be noticed that the ratio in Eq. (4.210) has the same structure as the weak values defined in Eq. (B.15). Indeed, it was shown in [100] that also the imaginary part of the weak values can give rise to witnesses of contextuality.

Indeed, this is the case for $\left\langle w^{Q}\right\rangle$ under some assumptions on the unitary U_t. In particular, define the superoperator $\mathcal{U}_\tau[\cdot] := U_\tau\,(\cdot)U_\tau^{\dagger}$. Suppose then that for g small enough it holds that:

$$\frac{1}{2}\left(\mathcal{U}_\tau + \mathcal{U}_\tau^{\dagger}\right) = (1-p_d)\,\mathbb{I} + p_d\,\mathcal{C}\,, \tag{4.211}$$

where $\mathcal{C}$ is some arbitrary quantum channel and $p_d = \mathcal{O}\left(g^2\right)$. It was proven in [99] that this can be guaranteed through the following theorem:

Theorem 22. *Define the matrix:*

$$J_{kj} := 1 - \frac{(\alpha_k - \alpha_j)^2}{C} \tag{4.212}$$

where $C > 0$ and α_i are the eigenvalues of the operator X defined above. Then, if there exists a C such that J is positive semidefinite, for g small enough the decomposition in Eq. (4.211) *holds.*

From an ontological model point of view, Eq. (4.211) means that there are two transformations T_τ and T_τ^{*}, not necessarily unitary, which are not too far from the trivial transformation $T_{\mathbb{I}}$ under uniform mixing. Moreover, since $\mathcal{C}$ is a channel, it also implies that there exists a third transformation $T_{\mathcal{C}}$ such that:

$$\frac{1}{2}\left(T_\tau + T_\tau^{*}\right) =_{\mathrm{op}} (1-p_d)\,T_{\mathbb{I}} + p_d\,T_{\mathcal{C}}\,. \tag{4.213}$$

Then, one can apply the following theorem:

Theorem 23. *Assuming the validity of Eq.* (4.213)*, for any non-contextual ontological model and any observable* $O := \sum_i O_i \, |i\rangle\langle i|$*, the average difference:*

$$\langle \Delta O \rangle_\tau := \sum_i O_i \left(p(i|T_\tau(\mathcal{P}), \mathcal{M}) - p(i|\mathcal{P}, \mathcal{M}) \right) , \tag{4.214}$$

corresponding to the difference between the average at time τ *and time* $t = 0$*, is bounded by:*

$$| \langle \Delta O \rangle_\tau | \leq 4 \, p_d \, O_{\max} , \tag{4.215}$$

where $O_{\max}$ *is the largest eigenvalue of* O*.*

We refer to [99] for the proof. The theorem just presented shows that for any non-contextual transformation that can be approximately reversed (in the sense of Eq. (4.213)) the linear response scales as $p_d = \mathcal{O}\left(g^2\right)$. On the other hand, from Eq. (4.210) we know that the average change in energy scales as $\mathcal{O}\left(g\right)$ if the imaginary part of the anomalous weak values for X is non-zero. Hence, for g small enough the work output $\left\langle w^Q \right\rangle$ can be bigger than the one generated by any non-contextual counterpart. Moreover, the power output in this context is given by:

$$P^Q = 2 \, g \sum_i E_0^{(i)} \, \mathrm{Im} \, \mathrm{Tr} \left[\rho \, \Pi_0^{(i)} \, X \right] + \mathcal{O}\left(g^2\right) . \tag{4.216}$$

Since there is no explicit dependence on time in this expression, no non-contextual ontological model can reproduce this power output (for g small enough) even if one allows for a longer time of simulation then the one of the original cycle.

This discussion concludes the chapter about thermodynamics close to equilibrium. As we saw, the assumption of slow driving allows for many simplifications and universal results, and, as it was shown in this last section, could still give rise to interesting genuinely quantum effects.

Chapter 5

Map of results and Outlook

Given the way in which this thesis was devised and written, it might be difficult to properly discern between the contributions of the author and the results that were already known in the literature. The rationale behind this choice is to favour the natural stream of thoughts and to provide an organic treatment of the subject, rather than to bend the exposition to the artificial need of partitioning between results of the author as opposed to the one of other scientists. Still, we take this section to provide a brief summary of the results of the author (and colleagues), divided by chapters to facilitate the identification in the main text. Moreover, this map of the results disseminated in the thesis is also complemented by the questions that still remain unresolved and that would be worth to look at in the future.

Results contained in Chapter 1: Fisher information and the geometry of quantum states

The discussion presented in Sec. 1.1–1.4 about the quantum Fisher information was already known in the literature. In [1] these results were reviewed and presented in a unified manner. Moreover, in the same publication (and in [4]) the dynamical nature of the Fisher information was discussed. The beginning of Sec. 1.5, together with Sec. 1.5.1–1.5.4, are completely new and derive from the two papers mentioned. Part of the characterisation of the detailed balance was already recognised in the literature, so it is fair to say that the only genuinely new result of Sec. 1.5.5 is the expression of the generator for a (Fisher) detailed balance dynamics in Eq. (1.343) (also contained in [1]).

Open questions:

- We know that for the Bures metric and the Wigner-Yanase skew information (defined in Sec. 1.4.2 and Sec. 1.4.5) a closed formula of the geodesic distance can be found. The particularly simple expressions connected to these two metrics arise from the minimality of the Bures metric, on the one hand, on the other, from the property of the Wigner-Yanase metric to have a constant curvature on all the space of states (which allows one to reduce the problem to the one of finding the geodesics on a sphere [29]). One question of particular interest is the following: is it possible to find a general solution for the geodesic distance for any standard monotone function? Or at least in the case of the relative entropy, or the family of Rényi-divergences? Such a solution would allow to give neat lower bounds, characterising, for example, the minimum average dissipation and similar quantities of operational relevance.

- One of the reason behind the interest for the Bures metric is that it is directly connected to the estimation of parameters encoded in unitary evolutions through the Cramer-Rao bound. Indeed, suppose the state ρ is evolved in the imaginary time with the Hamiltonian $\tilde{H}$. The susceptibility of the state at time $t = 0$ is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\rho(t)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\,e^{t\,\tilde{H}}\rho(0)e^{t\,\tilde{H}}\bigg|_{t=0} = 2\,\mathbb{J}_B\big|_{\rho(0)}[\tilde{H}]\,. \tag{5.1}$$

  If one compares the expression just obtained with the symmetric logarithmic derivative defined in Eq. (1.138), it is apparent that up to an unimportant factor 2, the two actually coincide. A similar behaviour also appears for another encoding, namely $\rho(t) := e^{H_0+t\tilde{H}}$, which is particularly useful in the context of linear response of thermal states. Indeed, in this case one obtains (see Eq. (4.26)):

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\rho(t)\bigg|_{t=0} = \mathbb{J}_L\big|_{\rho(0)}[\tilde{H}]\,. \tag{5.2}$$

  It is no coincidence that the two examples above are the only two for which the Cramer-Rao bound seems to have some physical relevance, as they are also the only two expressions (to the best of the author's knowledge) for which the following equation holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\rho(t)\bigg|_{t=0} = \mathbb{J}_f\big|_{\rho(0)}[\tilde{H}]\,. \tag{5.3}$$

States satisfying Eq. (5.3) are called exponential families associated to the function f. It is then an open question whether one can find a closed expression for the exponential family associated to some other f (and possibly for the general case).

- Despite the rich mathematical structure associated to a Riemannian manifold, it should be noticed that in our analysis of the relation between evolutions and Fisher information we only have analysed maps that contract the scalar product, or that are self-adjoint with respect to it. These are two of the most simple properties one can explore, while it is still an open problem what would happen if one looked at more sophisticated quantities, like scalar curvature or the Riemann tensor. The same type of problem will also arise in the context of Chapter 4, and it boils down to the question of whether there is a physical interpretation to the Riemannian properties of the Fisher metric.

- Finally, as it was mentioned in the main text, it would be interesting to study whether the equivalence between physical dynamics and maps that contract the Fisher information could be used in some form of reconstruction of quantum mechanics [101]. For example, once the Hilbert space formalism was obtained following axioms coming from information theory principles, one could define the possible evolutions just in terms of the Fisher information metric.

Results contained in Chapter 3: The second law

Most of the results presented in Chapter 3 were already well known in the literature. In the discussion of the Maxwell's demon, Sec. 3.1.2 closely follows (and partially expands) the work by Zurek [64, 65]. In particular, the proposal to use Levin's complexity (see Eq. (3.7)) is new to the best of the author's knowledge. On the same note, the discussion of the relation between Crooks relations and algorithmic coding theorem (Sec. 3.2.4) is also original. Finally, a more concrete result is the one discussed in Sec. 3.3, namely the axiomatic definition of Bayes' inversion, which was taken from [2]. The content of this paper is further explored in Appendix C.

Open questions:

- The treatment of Maxwell's demon presented in Sec. 3.1 hints at a subtle connection between information processing and thermo-

dynamics. Is it possible to make this relation more precise? For example, by looking at the exactly solvable model of Maxwell's demon in Sec. 3.1.1, one can see that there is a relation between the randomness of the bits stored in demon's memory and the amount of extracted work. Is it possible to rephrase this concept in purely algorithmic terms? Or, to say it with Maxwell's words, is this just a cunning example of nephelococcygia?

- In [2] we found numerical evidences that requiring a retrieval map to be involutive (see condition 5) is key to isolate Bayes' reverse map. Can this be analytically proven? Is it possible to give a similar result also for the quantum case?

Results contained in Chapter 4: Thermodynamic transformations close to equilibrium

All of the results in Chapter 4 are original work, apart from Sec. 4.8 about contextuality in linear response theory. In particular, the general expression of the cumulant generating function of work (see Sec. 4.2) was taken from [7], while the considerations about the metric structure associated to the average dissipation (Sec. 4.3) summarise the results from [6] and [9]. The chapter about fluctuations of work reviews the results from [8], and contains some expressions previously unpublished (in particular we define a whole family of positive quantum signatures in Eq. (4.146)). Sec. 4.5 and 4.6 are taken from [7], while Sec. 4.7 is from [6].

Open questions:

- The quantities discussed in Chapter 4 are defined in the TPM scheme. For this reason, one could argue that the effects that we refer as quantum are not genuinely so, as they admit a non-contextual model simulating them. This raises the question: is it possible to derive similar results to the one presented in Sec. 4.2–4.7 starting from the WWS, or with some construction in the spirit of Sec. 4.8? In particular, it should be noticed that Thm. 23 only discusses average quantities, while higher order statistics are not considered. Could one find some quantifier akin to $\mathcal{Q}$, defined in Eq. (4.131), whose positivity would witness contextuality?

- As discussed in the open questions of Chapter 1, despite the rich geometrical structure, we only make use of the concept of geodesics,

ignoring any other Riemannian properties of the manifold of thermal states. This approach descends from a tendency to not dwell on mathematical technicalities if they are not justified from a physical interpretation. Still, it is worth wondering whether one could give any physical meaning to the whole differential geometrical structure arising in the presence of a Riemannian metric. A result in this direction is the work [102] which connects the thermodynamic curvature of classical systems with their correlation length. Then, it would be interesting to verify whether this result can be extended to the quantum regime.

- One of the problems of the resource theory of thermodynamics, apart from the plethora of different definitions, is that it is hardly connected with what people usually call thermodynamic transformations. This difference in language could be noticed in Sec. 4.6.1 where, despite the effort to identify quantities defined in the two frameworks, the overall connection is still partly loose. In [103] continuous thermalising transformations were formalised in the framework of resource theory, providing a first step to bridge the two formalisms. In that work the system Hamiltonian needs to be constant during the transformation, implying that only the production of heat could be characterised, but not of work. Indeed, the general problem of constructing a resource theory of continuous thermodynamic operations seems rather challenging and still far away. Still, could restricting to transformations close to equilibrium help, given the many simplifications present in this regime?

Part III

Appendices

Appendix A

Dynamical properties of Fisher information

A.1 Derivation of the flux of Fisher information

We present here the derivation of Thm. 8. In particular, we want to study the evolution of the Fisher information:

$$\mathcal{F}_t := \mathrm{Tr}\left[\delta\rho_t \, \mathbb{J}_f^{-1}\big|_{\pi_t}[\delta\rho_t]\right] , \tag{A.1}$$

where $\pi_t := \Phi_t(\pi)$ and $\delta\rho_t := \Phi_t(\delta\rho)$. Using the integral expression in Eq. (1.78), we can rewrite the Fisher information as:

$$\mathcal{F}_t = 2\,\mathrm{Re}\int_0^1 \mathrm{d}N_g(s)\ \mathrm{Tr}\left[\delta\rho_t \, (\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]\right] , \tag{A.2}$$

where the real part comes from the fact that $(\mathbb{L}_\pi + s\mathbb{R}_\pi)^\dagger = (\mathbb{R}_\pi + s\mathbb{L}_\pi)$. This expression is particularly convenient for calculations, due to the simple form that the derivative of $(\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1}$ takes. In fact, this is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1} = -(\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1}(\mathbb{L}_{\dot\pi_t} + s\,\mathbb{R}_{\dot\pi_t})(\mathbb{L}_{\pi_t} + s\,\mathbb{R}_{\pi_t})^{-1} , \tag{A.3}$$

where $\dot\pi_t$ is simply the derivative of the state. This expression can be proved by noticing that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{L}_{\pi_t} = \mathbb{L}_{\dot\pi_t}$ (and similarly for $\mathbb{R}_{\pi_t}$) and by

taking the derivative of:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left((\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}\right)= \tag{A.4}$$

$$=(\mathbb{L}_{\dot{\pi}_t}+s\,\mathbb{R}_{\dot{\pi}_t})(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}+(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}= \tag{A.5}$$

$$=\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{I}=0\,, \tag{A.6}$$

which directly implies Eq. (A.3). Given this technical tool, we can start analysing the evolution of $\mathcal{F}_t$ under the dynamics generated by the Lindbladian:

$$\mathcal{L}_t[\rho]=-i[H_t,\rho]+\sum_{\alpha}^{d^2}\lambda_\alpha\left(A_\alpha(t)\,\rho\,A_\alpha(t)^\dagger-\frac{1}{2}\{A_\alpha(t)^\dagger A_\alpha(t),\rho\}\right). \tag{A.7}$$

Notice that, since $\mathcal{F}_t$ is invariant under unitary transformations, there is no contribution coming from the commutator in the previous equation. Moreover, since the derivative is linear, it decomposes into a sum of the form:

$$\dot{\mathcal{F}}_t=\sum_{\alpha}\lambda_\alpha(t)\;\mathcal{I}_\alpha(t)\,, \tag{A.8}$$

where each current $\mathcal{I}_\alpha(t)$ only contains the corresponding jump operator $A_\alpha(t)$, together with its adjoint. For this reason, without loss of generality, we consider here Lindblad operators generated by a single jump operator. In order to shorten the notation we also assume that the jump operator, denoted by A, is time independent, again without loss of generality. We start by rewriting the derivative of the Fisher information as:

$$\dot{\mathcal{F}}_t=2\,\mathrm{Re}\int_0^1\mathrm{d}N_g(s)\left(2\,\mathrm{Tr}\left[\delta\dot{\rho}_t(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]\right]+\right.$$
$$\left.+\,\mathrm{Tr}\left[\delta\rho_t\left(\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}\right)[\delta\rho_t]\right]\right). \tag{A.9}$$

The second term in the integral can be expanded as:

$$\mathrm{Tr}\left[\delta\rho_t\left(\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}\right)[\delta\rho_t]\right]= \tag{A.10}$$

$$=-\mathrm{Tr}\left[\delta\rho_t(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}(\mathbb{L}_{\dot\pi_t}+s\,\mathbb{R}_{\dot\pi_t})(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]\right]= \tag{A.11}$$

$$=-\mathrm{Tr}\left[B_s(t)^\dagger\,\dot\pi_t\,B_s(t)\right]-s\,\mathrm{Tr}\left[B_s(t)^\dagger B_s(t)\,\dot\pi_t\right]= \tag{A.12}$$

$$=-\mathrm{Tr}\left[B_s(t)^\dagger\,\mathcal{L}(\pi_t)\,B_s(t)\right]-s\,\mathrm{Tr}\left[B_s(t)^\dagger B_s(t)\,\mathcal{L}(\pi_t)\right], \tag{A.13}$$

where we introduced the notation $B_s(t):=(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]$. On the other hand, the first term in Eq. (A.9) simply gives:

$$2\,\mathrm{Tr}\left[\delta\dot\rho_t(\mathbb{L}_{\pi_t}+s\,\mathbb{R}_{\pi_t})^{-1}[\delta\rho_t]\right]= \tag{A.14}$$

$$=2\,\mathrm{Tr}\left[\mathcal{L}(\delta\rho_t)B_s(t)\right]=2\,\mathrm{Tr}\left[\mathcal{L}((\mathbb{R}_{\pi_t}+s\,\mathbb{L}_{\pi_t})[B_s(t)^\dagger])B_s(t)\right]= \tag{A.15}$$

$$=2\,\mathrm{Tr}\left[\mathcal{L}(B_s(t)^\dagger\pi_t)B_s(t)\right]+2s\,\mathrm{Tr}\left[\mathcal{L}(\pi_tB_s(t)^\dagger)B_s(t)\right], \tag{A.16}$$

where in the second line we have multiplied and divided by $(\mathbb{R}_{\pi_t}+s\,\mathbb{L}_{\pi_t})$ to obtain $B_s(t)^\dagger$. We can now proceed in summing up Eq. (A.13) and Eq. (A.16). Due to the number of terms that will appear, though, we first consider the first traces in both equations, and then the second ones. Hence, summing the first term in Eq. (A.13) and the first of Eq. (A.16), and explicitly expanding the Lindbladian, we obtain:

$$\begin{aligned}2\,\mathrm{Tr}\left[A\,B_s(t)^\dagger\pi_tA^\dagger B_s(t)\right]&-\cancel{\mathrm{Tr}\left[B_s(t)^\dagger\pi_tA^\dagger A\,B_s(t)\right]}\\&-\mathrm{Tr}\left[A^\dagger A\,B_s(t)^\dagger\pi_tB_s(t)\right]+\\-\mathrm{Tr}\left[B_s(t)^\dagger A\,\pi_tA^\dagger\,B_s(t)\right]&+\cancel{\frac{1}{2}\mathrm{Tr}\left[B_s(t)^\dagger\pi_tA^\dagger A\,B_s(t)\right]}\\&+\cancel{\frac{1}{2}\mathrm{Tr}\left[B_s(t)^\dagger A^\dagger A\,\pi_tB_s(t)\right]},\end{aligned} \tag{A.17}$$

where the first three terms come from Eq. (A.13), and the last three from Eq. (A.16). Notice that the second term corresponding to Eq. (A.13) cancels with the last two terms coming from Eq. (A.16). Indeed, we can take the conjugate of the last trace without affecting the result, thanks to the real part in Eq. (A.9). The remaining terms can be further simplified

to give:

$$2\mathrm{Tr}\left[\pi_t A^\dagger B_s(t) A\, B_s(t)^\dagger\right] - \mathrm{Tr}\left[\pi_t B_s(t) A^\dagger A\, B_s(t)^\dagger\right] - \mathrm{Tr}\left[\pi_t A^\dagger\, B_s(t) B_s(t)^\dagger A\right] = \tag{A.18}$$

$$= \mathrm{Tr}\left[\pi_t \left[A^\dagger, B_s(t)\right] A B_s(t)^\dagger\right] + \mathrm{Tr}\left[\pi_t\, A^\dagger B_s(t)\left[A, B_s(t)^\dagger\right]\right] = \tag{A.19}$$

$$= -\mathrm{Tr}\left[\pi_t\, B_s(t) A^\dagger \left[A, B_s(t)^\dagger\right]\right] + \mathrm{Tr}\left[\pi_t\, A^\dagger B_s(t)\left[A, B_s(t)^\dagger\right]\right] = \tag{A.20}$$

$$= \mathrm{Tr}\left[\pi_t \left[A^\dagger, B_s(t)\right]\left[A, B_s(t)^\dagger\right]\right] = \tag{A.21}$$

$$= -\mathrm{Tr}\left[\pi_t \left[A, B_s(t)^\dagger\right]^\dagger \left[A, B_s(t)^\dagger\right]\right] , \tag{A.22}$$

where in the first line we used the cyclicity of the trace to put in evidence π_t, then in the third line we took the complex conjugate of the trace (again exploiting the fact that only the real part contributes to the Fisher information), and used the identity $[X, Y]^\dagger = [Y^\dagger, X^\dagger] = -[X^\dagger, Y^\dagger]$.

We can now pass to examine the last two terms in Eq. (A.13) and Eq. (A.16). Since the steps are completely analogous to the ones in the previous derivation, we present them all together. Carrying them out gives:

$$2s\,\mathrm{Tr}\left[A\pi_t B_s(t)^\dagger A^\dagger B_s(t)\right] - \cancel{s\mathrm{Tr}\left[A^\dagger A \pi_t B_s(t)^\dagger B_s(t)\right]} - s\,\mathrm{Tr}\left[\pi_t B_s(t)^\dagger A^\dagger A\, B_s(t)\right] + \tag{A.23}$$

$$- s\,\mathrm{Tr}\left[B_s(t)^\dagger B_s(t) A \pi_t A^\dagger\right] + \cancel{\frac{s}{2}\,\mathrm{Tr}\left[B_s(t)^\dagger B_s(t) A^\dagger A \pi_t\right]} + \cancel{\frac{s}{2}\,\mathrm{Tr}\left[B_s(t)^\dagger B_s(t) \pi_t A^\dagger A\right]} = \tag{A.24}$$

$$= s\,\mathrm{Tr}\left[\pi_t \left[B_s(t)^\dagger, A^\dagger\right] B_s(t) A\right] + s\,\mathrm{Tr}\left[\pi_t\, B_s(t)^\dagger A^\dagger \left[B_s(t), A\right]\right] = \tag{A.25}$$

$$= -s\,\mathrm{Tr}\left[\pi_t\, A^\dagger B_s(t)^\dagger \left[B_s(t), A\right]\right] + s\,\mathrm{Tr}\left[\pi_t\, B_s(t)^\dagger A^\dagger \left[B_s(t), A\right]\right] = \tag{A.26}$$

$$= s\,\mathrm{Tr}\left[\pi_t \left[B_s(t)^\dagger, A^\dagger\right]\left[B_s(t), A\right]\right] = \tag{A.27}$$

$$= -s\,\mathrm{Tr}\left[\pi_t \left[B_s(t), A\right]^\dagger \left[B_s(t), A\right]\right] = \tag{A.28}$$

$$= -s\,\mathrm{Tr}\left[\pi_t \left[A, B_s(t)\right]^\dagger \left[A, B_s(t)\right]\right] . \tag{A.29}$$

This concludes the proof of Thm. 8. In fact, generalising again to the

case of many time dependent jump operators, we finally obtain:

$$\dot{\mathcal{F}}_t = \sum_\alpha \lambda_\alpha(t)\, \mathcal{I}_\alpha(t) = \tag{A.30}$$

$$= -2 \sum_\alpha \lambda_\alpha(t) \int_0^1 \mathrm{d}N_g(s) \left(\mathrm{Tr}\left[\pi_t \left[A_\alpha(t), B_s(t)^\dagger\right]^\dagger [A_\alpha(t), B_s(t)^\dagger]\right] + \tag{A.31}$$

$$+ s\, \mathrm{Tr}\left[\pi_t \left[A_\alpha(t), B_s(t)\right]^\dagger [A_\alpha(t), B_s(t)]\right] \Bigg), \tag{A.32}$$

where we can drop the real part, as both the terms in the integral are positive definite.

A.2 Markovianity for classical evolutions

Despite the richness of the Fisher information geometry for quantum systems when compared to the single metric in the classical case, all the results characterising the relation between Fisher information and Markovianity can be proved by reducing to essentially classical systems. For this reason, we briefly present a useful characterisation of Markovian stochastic maps.

Classical dissipative evolutions are described by stochastic maps, i.e., matrices Φ satisfying the two conditions:

$$\sum_i (\Phi)_{i,j} = 1\,; \tag{A.33}$$

$$(\Phi)_{i,j} \geq 0 \qquad \forall\ \{i,j\}\,, \tag{A.34}$$

where the first condition ensures the conservation of total probability, while the second is needed to make sure that states are mapped into states. In complete analogy to the quantum case, a family of stochastic maps Φ_t depending smoothly on t is called divisible if for any two times t and s ($t \geq s$) one can define an intermediate map $\Phi_{t,s}$ satisfying the relation $\Phi_t = \Phi_{t,s} \circ \Phi_s$. A divisible stochastic dynamics is Markovian if all the intermediate maps $\Phi_{t,s}$ are stochastic.

The smoothness in t allows to define the rate matrix R_t through the equation

$$R_t := \lim_{\varepsilon \to 0} \frac{\Phi_{t+\varepsilon,t} - \mathbb{I}}{\varepsilon}\,. \tag{A.35}$$

Then, since the composition of two stochastic maps is again stochastic, Markovianity holds if and only if R_t generates a stochastic evolution for any time t. For this reason, it is useful to characterise this kind of rate matrices. To this end, consider the matrix $\Phi_{t+\varepsilon,t} \simeq \mathbb{I}+\varepsilon\, R_t$. For the global evolution to be Markovian, this matrix should satisfy the two conditions in Eq. (A.33) and Eq. (A.34), namely:

$$\sum_i \left(\delta_{ij} + \varepsilon\,(R_t)_{i,j}\right) = 1 + \varepsilon \sum_i (R_t)_{i,j} = 1\,, \tag{A.36}$$

$$\delta_{ij} + \varepsilon\,(R_t)_{i,j} \geq 0 \qquad \forall\ \{i,j\}\,, \tag{A.37}$$

where δ_{ij} denotes the Kronecker delta. From the first condition we can deduce that $\sum_i (R_t)_{i,j} = 0$. In particular, highlighting the diagonal terms, one obtains $R^{(t)}_{j,j} = -\sum_{i\neq j} R^{(t)}_{i,j}$. Matrices satisfying this constraint can be decomposed as:

$$R_t = \sum_{i\neq j} a^{(t)}_{i\leftarrow j}\,(\,|i\rangle\langle j\,| - |j\rangle\langle j\,|) \tag{A.38}$$

for $a^{(t)}_{i\leftarrow j}$ some real coefficients. We assume that this condition is always satisfied, also for non-Markovian evolutions, as it corresponds to the requirement that the dynamics preserves the normalisation. In fact, since non-Markovian evolutions are trace preserving on their domain, one can argue by linearity that this condition can be extended to the whole space of states.

Moreover, the condition in Eq. (A.37) implies that $(R_t)_{i,j} \geq 0$ whenever $i \neq j$. In the parametrisation above this means that the rates satisfy $a^{(t)}_{i\leftarrow j} \geq 0$. Hence, Markovianity corresponds to the request of having positive rates $a^{(t)}_{i\leftarrow j}$ for all times.

It should be pointed out that the derivation just presented is completely analogous to the one for the Lindbladian operators. In fact, in the same way we characterised trace preserving maps with Thm. 7, one needs to do the same for stochastic matrices, leading to Eq. (A.38). Once this parametrisation is obtained, it is then straightforward to connect Markovianity and positivity of rates.

A.3 Use of the trace distance in non-Markovianity

The study of non-Markovianity is mainly carried out in terms of information quantifiers, like, for example, distinguishability distances. These are

functions devised to quantify how statistically distinguishable two states are, and they usually decrease under physical evolutions. Then, as this leads to a monotonic decrease in these distances under Markovian dynamics, any increase thereof signals the appearance of non-Markovianity. In this context, the most used example is given by the trace distance:

$$D_{\mathrm{Tr}}(\rho,\sigma) = \mathrm{Tr}\left[|\rho-\sigma|\right] . \tag{A.39}$$

This quantity is connected to the maximal probability p_d of distinguishing ρ from σ in a single shot measurement, thanks to the relation $p_d(\rho,\sigma) = (1 + D_{\mathrm{Tr}}(\rho,\sigma))/2$ [17]. Moreover, since it is translational invariant, it is particularly appealing for analytical calculations. In particular, suppose ρ and σ are two probability vectors, and define $\delta\rho := \sigma - \rho$. Then, the trace distance is given by:

$$D_{\mathrm{Tr}}(\rho,\sigma) = D_{\mathrm{Tr}}(\rho,\rho+\delta\rho) = \mathrm{Tr}\left[|\delta\rho|\right] . \tag{A.40}$$

If the evolution is described by the rate matrix R_t, it is straightforward to see that:

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{Tr}}(\rho,\sigma) = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Tr}\left[|\delta\rho|\right] = \sum_i \frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho_i| = \tag{A.41}$$

$$= \sum_i \mathrm{sign}(\delta\rho_i)\delta\dot{\rho}_i = \sum_i \mathrm{sign}(\delta\rho_i)\sum_j (R_t)_{i,j}\, d_j = \tag{A.42}$$

$$= \sum_{i\neq j} \mathrm{sign}(\delta\rho_i)\left(a^{(t)}_{i\leftarrow j}\delta\rho_j - a^{(t)}_{j\leftarrow i}\delta\rho_i\right) = \tag{A.43}$$

$$= \sum_{i\neq j} a^{(t)}_{j\leftarrow i}\,\delta\rho_i\left(\mathrm{sign}(\delta\rho_j) - \mathrm{sign}(\delta\rho_i)\right) . \tag{A.44}$$

where we used the parametrisation of the rate matrix in Eq. (A.38), and in the last line we swapped the indexes of the first term in order to put the coefficient $a^{(t)}_{j\leftarrow i}$ in evidence. It should be noticed that if all the rates are positive, then the sum will be negative: in fact, either $\mathrm{sign}(\delta\rho_j) = \mathrm{sign}(\delta\rho_i)$, in which case the term inside the parenthesis is zero, or $\mathrm{sign}(\delta\rho_j) = -\mathrm{sign}(\delta\rho_i)$, so that $\delta\rho_i(\mathrm{sign}(\delta\rho_j) - \mathrm{sign}(\delta\rho_i)) = -2\,\delta\rho_i\,\mathrm{sign}(\delta\rho_i) = -2|\delta\rho_i|$. This calculation shows explicitly how the trace distance decreases under Markovian maps.

Interestingly, it is a well known fact that an evolution is Markovian if and only if the trace norm of any vector $\boldsymbol{v}$ decreases monotonically [35]. For this reason, it would be natural to expect the same to hold for the

trace distance as well. It is easy to see, though, that this is false: one simple counterexample can be given in dimension $d = 2$, with $a_{1\leftarrow 2} < 0$ and $a_{2\leftarrow 1} > 0$, and the additional condition that $|a_{2\leftarrow 1}| > |a_{1\leftarrow 2}|$. Plugging such choice in Eq. (A.44), and using the fact that in dimension 2 in order for $\delta\rho$ to be traceless it must satisfy $\delta\rho_1 = -\delta\rho_2$, it is easy to check that the derivative of the trace distance stays negative. Still, there is no contradiction between the two results: indeed, the trace distance can only access vectors of the form $\delta\rho = \sigma - \rho$, i.e., that are traceless. This reduces the dimension of the vectors tested by one. In fact, by choosing the traceful vector $v_i = \delta_{i2}$, one is able to witness non-Markovianity in the counterexample just presented.

Suppose now to have access to extra ancillary degrees of freedom on which the dynamics acts trivially, i.e., the global evolution is given by $\Phi_t \otimes \mathbb{I}_{d_A}$, where d_A is the dimension of the ancilla. Then, one can always find product states ρ and σ on the composite space such that $\mathrm{Tr}_S[\rho - \sigma] \neq 0$ while $\mathrm{Tr}_A[\rho - \sigma] = 0$, so that the total trace is zero. In this way, ancillary degrees of freedom give access to traceful vectors, and so to the possibility of witnessing non-Markovianity. A similar argument was also presented in [39] for the case of quantum dynamics. These results lead to the following:

Theorem 24. *Given a divisible dynamics Φ_t, there always exists an ancilla of finite dimension d_A on which the dynamics acts trivially (i.e., the total evolution is given by $\Phi_t \otimes \mathbb{I}_{d_A}$) such that one can witness any non-Markovianity in Φ_t via revivals in the trace distance between initially prepared states.*

For quantum states $d_A = d+1$ [39]. It should be noticed, though, that in order to actually speak about complete positivity one always needs an ancilla of dimension at least d (as complete positivity coincides with the positivity of the map $\Phi \otimes \mathbb{I}_d$). In this way, the trace distance needs one extra dimension than the minimum in order to witness non-Markovianity. A similar result actually holds also for classical systems. In this case, one only needs to enlarge the state space by one extra dimension. Then, for any vector $\boldsymbol{v}$ on the original space, one can always construct a traceless state on this extended space as:

$$\delta\rho = \begin{cases} \delta\rho_i = v_i & \text{if } i \in \{1, \dots, d\}\,; \\ \delta\rho_i = -\sum_{j=1}^{d} \delta\rho_j & \text{if } i \equiv d+1\,. \end{cases} \tag{A.45}$$

This state satisfies $\mathrm{Tr}[|\delta\dot{\rho}|] = \mathrm{Tr}[|\dot{\boldsymbol{v}}|]$, which proves the claim. Still, in this case it should be noticed that this construction cannot be carried

out by adjoining an ancilla, as in that case the minimum dimension of the state space is $d_A = 2$. Nonetheless, this example is useful as it shows that two dimensional ancillas are enough, and that in principle it would be sufficient to embed the dynamics Φ_t in a space only one dimension bigger than the original space in order to witness non-Markovianity.

A.3.1 Relation between trace distance and Fisher information

In this section we present two technical results that are fundamental for the rest of the proofs in this section. First, we show in which cases one can reduce the study of the dynamics of the quantum Fisher information to its classical counterpart; secondly, we define a set of states and perturbations such that the Fisher information actually coincides with the trace distance.

The first result we present is:

Lemma 2. *Consider a state ρ and a perturbation $\delta\rho$ of the form $\delta\rho = \delta\rho_\Delta + \mathrm{d}t\,\delta\rho_C$, where $\delta\rho_\Delta$ contains all the elements on the diagonal of $\delta\rho$, while $\delta\rho_C$ only off-diagonal terms, and $\mathrm{d}t \ll 1$. Then, it follows that:*

$$\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_f^{-1}\big|_\rho[\delta\rho]\right] = \mathrm{Tr}\left[\delta\rho_\Delta\,\mathbb{J}_f^{-1}\big|_\rho[\delta\rho_\Delta]\right] + \mathcal{O}\left(\mathrm{d}t^{\,2}\right) = \tag{A.46}$$

$$= \mathrm{Tr}\left[(\delta\rho_\Delta)^2\,\rho^{-1}\right] + \mathcal{O}\left(\mathrm{d}t^{\,2}\right). \tag{A.47}$$

Moreover, if $\delta\rho = \delta\rho_\Delta$ at time zero, then the derivative of the quantum Fisher information reduces to its classical counterpart:

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_f^{-1}\big|_\rho[\delta\rho]\right] = \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\left[(\delta\rho_\Delta)^2\,\rho^{-1}\right]. \tag{A.48}$$

In this way, as long as only first derivatives are involved, all the quantum Fisher informations collapse on the classical one for commuting ρ and $\rho + \delta\rho$.

Proof. The first part of the result is quite straightforward, since

$$\mathrm{Tr}\left[\delta\rho_\Delta\,\mathbb{J}_f^{-1}\big|_\rho[\delta\rho_C]\right] = \mathrm{Tr}\left[\delta\rho_C\,\mathbb{J}_f^{-1}\big|_\rho[\delta\rho_\Delta]\right] = 0\,, \tag{A.49}$$

thanks to the fact that $\mathbb{J}_f^{-1}\big|_\rho[\delta\rho_\Delta]$ only has diagonal terms, while $\delta\rho_C$ is non zero only away from the diagonal.

The second part of the Lemma is also quite straightforward to prove: denote by $\rho_{\mathrm{d}t} + \delta\rho_{\mathrm{d}t}$ the perturbed state after an infinitesimal evolution of time $\mathrm{d}t$. Then, one can expand $\delta\rho_{\mathrm{d}t}$ in Taylor series to obtain $\delta\rho_{\mathrm{d}t} = \delta\rho_\Delta + \mathrm{d}t\,\delta\dot{\rho}_\Delta + \mathrm{d}t\,\delta\rho_C$, where $\delta\dot{\rho}_\Delta$ is the change along the diagonal, while $\delta\rho_C$ only contains off-diagonal terms (notice that it doesn't matter whether we define the diagonal with respect to time $t = 0$ or $\mathrm{d}t$, as this difference only contributes to order $\mathcal{O}(\mathrm{d}t)$). This implies that the Fisher information reduces to its classical counterpart also for infinitesimal times $\mathrm{d}t$. Hence, the derivative also coincides with the derivative of the classical Fisher metric, proving Eq. (A.49). □

The Lemma above will be used to lift the proof for classical stochastic systems to the quantum regime. Moreover, it also allows us to prove:

Lemma 3. *Given an arbitrary perturbation $\delta\rho$, define the state $\rho_{\delta\rho} = \frac{|\delta\rho|}{\mathrm{Tr}[|\delta\rho|]}$. Then, the square of the trace distance between $\rho_{\delta\rho}$ and $\rho_{\delta\rho} + \delta\rho$ coincides with the Fisher metric at that point, i.e.,*

$$D^2_{\mathrm{Tr}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \mathrm{Tr}\left[|\delta\rho|\right]^2 = \mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho}}[\delta\rho]\right] . \tag{A.50}$$

Moreover, this relation also holds for their first derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\, D^2_{\mathrm{Tr}}(\rho_{\delta\rho}, \rho_{\delta\rho} + \delta\rho) = \frac{\mathrm{d}}{\mathrm{d}t}\, \mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho}}[\delta\rho]\right] . \tag{A.51}$$

Proof. First notice that $[\rho_{\delta\rho}, \delta\rho] = 0$, so that the Fisher information is simply given by:

$$\mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho}}[\delta\rho]\right] = \mathrm{Tr}\left[(\delta\rho)^2\, \rho_{\delta\rho}^{-1}\right] = \mathrm{Tr}\left[|\delta\rho|\right]\, \mathrm{Tr}\left[\frac{(\delta\rho)^2}{|\delta\rho|}\right] = \mathrm{Tr}\left[|\delta\rho|\right]^2 , \tag{A.52}$$

which proves the first part of the claim. Moreover, thanks to Lemma 2 this relation also carries out to the first derivative, as $[\rho_{\delta\rho}, \delta\rho] = 0$. □

The fact that not only one can locally identify the Fisher distance and the trace distance, but also their first derivatives, allows us to lift many of the construction made for the latter to the first, which will prove of key importance in most of our derivations.

A.4 Additional proofs

In the following we provide the proof of Thm. 9-11 from the main text.

A.4.1 Proof of Theorem 9

The statement of the theorem reads:

Theorem. *A classical divisible evolution Φ_t acting on a d-dimensional state space is Markovian if and only if it induces a monotonic decrease in the Fisher information at all times and on the whole domain. In formulae, Markovianity of Φ_t is equivalent to the condition:*

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\Phi_{s,t}(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\Phi_{s,t}(\rho)}[\Phi_{s,t}(\delta\rho)]\right]\Big|_{s=t} \leq 0 \quad \forall\, t\,,\rho\,,\delta\rho\,, \tag{A.53}$$

where ρ and $\rho+\delta\rho$ are two arbitrary close by states in the interior of the state space.

For quantum dynamics the same equivalence holds if one augments the space with a d-dimensional ancillary system on which the dynamics acts trivially, i.e., the evolution on the composite space is given by $\mathbb{I}_d\otimes\Phi_t$.

It should be noticed that if Φ_t is non-Markovian, then there is no guarantee that $\Phi_{s,t}$ maps states to states when acting on points outside of the image of Φ_t. Still, since $\Phi_{t+\mathrm{d}t,\,t} \simeq \mathbb{I} + \mathrm{d}t\ \mathcal{L}_t$, for each point in the interior of the state space there is a $\mathrm{d}t$ small enough so that its image under $\Phi_{t+\mathrm{d}t,\,t}$ is still a state. For this reason, one can define the derivative in Eq. (A.53) without problems when ρ is an interior state.

In the theorem above one side of the implication is trivial, namely the fact that Markovian maps uniformly contracts the Fisher information. In the following we prove the other direction.

Proof (classical case). We need to show that any non-Markovianity is reflected in an increase of Fisher information. First, we can use Eq. (1.278) to express the derivative of the Fisher information, obtaining:

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\Phi_{s,t}(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\Phi_{s,t}(\rho)}[\Phi_{s,t}(\delta\rho)]\right]\Big|_{s=t} = -\sum_{i\neq j} a^{(t)}_{i\leftarrow j}\,\left(\frac{\delta\rho_j}{\rho_j}-\frac{\delta\rho_i}{\rho_i}\right)^2\rho_j\,. \tag{A.54}$$

Assume without loss of generality that $a^{(t)}_{1\leftarrow 2} < 0$. Consider then the state

ρ and the perturbation $\delta\rho$ to be given by:

$$\rho = \begin{pmatrix} \mathcal{O}(\varepsilon) \\ 1-\epsilon \\ \mathcal{O}(\varepsilon) \\ \vdots \\ \mathcal{O}(\varepsilon)\,, \end{pmatrix}, \quad \delta\rho = \begin{pmatrix} \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon) \\ \mathcal{O}(\varepsilon^2) \\ \vdots \\ \mathcal{O}(\varepsilon^2) \end{pmatrix}, \tag{A.55}$$

where ε is an arbitrary small number and we assume that both ρ and $\rho+\delta\rho$ are properly normalised. We also assume, without loss of generality, that both states have strictly positive components, meaning that they sits in the interior of the state space. Now, with this choice, the only terms of $\mathcal{O}(1)$ in the sum in Eq. (A.54) comes from setting $i = 1$ and $j = 2$. Hence, at leading order we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\Phi_{s,t}(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\Phi_{s,t}(\rho)}[\Phi_{s,t}(\delta\rho)]\right]\Big|_{s=t} = -a^{(t)}_{1\leftarrow 2}\,\frac{\delta\rho_1^2}{\rho_1^2} + \mathcal{O}(\varepsilon) > 0\,, \tag{A.56}$$

for ε small enough, as $a^{(t)}_{1\leftarrow 2} < 0$ by assumption. This proves the claim. □

Proof (quantum case). In the case of a quantum map non-Markovianity means that Φ_t is not CP-divisible. This implies that there exists a time t for which the Choi state $\mathbb{I}_d \otimes \Phi_{t+\mathrm{d}t,\,t}[|\Omega\rangle\langle\Omega|]$ develops some negative eigenvalue. Call $|v\rangle$ the corresponding eigenvector. Since $\Phi_{t+\mathrm{d}t,\,t}$ is infinitesimal, $\mathcal{C}^{\Phi_{t+\mathrm{d}t,\,t}}$ must be close to $|\Omega\rangle\langle\Omega|$. But then, in order for $|v\rangle$ to correspond to a negative eigenvalue it must contain a non-zero component $|v_\perp\rangle$ orthogonal to $|\Omega\rangle$. To see that this is true, assume the opposite, i.e., $|v\rangle \equiv |\Omega\rangle$ (as it is parallel to $|\Omega\rangle$ and normalised); then, from perturbation theory we know that the corresponding eigenvalue must be $1 + \mathcal{O}(\mathrm{d}t) > 0$. This contradicts the assumption that $|v\rangle$ is associated to a negative eigenvalue. Hence $\langle v_\perp|\Omega\rangle = 0$.

Consider now the state and the perturbation to be given by:

$$\rho = |\Omega\rangle\langle\Omega|\,; \qquad \delta\rho = \varepsilon\,(|v_\perp\rangle\langle v_\perp| - |\Omega\rangle\langle\Omega|)\,. \tag{A.57}$$

With this choice we have that $[\rho, \delta\rho] = 0$. Moreover, the evolution of $\delta\rho$ through $\Phi_{t+\mathrm{d}t,\,t}$ can only generate an off-diagonal component of order $\mathcal{O}(\mathrm{d}t)$. Hence, we are in the situation of Lemma 2, so that we can neglect any off-diagonal contributions. Then, since all the transitions are along

the diagonal, the dynamics is effectively classical. This brings us back to the proof above, since we know that the rate $a_{v_\perp \leftarrow \psi^+}$ is negative by assumption. Then by padding the rest of the diagonal elements of the state with $\mathcal{O}(\varepsilon)$ terms, and with $\mathcal{O}(\varepsilon^2)$ terms for the perturbation, the resulting ρ and $\delta\rho$ are exactly of the form in Eq. (A.55). Hence, one can carry out the same argument as in the classical case. This concludes the proof. □

A.4.2 Proof of Theorem 10

In Sec. A.3 we showed that the contractivity of the trace distance is not sufficient to imply the Markovianity of a smooth evolution, but still it is possible to supplement the system with a finite number of ancillas which allows to witness any non-Markovian dynamics. In the case of the Fisher information the situation is completely reversed: despite what we showed in Sec. A.4.1, namely that the Markovianity of an evolution is in bijection with the monotonous contractivity of the Fisher distance, we prove here that there is no way of using the latter to witness non-Markovianity in the standard way (that is, by preparing initial states $\{\rho, \rho + \delta\rho\}$ at time $t = 0$ that will experience a local increase in distance at some later time), not even by allowing the use of finite, but arbitrary, number of copies of the dynamics and of ancillas:

Theorem. *Given a divisible evolution Φ_t, no ancillary degree of freedom of finite dimensions or finite number of copy of the dynamics are sufficient to witness all possible non-Markovian evolutions via revivals of the Fisher distance between two initially prepared states.*

This difference in behaviour can be explained by the heavy dependence of the Fisher metric on the base-point. In fact, consider the case in which the first instance of non-Markovianity happens at some time t. In principle, the image of Φ_t might be arbitrarily small, forcing the physically accessible states to be a limited subset of the full space. Whereas this is not a problem for the trace distance, as it is translational invariant, it severely reduces the space of witnesses for the Fisher distance.

Proof (classical case). Consider an initial evolution Φ_t that maps any state into a ball close to some state π (e.g., a map of the form $\Phi_t[\rho] = \pi(1-\varepsilon) + \varepsilon\rho$). Suppose one attaches an arbitrary ancilla of some dimension d_A to the system at time $t = 0$, so that the dynamics is given by the

map $\Phi_t \otimes \mathbb{I}_{d_A}$, and the initial state ρ is part of the tensor product state space. Then, after time t, the global state will be ε-close to:

$$\rho(t) := \Phi_t \otimes \mathbb{I}(\rho) \simeq \pi \otimes \rho_A + \mathcal{O}(\varepsilon), \tag{A.58}$$

where we defined $\rho_A := \mathrm{Tr}_S[\rho]$ to be the reduced state on the ancillary space at time $t = 0$. It should also be noticed that the global rate matrix $\tilde{R}_t$ has the following coordinate expression:

$$\tilde{R}_t = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_t \otimes \mathbb{I}_{d_A} = R_t \otimes \mathbb{I}_{d_A} \quad \Longrightarrow \quad (\tilde{R}_t)_{ij,\alpha\beta} = (R_t)_{ij}\,\delta_{\alpha\beta}, \tag{A.59}$$

where R_t indicates the original rate matrix, while we denote by latin letters the coordinates on the system space, and by greek ones the one for the ancillary system. From this expression we can deduce that the rates for $\tilde{R}_t$ factorise as $\tilde{a}^{(t)}_{i\alpha\leftarrow j\beta} = a^{(t)}_{i\leftarrow j}\delta_{\alpha\beta}$.

Consider now the derivative of the Fisher information at time t between the state $\Phi_t(\rho)$ and $\Phi_t(\rho+\delta\rho)$. Using Eq. (1.278) again we obtain:

$$\dot{\mathcal{F}}_t = -\sum_{i\neq j,\alpha} a^{(t)}_{i\leftarrow j}\left(\frac{(\delta\rho(t))_{i,\alpha}}{(\rho(t))_{i,\alpha}} - \frac{(\delta\rho(t))_{j,\alpha}}{(\rho(t))_{j,\alpha}}\right)^2 (\rho(t))_{j,\alpha} = \tag{A.60}$$

$$= -\sum_{i\neq j,\alpha} a^{(t)}_{i\leftarrow j}\left(\frac{(\delta\rho(t))_{i,\alpha}}{(\rho_A)_\alpha\pi_i} - \frac{(\delta\rho(t))_{j,\alpha}}{(\rho_A)_\alpha\pi_j}\right)^2 (\rho_A)_\alpha\pi_j + \mathcal{O}(\varepsilon). \tag{A.61}$$

The expression above prevents the witnessing of all non-Markovian behaviours. Suppose, in fact, that at time t a single rate becomes negative, for definiteness say $a^{(t)}_{1\leftarrow 2} < 0$. Then, it is sufficient that $a^{(t)}_{2\leftarrow 1}\pi_1 > |a^{(t)}_{1\leftarrow 2}|\,\pi_2$ for the sum in Eq. (A.60) to be strictly negative in the limit $\varepsilon \to 0$. Hence, even if the dynamics is non-Markovian, there is no increase in Fisher distance on the image of Φ_t.

In the same way, even using multiple copies of the channel does not help finding a witness. Consider in fact the dynamics $\tilde{\Phi}_t := \Phi_t^{\otimes n} \otimes \mathbb{I}_{d_A}$ acting on the state space given by n identical copies of the system and by some ancillary space of dimension d_A. The rate matrix in this case is given by

$$\tilde{R}_t = \frac{\mathrm{d}}{\mathrm{d}t}\,\tilde{\Phi}_t = \sum_{l=1}^{n} \mathbb{I}_d^{\otimes(l-1)} \otimes R_t \otimes \mathbb{I}_d^{\otimes(n-l)} \otimes \mathbb{I}_{d_A}. \tag{A.62}$$

Denoting again by latin letters the indexes on the copies of the system space, and by greek letters the indexes of the ancillary space, we can express the rates as

$$\tilde{a}^{(t)}_{(i_1 i_2 \dots i_n,\alpha)\leftarrow(j_1 j_2 \dots j_n,\beta)} = \sum_{l=1}^{n} a^{(t)}_{i_l \leftarrow j_l} \left(\delta_{i_1 j_1} \dots \delta_{i_{l-1} j_{l-1}} \delta_{i_{l+1} j_{l+1}} \dots \delta_{i_n j_n} \delta_{\alpha\beta}\right) . \tag{A.63}$$

In order to shorten the notation, in the following we denote the multi-index $(i_1 i_2 \dots i_n, \alpha)$ by $(\{i_k\}, \alpha)$.

Considering the same initial dynamics Φ_t as in the previous part of the proof, we see that any state gets mapped in the vicinity of:

$$\rho(t) := \Phi_t^{\otimes n} \otimes \mathbb{I}(\rho) \simeq \pi^{\otimes n} \otimes \rho_A + \mathcal{O}(\varepsilon) , \tag{A.64}$$

where again ρ_A is the reduced of the initial state on the ancillary space. Thanks to the expression of the rates in Eq. (A.63), we can express the derivative of the Fisher information as:

$$\dot{\mathcal{F}}_t = \\ = -\sum_{\{i_k\}\neq\{j_k\},\alpha} \tilde{a}^{(t)}_{(\{i_k\},\alpha)\leftarrow(\{j_k\},\beta)} \left(\frac{(\delta\rho(t))_{\{i_k\},\alpha}}{(\rho(t))_{\{i_k\},\alpha}} - \frac{(\delta\rho(t))_{\{j_k\},\alpha}}{(\rho(t))_{\{j_k\},\alpha}}\right)^2 (\rho(t))_{\{j_k\},\alpha} , \tag{A.65}$$

where $(\rho(t))_{\{j_k\},\alpha} \simeq \pi_{j_1} \dots \pi_{j_n} (\rho_A)_\alpha$. Even in this case Eq. (A.65) can stay negative, despite the inset of non-Markovianity. Again, suppose that $a^{(t)}_{1\leftarrow 2}$ is the only negative rate of R_t. Then, the negative contributions to Eq. (A.65) are given by:

$$-\sum_{\{j_k\}} \sum_{l=1}^{n} a^{(t)}_{2\leftarrow 1} \left(\frac{(\delta\rho(t))_{\{j_k\}_{l=2},\alpha}}{(\rho(t))_{\{j_k\}_{l=2},\alpha}} - \frac{(\delta\rho(t))_{\{j_k\}_{l=1},\alpha}}{(\rho(t))_{\{j_k\}_{l=1},\alpha}}\right)^2 (\rho(t))_{\{j_k\}_{l=1},\alpha} , \tag{A.66}$$

where we denoted by $\{j_k\}_{l=x}$ the string of indexes $(j_1, \dots, j_{l-1}, x, j_{l+1}, \dots, j_n)$ and we used the expression of the rates in Eq. (A.63). We can compare the sum above with

$$-\sum_{\{j_k\}} \sum_{l=1}^{n} a^{(t)}_{1\leftarrow 2} \left(\frac{(\delta\rho(t))_{\{j_k\}_{l=1},\alpha}}{(\rho(t))_{\{j_k\}_{l=1},\alpha}} - \frac{(\delta\rho(t))_{\{j_k\}_{l=2},\alpha}}{(\rho(t))_{\{j_k\}_{l=2},\alpha}}\right)^2 (\rho(t))_{\{j_k\}_{l=2},\alpha} , \tag{A.67}$$

which is positive by the assumption that $a^{(t)}_{1\leftarrow 2}$ is the only negative rate. Then, summing up Eq. (A.66) and Eq. (A.67) then turns out to be negative whenever

$$a^{(t)}_{2\leftarrow 1}\left(\rho(t)\right)_{\{j_k\}_{l=1},\alpha} > |a^{(t)}_{1\leftarrow 2}|\left(\rho(t)\right)_{\{j_k\}_{l=2},\alpha}\,, \tag{A.68}$$

i.e., if $a^{(t)}_{2\leftarrow 1}\pi_1 > |a^{(t)}_{1\leftarrow 2}|\,\pi_2$, in complete analogy with what happened in the single copy case. This proves the general claim in Thm. 10. □

Proof (quantum case). We can prove Thm. 10 for quantum dynamics by using again Lemma 2 to reduce it to the classical scenario. Before doing so, though, we need to extend it to the multi-copies scenario with ancillas.

Consider a dynamics Φ_t which moves all the states close to a target one, say π, and then it dephases them in the corresponding eigenbasis. For definiteness, this can be expressed as:

$$\Phi_t[\rho] = \mathcal{D}^{(\varepsilon_2)} \circ \mathcal{F}^{(\varepsilon_1)}[\rho]\,, \tag{A.69}$$

where

$$\mathcal{F}^{(\varepsilon_1)}[\rho] = (1-\varepsilon_1)\pi + \varepsilon_1\rho\,, \tag{A.70}$$

$$\mathcal{D}^{(\varepsilon_2)}[\rho] = (1-\varepsilon_2)\rho_D + \varepsilon_2\rho\,, \quad \rho_D = \sum_i\ |i\rangle\langle i|\,\rho\,|i\rangle\langle i|\ , \tag{A.71}$$

and $|i\rangle$ is an eigenbasis of π (so that $\mathcal{D}$ is the dephasing operator in the basis of π).

We directly study the multiple copies scenario. Consider any initially prepared ρ and $\rho+\delta\rho$ quantum states of $\mathcal{H}^{\otimes n}\otimes\mathbb{C}^{d_A}$, where $\mathcal{H}$ is the d-dimensional state space of the single-copy channel, and we supply a d_A-dimensional ancilla. Then the total evolution is given by:

$$\tilde{\Phi}_t = \Phi_t^{\otimes n}\otimes\mathbb{I}_{d_A}\,, \tag{A.72}$$

where $\mathcal{T}^{(t,0)}$ is as in Eq. (A.69). By defining ρ_A to be the reduced state of ρ on the ancillary degrees of freedom, it is easy to see that at leading order one has

$$\tilde{\Phi}_t[\rho] = \pi^{\otimes n}\otimes\rho_A + \mathcal{O}\left(\varepsilon_1\right). \tag{A.73}$$

Moreover, it also holds that

$$\tilde{\Phi}_t[\delta\rho] = \varepsilon_1\,\delta\rho_D + \mathcal{O}\left(\varepsilon_1\varepsilon_2\right), \tag{A.74}$$

where $\delta\rho_D = \mathcal{D}^{(0)^{\otimes n}} \otimes \mathbb{I}_{d_A}[\delta\rho]$ is the dephased perturbation. As such it can be expressed as

$$\delta\rho_D = \sum_\gamma \theta_D^{(\gamma)} \otimes \omega^{(\gamma)}\,, \tag{A.75}$$

where $\theta_D^{(\gamma)}$ are operators on $\mathcal{H}^{\otimes n}$ satisfying $[\theta_D^{(\gamma)}, \pi^{\otimes n}] = 0$, and $\omega^{(\gamma)}$ are operators on $\mathbb{C}^{d_A}$. The expression of the quantum Fisher information is then given by:

$$\mathrm{Tr}\left[\delta\rho\, \mathbb{J}_f^{-1}\big|_\rho[\delta\rho]\right] = \varepsilon_1^2\, \mathrm{Tr}\left[\delta\rho_D\, \mathbb{J}_f^{-1}\big|_{\pi^{\otimes n}\otimes\rho_A}[\delta\rho_D]\right] + \mathcal{O}\left(\varepsilon_1^2\,(\varepsilon_2+\varepsilon_1)\right). \tag{A.76}$$

If we now use the expression for $\delta\rho_D$ given in Eq. (A.75), we can rewrite the leading term in the above equation as:

$$\varepsilon_1^2 \sum_{\gamma,\gamma'} \mathrm{Tr}\left[\theta_D^{(\gamma)} \mathbb{J}_f^{-1}\big|_{\pi^{\otimes n}}[\theta_D^{(\gamma')}]\right] \mathrm{Tr}\left[\omega^{(\gamma)} \mathbb{J}_f^{-1}\big|_{\rho_A}[\omega^{(\gamma')}]\right] =: \varepsilon_1^2 \sum_{\gamma,\gamma'} \mathcal{M}^{(1)}_{\gamma\gamma'} \mathcal{M}^{(2)}_{\gamma\gamma'}\,, \tag{A.77}$$

where we implicitly defined $\mathcal{M}^{(1)}/\mathcal{M}^{(2)}$ to be the matrix with components given by the first/second trace. It should be noticed that $\mathcal{M}^{(2)}$ is time independent, as the evolution on the ancillary degrees of freedom is trivial. For this reason, the instantaneous derivative of the Fisher information can be rewritten as:

$$\dot{\mathcal{F}}_t = \varepsilon_1^2 \sum_{\gamma,\gamma'} \left(\frac{\mathrm{d}}{\mathrm{d}t}\, \mathcal{M}^{(1)}_{\gamma\gamma'}\right) \mathcal{M}^{(2)}_{\gamma\gamma'}\,, \tag{A.78}$$

that is, as the trace-product of two matrices, $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Since $\mathcal{M}^{(2)}$ is positive definite, it is enough for $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ to be negative definite in order for the product to be ≤ 0.

We can now prove the quantum version of Theorem 10. In fact, we just need to show that there exists a non-Markovian evolution for which $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{M}^{(1)}$ is negative definite. First, notice that $\mathcal{M}^{(1)}$ is a Hermitian matrix, thanks to the self-adjointness of the operator $\mathbb{J}_f$. For this reason, without loss of generality, we can assume it to be given in diagonal form. Consider then the semi-classical Lindbladian with rates $a^{(t)}_{i\leftarrow j}$ and jump operators $|i\rangle\langle j\,|$, i.e.,

$$\mathcal{L}[\rho] = \sum_{i\neq j} a^{(t)}_{i\leftarrow j} \left(|i\rangle\langle j\,|\,\rho\,|j\rangle\langle i\,| - \frac{1}{2}\{\,|j\rangle\langle j\,|\,,\rho\}\right)\,, \tag{A.79}$$

and suppose $a^{(t)}_{1\leftarrow 2} < 0$. Thanks to the condition $[\theta^{(\gamma)}_D, \pi^{\otimes n}] = 0$, each component of $\mathcal{M}^{(1)}$ can be expressed in terms of the classical Fisher information. Hence, thanks to the proof for the classical case, one can always find π such that the Fisher information rates are negative, irrespectively of the expression of the $\theta^{(\gamma)}_D$. This concludes the proof. □

A.4.3 Proof of Theorem 11

Theorem. *Given an evolution Φ_t, for any state ρ and perturbation $\delta\rho$ defined on the system space and on some ancillary degrees of freedom, it is possible to implement a class of transformations $F^{(t)}_{\delta\rho}$ depending on Φ_t and $\delta\rho$ that can be used to witness non-Markovianity at time t by using the Fisher information. This means that if the evolution $\Phi_{t+\mathrm{d}t,t}$ is Markovian, then*

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{F^{(t)}_{\delta\rho}\circ\Phi_s(\rho)}[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)]\right]\Big|_{s=t}\leq 0\,, \qquad \text{(A.80)}$$

whereas in the presence of non-Markovianity there always exists at least one $\delta\rho$ for which the inequality is reversed.

The minimal dimension of the ancilla for classical systems is $d_A = 2$, while for quantum maps one needs $d_A = d + 1$.

Proof. Suppose at time t the divisible dynamics Φ_t is non-Markovian. Then, thanks to Thm. 24, one can always supplement the system with an ancilla of dimension d_A, so that there always exists some traceless $\delta\rho$ such that $\frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho|^2_{\mathrm{Tr}}$ is increasing in time. For classical system $d_A = 2$, while for quantum ones $d_A = d + 1$.

Choose, then, $\delta\rho$ so that it shows backflow in the trace norm at time t. Define now $\rho_{\delta\rho_t} := \frac{|\delta\rho_t|}{\mathrm{Tr}[|\delta\rho_t|]}$, where we used the notation $\delta\rho_t := \Phi_t(\delta\rho)$. Thanks to Lemma 3 we have that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\left[\delta\rho_t\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\delta\rho_t]\right] = \frac{\mathrm{d}}{\mathrm{d}t}\,D^2_{\mathrm{Tr}}(\rho_{\delta\rho_t},\rho_{\delta\rho_t}+\delta\rho_t) > 0\,, \qquad \text{(A.81)}$$

So that the same backflow appears also in the Fisher information. In order to obtain some insights on what's going on, it is useful to repeat

the computation for the derivative here. Doing so, we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Tr}\left[\delta\rho_t\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\delta\rho_t]\right] = \frac{\mathrm{d}}{\mathrm{d}t}\sum_i \frac{(\delta\rho_t)_i^2}{(\rho_{\delta\rho_t})_i} \tag{A.82}$$

$$= \sum_i 2\left(\frac{(\delta\rho_t)_i\,(\delta\dot{\rho}_t)_i}{(\rho_{\delta\rho_t})_i} - \frac{1}{2}\frac{(\delta\rho_t)_i^2\,(\dot{\rho}_{\delta\rho_t})_i}{(\rho_{\delta\rho_t})_i^2}\right). \tag{A.83}$$

It should be noticed that the last term in the sum is zero: in fact, the first term already gives $\frac{\mathrm{d}}{\mathrm{d}t}|\delta\rho|^2_{\mathrm{Tr}}$, so the last term must be zero due to Lemma 3. On the other hand, it is not difficult to carry out the explicit calculation, giving

$$\sum_i \frac{(\delta\rho_t)_i^2\,(\dot{\rho}_{\delta\rho_t})_i}{(\rho_{\delta\rho_t})_i^2} = \left(\sum_i \frac{\cancel{(\delta\rho_t)_i^2}\,(\dot{\rho}_{\delta\rho_t})_i}{\cancel{(\delta\rho_t)_i^2}}\right)\left(\sum_j |(\delta\rho_t)_j|\right)^2 = 0\,, \tag{A.84}$$

where the result is obtained by noticing that $\rho_{\delta\rho_t}$ is normalised, so that $\dot{\rho}_{\delta\rho_t}$ is traceless. This shows that despite the explicit dependence of the Fisher information on $\rho_{\delta\rho_t}$, there is no contribution arising from its derivative in Eq. (A.83). In some sense, the base-point acts as it were frozen, without affecting the ability of the Fisher information to present a backflow.

This intuition inspires the following construction. Define the filtering operator $F^{(t)}_{\delta\rho}$ as

$$F^{(t)}_{\delta\rho}(\sigma) = (1-\varepsilon)\,\rho_{\delta\rho_t} + \varepsilon\,\sigma\,, \tag{A.85}$$

where ε is some arbitrary small parameter. It should be noticed that on traceless states, it acts as:

$$F^{(t)}_{\delta\rho}(\delta\sigma) = F^{(t)}_{\delta\rho}(\sigma+\delta\sigma) - F^{(t)}_{\delta\rho}(\sigma) = \varepsilon\,\delta\sigma\,. \tag{A.86}$$

In particular, the Fisher information between σ and $\sigma+\delta\sigma$ transforms under the application of this filter as

$$\mathrm{Tr}\left[F^{(t)}_{\delta\rho}(\delta\sigma)\,\mathbb{J}_f^{-1}\big|_{F^{(t)}_{\delta\rho}(\sigma)}[F^{(t)}_{\delta\rho}(\delta\sigma)]\right] = \varepsilon^2\,\mathrm{Tr}\left[\delta\sigma\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\delta\sigma]\right] + \mathcal{O}\left(\varepsilon^3\right), \tag{A.87}$$

that is $F^{(t)}_{\delta\rho}$ moves the base-point of the Fisher information to $\rho_{\delta\rho_t}$. For this reason, the filtering defined in Eq. (A.85) allows to witness non-Markovianity. This can be verified by explicitly computing the derivative in Eq. (A.80):

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{F^{(t)}_{\delta\rho}\circ\Phi_s(\rho)}[F^{(t)}_{\delta\rho}\circ\Phi_s(\delta\rho)]\right]\Big|_{s=t}= \tag{A.88}$$

$$=\varepsilon^2\,\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\Phi_s(\delta\rho)\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\Phi_s(\delta\rho)]\right]\Big|_{s=t}= \tag{A.89}$$

$$=\varepsilon^2\,\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{Tr}\left[\delta\rho_s\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\delta\rho_s]\right]\Big|_{s=t}= \tag{A.90}$$

$$=\varepsilon^2\,\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\left[\delta\rho_t\,\mathbb{J}_f^{-1}\big|_{\rho_{\delta\rho_t}}[\delta\rho_t]\right]=\varepsilon^2\,\frac{\mathrm{d}}{\mathrm{d}t}\,D^2_{\mathrm{Tr}}(\rho_{\delta\rho_t},\rho_{\delta\rho_t}+\delta\rho_t)\,, \tag{A.91}$$

where we neglected terms of order $\mathcal{O}\left(\varepsilon^3\right)$, and in the last line we added a time dependence on the base-point, thanks to the considerations around Eq. (A.84). For this reason, the quantity in Eq. (A.85) has the same behaviour as the trace distance: it contracts for any Markovian dynamics, while whenever this condition is not satisfied one can always find a $\delta\rho$ that witness the inset of non-Markovianity. □

A.5 Detailed balance in the quantum regime

In this appendix we will first prove Thm. 14 and then we will provide the structural characterisation of Fisher detailed balance Lindbladians presented in Eq. (1.343).

Due to the amount of different notions of adjoints used in the following, we remind the reader about the notation used. There are three different scalar product used, namely the Hilbert-Schmidt one, K^o_π and $K_{f,\pi}$, which induce the following adjoints:

$$\text{Hilbert-Schmidt:}\qquad \mathrm{Tr}\left[AX(B)\right]=\mathrm{Tr}\left[X^\dagger(A)B\right]\,; \tag{A.92}$$

$$K^o_\pi:\qquad \mathrm{Tr}\left[AO(B)\pi\right]=\mathrm{Tr}\left[\widetilde{O}^o(A)B\pi\right]\,; \tag{A.93}$$

$$K_{f,\pi}:\qquad \mathrm{Tr}\left[A\mathbb{J}_f^{-1}[\mathcal{O}B]\right]=\mathrm{Tr}\left[(\widetilde{\mathcal{O}}_f A)\mathbb{J}_f^{-1}[B]\right]\,, \tag{A.94}$$

where O and $\mathcal{O}$ are a superoperator on the space of observables or on the state space, respectively, while X is a generic bounded operator. As it is explained in the main text, one can formulate the adjoint with respect

to K^o_π or $K_{f,\pi}$ in terms of the Hilbert-Schmidt adjoint, thanks to the relation:

$$\widetilde{O}^o = \mathbb{R}_\pi^{-1} \circ O^\dagger \circ \mathbb{R}_\pi \,; \tag{A.95}$$

$$\widetilde{\mathcal{O}}_f = \mathbb{J}_f\big|_\pi \circ \mathcal{O}^\dagger \circ \mathbb{J}_f^{-1}\big|_\pi \,, \tag{A.96}$$

as it can be verified directly from the definition. From this relations it is easy to see that the self-adjoint condition corresponds to the two equality $\mathbb{R}_\pi \circ O = O^\dagger \circ \mathbb{R}_\pi$ in the K^o_π case, and $\mathcal{O} \circ \mathbb{J}_f\big|_\pi = \mathbb{J}_f\big|_\pi \circ \mathcal{O}^\dagger$ for the Fisher one. Thanks to this characterisation we can prove the following useful result:

Lemma 4. *Suppose $\mathcal{O}$ and O are adjoint of each other, $\mathcal{O}^\dagger \equiv O$. Then, if $[\mathcal{O}, \mathbb{L}_\pi \mathbb{R}_\pi^{-1}] = 0$, the two conditions of self-adjointness and skew-self-adjointness with respect to $K_{f,\pi}$ and K^o_π coincide.*

Proof. First of all, it should be noticed that both the auto-modular operator $(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})^\dagger = \mathbb{L}_\pi^{-1}\mathbb{R}_\pi = (\mathbb{L}_\pi \mathbb{R}_\pi^{-1})^{-1}$ and $(\mathbb{R}_\pi)^\dagger = \mathbb{L}_\pi$. Using these property we can show that the (skew-)self-adjointness with respect to $K_{f,\pi}$ is equivalent to the corresponding notion for K^o_π. In fact, the following relations are equivalent:

$$\widetilde{\mathcal{O}}_f = \pm\mathcal{O} \iff (\mathcal{O}\mathbb{J}_f\big|_\pi)^\dagger = \pm\,\mathcal{O}\mathbb{J}_f\big|_\pi \iff \tag{A.97}$$

$$\iff f((\mathbb{L}_\pi \mathbb{R}_\pi^{-1})^{-1})\mathbb{L}_\pi O = \pm\,\mathcal{O}\mathbb{R}_\pi f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1}) \iff \tag{A.98}$$

$$\iff \cancel{f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})}\mathbb{R}_\pi O = \pm\cancel{f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})}O^\dagger \mathbb{R}_\pi \iff \widetilde{O}^o = \pm O \,, \tag{A.99}$$

where in the last line we used the property $f(x) = xf(x^{-1})$, together with the commutation between $\mathcal{O}$ and $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$ to push $f(\mathbb{L}_\pi \mathbb{R}_\pi^{-1})$ to the left of both equations. Finally, the last step is simply the definition of $\mathcal{O}^\dagger \equiv O$. □

This lemma is particularly useful because it allows to reduce the question about the equivalence of Def. 1 and Def. 3 to the decision about the commutation of the Lindbladian with the auto-modular operator $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$.

A.5.1 Proof of Theorem 14

The aim of this section is to prove Thm. 14:

Theorem. *The following conditions are equivalent:*

1. *the generator of the dynamics in the Heisenberg picture $\mathcal{L}^\dagger$ satisfies the adjointness relations in Def. 1*

2. *the Lindbladian $\mathcal{L}$ satisfies the structural characterisation in Def. 2.*

These conditions imply:

3. *the generator of the dynamics in the Schroedinger picture $\mathcal{L}$ satisfies the adjointness relations in Def. 3.*

Moreover, if the Hamiltonian H is non-degenerate the three conditions are equivalent.

First, it should be noticed that the equivalence between condition 1 and 2 was already proven by Alicki in [41], so we postpone the proof to App. A.5.3 where the characterisation of Lindbladians satisfying condition 3 is given. In the particular case in which H is non-degenerate, this will provide a proof of the structural definition of condition 2.

Proof. First it should be noticed that if condition 2 is satisfied, the Lindbladian commutes with the auto-modular operator $\mathbb{L}_\pi\mathbb{R}_\pi^{-1}$. In fact, starting from the characterisation:

$$\mathcal{L}(\rho) = -i[H,\rho] + \sum_{\omega,i} \lambda_i^\omega \left(A_i^\omega \, \rho \, (A_i^\omega)^\dagger - \frac{1}{2}\{(A_i^\omega)^\dagger A_i^\omega, \rho\} \right) , \quad \text{(A.100)}$$

it is a matter of straightforward calculations to verify that:

$$\mathbb{L}_\pi\mathbb{R}_\pi^{-1}(\mathcal{L}(\rho)) = -i\pi[H,\rho]\pi^{-1} +$$
$$+ \sum_{\omega,i} \lambda_i^\omega \left((\pi \, A_i^\omega \pi^{-1}) \, \pi \, \rho \, \pi^{-1} (\pi^{-1} A_i^\omega \, \pi)^\dagger - \frac{1}{2}\pi\{(A_i^\omega)^\dagger A_i^\omega, \rho\}\pi^{-1} \right) = \quad \text{(A.101)}$$
$$= -i[H, \pi\rho\pi^{-1}] +$$
$$+ \sum_{\omega,i} \lambda_i^\omega \left(\cancel{e^\omega e^{-\omega}} \, A_i^\omega \, \pi \, \rho \, \pi^{-1} (A_i^\omega)^\dagger - \frac{1}{2}\{(A_i^\omega)^\dagger A_i^\omega, \pi\,\rho\pi^{-1}\} \right) = \quad \text{(A.102)}$$
$$= \mathcal{L}(\mathbb{L}_\pi\mathbb{R}_\pi^{-1}(\rho)) \quad \text{(A.103)}$$

where we used the condition $[H,\pi] = 0$, together with $\mathbb{L}_\pi\mathbb{R}_\pi^{-1}(A_i^\omega) = e^\omega \, A_i^\omega$ and $(A_i^\omega)^\dagger = A_i^{-\omega}$. Since condition 2 is equivalent to condition 1, and $[\mathcal{L}, \mathbb{L}_\pi\mathbb{R}_\pi^{-1}] = 0$, thanks to Lemma 4 this means that $\mathcal{L}$ has the same

self-adjointness properties with respect to K_π^o and $K_{f,\pi}$. This proves the forward implication.

Let us prove the reverse, namely that condition 3 is equivalent to condition 1 for non-degenerate Hamiltonians. Let us first focus on the unitary part. Then, using $\mathcal{U}^\dagger = -\mathcal{U}$ we can rewrite the skew-adjointness condition as:

$$\mathcal{U} \circ \mathbb{J}_f\big|_\pi = -\mathbb{J}_f\big|_\pi \circ \mathcal{U}^\dagger \qquad \Longleftrightarrow \qquad \mathcal{U} \circ \mathbb{J}_f\big|_\pi = \mathbb{J}_f\big|_\pi \circ \mathcal{U}\,. \tag{A.104}$$

Then, applying the two operators in the last equation to the identity, we can verify that:

$$\mathcal{U} \circ \mathbb{J}_f\big|_\pi(\mathbb{1}) = \mathbb{J}_f\big|_\pi \circ \mathcal{U}(\mathbb{1}) \implies \mathcal{U}(\pi) = -i\,\mathbb{J}_f\big|_\pi([H,\mathbb{1}]) \implies \tag{A.105}$$

$$\implies [H,\pi] = 0\,. \tag{A.106}$$

Notice that this result is generic, i.e., no assumptions on the spectrum of H need to be made. This directly implies the commutation $[\mathcal{U}, \mathbb{L}_\pi \mathbb{R}_\pi^{-1}] = 0$, so again thanks to Lemma 4 we have that for the unitary part condition 3 and 1 are always equivalent.

Let us now focus on Eq. (A.106). Since H and π commute, we can find a common set of eigenvectors, which we denote by $\{|\alpha\rangle\}$. Moreover, the basis $\{|\alpha\rangle\langle\beta|\}$ gives a set of common eigenvectors to $\mathcal{U}$ and $\mathbb{L}_\pi\mathbb{R}_\pi^{-1}$, as it can be verified by direct calculation:

$$\mathcal{U}(|\alpha\rangle\langle\beta|) = -i(H_\alpha - H_\beta)\,|\alpha\rangle\langle\beta|\,; \tag{A.107}$$

$$\mathbb{L}_\pi\mathbb{R}_\pi^{-1}(|\alpha\rangle\langle\beta|) = \frac{\pi_\alpha}{\pi_\beta}\,|\alpha\rangle\langle\beta|\,. \tag{A.108}$$

Both superoperators have constant eigenvalues for all eigenvectors of the form $|\alpha\rangle\langle\alpha|$. Under the assumption of continuity under small perturbations, we can also assume that each eigenvector such that $\alpha \neq \beta$ has a different eigenvalue (non-degenerate gap condition). These two observations together then imply that any superoperator commuting with $\mathcal{U}$ needs to commute with $\mathbb{L}_\pi\mathbb{R}_\pi^{-1}$ as well. In the main text we showed how normality of $\mathcal{L}$ implies $[\mathcal{U}, \mathcal{L}_\mathcal{D}] = 0$. Then, thanks to the considerations above, we also have that:

$$[\mathcal{L}_\mathcal{D}, \mathbb{L}_\pi\mathbb{R}_\pi^{-1}] = 0\,. \tag{A.109}$$

Hence we can use Lemma 4, which implies that if $(\widetilde{\mathcal{L}_\mathcal{D}})_f = \mathcal{L}_\mathcal{D}$, we also have that $(\widetilde{\mathcal{L}_\mathcal{D}^\dagger})^o = \mathcal{L}_\mathcal{D}^\dagger$. This concludes the proof. □

It should be noticed that the non-degeneracy of the spectrum is needed to prove the commutation relation in Eq. (A.109). The same equivalence can be proven if $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$ has some functional dependence on $\mathcal{U}$. Take for example the thermal scenario, i.e., $\pi \propto \exp[-\beta H]$. Then the auto-modular operator takes the form $\mathbb{L}_\pi \mathbb{R}_\pi^{-1} = \exp[-i\beta\mathcal{U}]$. Due to normality of the Lindbladian (which implies $[\mathcal{U}, \mathcal{L}_\mathcal{D}] = 0$), we directly obtain Eq. (A.109), proving the equivalence without further assumptions on the spectrum of π or H.

A.5.2 Def. 3 is weaker in general

Whereas the constraint coming from Def. 1 imply the one in Def. 3, the reverse does not hold in general. In fact, this is connected with the commutation between $\mathcal{L}_\mathcal{D}$ and $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$. Whereas in the first definition of detailed balance these two operators always commute, this is not the case for the Fisher detailed balance dissipators. This leads to a less constrained evolution of the coherences, as it will be shown in the following.

First, as it was discussed in the proof in the previous section, if $\{|\alpha\rangle\}$ is an eigenbasis for π, then $\{|\alpha\rangle\langle\beta|\}$ are eigenvectors for $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$, with eigenvalues:

$$\mathbb{L}_\pi \mathbb{R}_\pi^{-1}(|\alpha\rangle\langle\beta|) = \frac{\pi_\alpha}{\pi_\beta} |\alpha\rangle\langle\beta| \,. \tag{A.110}$$

Since the steady state π is always assumed to be full rank, proving that $\mathcal{L}_\mathcal{D}$ commutes with $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$ is equivalent to requiring that the matrix elements

$$(\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta} \equiv \langle\gamma| \mathcal{L}_\mathcal{D}(|\alpha\rangle\langle\beta|) |\delta\rangle \tag{A.111}$$

satisfy the following condition

$$[\mathcal{L}_\mathcal{D}, \Phi_\pi] = 0 \quad \iff \quad (\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta} \left(\frac{\pi_\alpha}{\pi_\beta} - \frac{\pi_\gamma}{\pi_\delta} \right) = 0 \,. \tag{A.112}$$

Equivalently, this means that matrix elements of $\mathcal{D}$ can be nonzero only if:

$$(\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta} \neq 0 \quad \implies \quad \frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\delta} \,. \tag{A.113}$$

In the following we show that a slightly more general condition follows from the requirement that $(\widetilde{\mathcal{L}_\mathcal{D}})_f = \mathcal{L}_\mathcal{D}$ for all standard monotone

functions f. Indeed, this condition can be written in coordinates as:

$$\mathcal{L}_{\mathcal{D}} \circ \mathbb{J}_f\big|_{\pi} = \mathbb{J}_f\big|_{\pi} \circ \mathcal{L}_{\mathcal{D}}^{\dagger} \quad \Longleftrightarrow \tag{A.114}$$

$$\Longleftrightarrow \quad (\mathcal{L}_{\mathcal{D}})_{\delta|\beta}^{\gamma|\alpha} f\left(\frac{\pi_\alpha}{\pi_\beta}\right)\pi_\beta = (\mathcal{L}_{\mathcal{D}})_{\alpha|\gamma}^{\beta|\delta} f\left(\frac{\pi_\gamma}{\pi_\delta}\right)\pi_\delta\,. \tag{A.115}$$

At this point it is useful to introduce the notation $e^{-\omega_1} := \pi_\alpha/\pi_\beta$, and $e^{-\omega_2} := \pi_\gamma/\pi_\delta$. Grouping the functional dependence on one side of the equation we obtain:

$$\frac{(\mathcal{L}_{\mathcal{D}})_{\delta|\beta}^{\gamma|\alpha}\,\pi_\beta}{(\mathcal{L}_{\mathcal{D}})_{\alpha|\gamma}^{\beta|\delta}\,\pi_\delta} = \frac{f\left(e^{-\omega_2}\right)}{f\left(e^{-\omega_1}\right)}\,. \tag{A.116}$$

It should be noticed that the left hand side of the equation does not depend on the function f, so the coordinates of $\mathcal{L}_{\mathcal{D}}$ are zero unless $\omega_1 = \pm\omega_2$. Notice that the second condition follows from the symmetry of standard monotone functions $f(x) = x\,f(x^{-1})$.

Hence, the only non-zero elements of $\mathcal{L}_{\mathcal{D}}$ are satisfy either of the conditions:

$$(\mathcal{L}_{\mathcal{D}})_{\delta|\beta}^{\gamma|\alpha} \neq 0 \implies \left(\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\delta}\right) \vee \left(\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\delta}{\pi_\gamma}\right)\,. \tag{A.117}$$

Comparing this result with Eq. (A.113) directly shows that $\mathcal{L}_{\mathcal{D}}$ does not commute in general with $\mathbb{L}_\pi\mathbb{R}_\pi^{-1}$, so Def. 3 is weaker than Def. 1 (at the end of the section we present an explicit example showing this). For this reason, it is interesting to explore which constraints Eq. (A.117) imposes on the Lindbladian. First, it should be noticed that since $\mathcal{L}_{\mathcal{D}}$ is adjoint preserving, its coordinates satisfy

$$\mathcal{L}_{\mathcal{D}}(A^\dagger) = \mathcal{L}_{\mathcal{D}}(A)^\dagger \quad \Longleftrightarrow \quad (\mathcal{L}_{\mathcal{D}})_{\delta|\beta}^{\gamma|\alpha} = (\bar{\mathcal{L}}_{\mathcal{D}})_{\gamma|\alpha}^{\delta|\beta}\,. \tag{A.118}$$

Then, combining Eq. (A.116), Eq. (A.117), and Eq. (A.118), we can see that:

- populations and coherences do not mix, and the populations on the diagonal satisfy the classical detailed balance. In fact, from Eq. (A.117) we see that from $\alpha = \beta$ it follows that $\gamma = \delta$ (assuming π to be non-degenerate), and the currents are related by the standard detailed balance condition:

$$(\mathcal{L}_{\mathcal{D}})_{\gamma|\alpha}^{\gamma|\alpha}\,\pi_\alpha = (\mathcal{L}_{\mathcal{D}})_{\alpha|\gamma}^{\alpha|\gamma}\,\pi_\gamma\,. \tag{A.119}$$

- The dynamics of the coherences can be split in two cases: the one for which $\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\delta}$, implying the following relation:

$$(\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta}\,\pi_\beta = (\mathcal{L}_\mathcal{D})^{\beta|\delta}_{\alpha|\gamma}\,\pi_\delta = (\bar{\mathcal{L}}_\mathcal{D})^{\alpha|\gamma}_{\beta|\delta}\,\pi_\delta\,. \tag{A.120}$$

  This property is satisfied also in the Alicki's definition of detailed balance.

- Moreover, one has additional transitions between coherences with characterised by the relation $\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\delta}{\pi_\gamma}$. In this case the rate are given by:

$$(\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta}\,\pi_\beta = (\mathcal{L}_\mathcal{D})^{\beta|\delta}_{\alpha|\gamma}\,\pi_\gamma = (\bar{\mathcal{L}}_\mathcal{D})^{\alpha|\gamma}_{\beta|\delta}\,\pi_\gamma\,. \tag{A.121}$$

  These rates are the only novelty compared with the one coming from Def. 1, and are the cause of the failure of $\mathcal{L}_\mathcal{D}$ from commuting with $\mathbb{L}_\pi\mathbb{R}_\pi$.

In order to justify the preference for Def. 3 as the quantum generalisation of detailed balance, we argue here that this last case is still physically sensible. Consider indeed two coherences terms $|\alpha\rangle\langle\beta|$ and $|\gamma\rangle\langle\delta|$ such that $\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\delta}$. Then, from Eq. (A.120) it follows that the ratio between the currents induced between the two coherences is given by:

$$\frac{\left|(\mathcal{L}_\mathcal{D})^{\gamma|\alpha}_{\delta|\beta}\right|}{\left|(\bar{\mathcal{L}}_\mathcal{D})^{\alpha|\gamma}_{\beta|\delta}\right|} = \frac{\pi_\delta}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\alpha}\,. \tag{A.122}$$

The additional freedom given by Eq. (A.121) corresponds to the possibility of the matrix element $(\mathcal{L}_\mathcal{D})^{\delta|\alpha}_{\gamma|\beta}$ to be non-zero. It should be noticed, though, that the current between the two coherences is consistent with Eq. (A.122):

$$\frac{\left|(\mathcal{L}_\mathcal{D})^{\delta|\alpha}_{\gamma|\beta}\right|}{\left|(\bar{\mathcal{L}}_\mathcal{D})^{\alpha|\delta}_{\beta|\gamma}\right|} = \frac{\pi_\delta}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\alpha}\,. \tag{A.123}$$

Thus, the difference between Def. 1 and the Fisher one (i.e., Def. 3), is that the first allows the coherences $|\alpha\rangle\langle\beta|$ and $|\gamma\rangle\langle\delta|$ to communicate but prohibits interaction between $|\alpha\rangle\langle\beta|$ and $|\delta\rangle\langle\gamma|$, while the latter allows the dynamics to connect both off-diagonal elements, while keeping the ratio between the two currents detailed balanced.

Finally, before moving on to the characterisation of the Lindbladian operators satisfying Def. 3 we present here an example of $\mathcal{L}$ which is detailed balance in the Fisher sense, but not according to Def. 1. Consider a two-levels system thermalizing to the Gibbs state

$$\pi_\beta = \frac{|0\rangle\langle 0| + e^{-\beta}\,|1\rangle\langle 1|}{1+e^{-\beta}}\,. \tag{A.124}$$

We consider the Hamiltonian of the system to be completely degenerate, i.e., $H \propto \mathbb{1}$, which implies $\mathcal{U} = 0$. Then, it is easy to see that the Lindbladian

$$\mathcal{L}(\rho) = A\rho A^\dagger - \frac{1}{2}\{\rho, A^\dagger A\}\,, \tag{A.125}$$

with the jump operator given by $A = |0\rangle\langle 1| + \sqrt{e^{-\beta}}\,|1\rangle\langle 0|$ satisfies $\widetilde{\mathcal{L}}_f = \mathcal{L}$, but not $\widetilde{\mathcal{L}}^o = \mathcal{L}$.

A.5.3 Structural characterisation of Def. 3

This section is devoted to the derivation of the structural form of a detailed balanced Lindbladian according to the Def. 3. Restricting the derivation to the case in which $\mathcal{L}$ commutes with $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$ also proves the equivalence between Def. 1 and Def. 2.

With the hindsight of the previous section it is useful to expand the Lindbladian operator in terms of the following eigenbasis:

$$F_\alpha^\omega := \left\{ |\gamma\rangle\langle\alpha| \;\middle|\; \frac{\pi_\gamma}{\pi_\alpha} = e^\omega \right\}. \tag{A.126}$$

It is also useful to introduce the function $\beta_\alpha(\omega) := \{\beta|\,\pi_\beta = \pi_\alpha e^\omega\}$, namely a function that returns the index β such that $\frac{\pi_\beta}{\pi_\alpha} = e^\omega$. In order to keep the notation clear we also define $\gamma_\alpha(\omega)$ and $\delta_\alpha(\omega)$ exactly in the same way. The elements of the eigenbasis in Eq. (A.126) have the property that:

$$(F_\alpha^\omega)^\dagger = F_{\gamma_\alpha(\omega)}^{-\omega}\,; \tag{A.127}$$

$$\pi F_\alpha^\omega = e^\omega\, F_\alpha^\omega\, \pi\,. \tag{A.128}$$

Generically, one can express the action of the dissipator as:

$$\mathcal{L}_{\mathcal{D}}(\rho) := \sum_{\substack{\alpha,\beta,\\ \gamma,\delta}} (\mathcal{L}_{\mathcal{D}})^{\gamma|\alpha}_{\delta|\beta} \ |\gamma\rangle\langle\alpha| \, \rho \, (|\delta\rangle\langle\beta|)^\dagger = \tag{A.129}$$

$$= \sum_{\substack{\alpha,\omega,\\ \omega_1,\omega_2}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega+\omega_2)|\beta_\alpha(\omega_1)} \, F^\omega_\alpha \, \rho \, (F^{\omega-\omega_1+\omega_2}_{\beta_\alpha(\omega_1)})^\dagger \,, \tag{A.130}$$

where we implicitly defined $\pi_\gamma/\pi_\alpha =: e^{\omega}$, $\pi_\beta/\pi_\alpha =: e^{\omega_1}$ and $\pi_\delta/\pi_\gamma =: e^{\omega_2}$. This expression is particularly useful because it allows for a straightforward application of the constraints in Eq. (A.117). In fact, we have that:

$$\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\gamma}{\pi_\delta} \qquad \Longleftrightarrow \qquad \omega_1 = \omega_2 \,; \tag{A.131}$$

$$\frac{\pi_\alpha}{\pi_\beta} = \frac{\pi_\delta}{\pi_\gamma} \qquad \Longleftrightarrow \qquad \omega_1 = -\omega_2 \,. \tag{A.132}$$

Hence, the sum above can be restricted to the case $\omega_1 = \pm\omega_2$, giving:

$$\mathcal{L}_{\mathcal{D}}(\rho) = \sum_{\substack{\alpha,\omega\\ \omega_1}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega+\omega_1)|\beta_\alpha(\omega_1)} \, F^\omega_\alpha \, \rho \, (F^\omega_{\beta_\alpha(\omega_1)})^\dagger + \tag{A.133}$$

$$+ \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega-\omega_1)|\beta_\alpha(\omega_1)} \, F^\omega_\alpha \, \rho \, (F^{\omega-2\omega_1}_{\beta_\alpha(\omega_1)})^\dagger \,. \tag{A.134}$$

It should be noticed that the case $\omega_1 = 0$ is included in the first sum, so that we have to impose the constraint $\omega_1 \neq 0$ in the second line. In this way $\mathcal{L}_{\mathcal{D}}$ naturally splits in two parts, $\mathcal{L}_{\mathcal{D}_1}$ (i.e., the operator in the first line) corresponding to the component of the dissipator commuting with $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$, and $\mathcal{L}_{\mathcal{D}_2}$, given by the sum in the second line. Interestingly, the condition $\widetilde{\mathcal{L}}^o_{\mathcal{D}} = \mathcal{L}_{\mathcal{D}}$ implies $\mathcal{L}_{\mathcal{D}} \equiv \mathcal{L}_{\mathcal{D}_1}$, so characterising the latter provides the structural form of detailed balance Lindbladian according to Def. 1.

Whereas $\mathcal{L}_{\mathcal{D}_1}$ directly commutes with $\mathbb{L}_\pi \mathbb{R}_\pi^{-1}$, we can apply a transformation to $\mathcal{L}_{\mathcal{D}_2}$ to make it commuting. In particular, it should be noticed

that $\mathcal{L}_{\mathcal{D}_2}$ transforms under the transposition superoperator T as:

$$T\,\mathcal{L}_{\mathcal{D}_2}(\rho) = \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega-\omega_1)|\beta_\alpha(\omega_1)} \left(F^\omega_\alpha\,\rho\,(F^{\omega-2\omega_1}_{\beta_\alpha(\omega_1)})^\dagger\right)^T = \tag{A.135}$$

$$= \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega-\omega_1)|\beta_\alpha(\omega_1)} \left|\pi_\alpha e^{\omega-\omega_1}\right\rangle\!\left\langle\pi_\alpha\right|\rho\left|\pi_\alpha e^{\omega_1}\right\rangle\!\left\langle\pi_\alpha e^{\omega}\right| = \tag{A.136}$$

$$= \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega-\omega_1)|\beta_\alpha(\omega_1)}\, F^{\omega-\omega_1}_\alpha\,\rho\,(F^{\omega-\omega_1}_{\beta_\alpha(\omega_1)})^\dagger = \tag{A.137}$$

$$= \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega+\omega_1)|\alpha}_{\delta_\alpha(\omega)|\beta_\alpha(\omega_1)}\, F^{\omega}_\alpha\,\rho\,(F^{\omega}_{\beta_\alpha(\omega_1)})^\dagger\,, \tag{A.138}$$

where in the second line we used the abuse of notation $|\pi_\alpha\rangle$ for $|\alpha\rangle$, and in the last line we made an implicit change of variables. From Eq. (A.138) we can see that $T\mathcal{L}_{\mathcal{D}_2}$ takes a form completely analogous to $\mathcal{L}_{\mathcal{D}_1}$. For this reason, it directly follows that:

$$[T\mathcal{L}_{\mathcal{D}_2}, \mathbb{L}_\pi\mathbb{R}_\pi^{-1}] = 0\,. \tag{A.139}$$

This condition allows to lift the characterisation for $\mathcal{L}_{\mathcal{D}_1}$ to $\mathcal{L}_{\mathcal{D}_2}$, as it will be shown in the following.

We begin by studying $\mathcal{L}_{\mathcal{D}_1}$. Notice that the coordinates of $\mathcal{L}_{\mathcal{D}}$ are related by:

$$(\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\alpha(\omega+\omega_1)|\beta_\alpha(\omega_1)} = e^\omega\,(\mathcal{L}_{\mathcal{D}})^{\beta_\alpha(\omega_1)|\delta_\alpha(\omega+\omega_1)}_{\alpha|\gamma_\alpha(\omega)}\,, \tag{A.140}$$

as it can be verified from Eq. (A.116). Then, we can diagonalise $\mathcal{L}_{\mathcal{D}_1}$ in a way completely analogous to the procedure carried out in Sec. 1.5.1. To this end, it is useful to introduce a new basis of operators, given by $\{X^\omega_m\} = \{\Delta_i\}_{1\leq i\leq d}\cup\{F^\omega_\alpha\}_{\omega\neq 0}$, where $\{\Delta_i\}_{1\leq i\leq d}$ is an orthonormal basis for the diagonal matrices, and $\Delta_1 = \mathbb{1}/\sqrt{d}$. Then, we can rewrite $\mathcal{L}_{\mathcal{D}_1}$ in this basis as:

$$\mathcal{L}_{\mathcal{D}_1}(\rho) = \sum_{\alpha,\beta} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(0)|\alpha}_{\delta_\beta(0)|\beta}\, F^0_\alpha\,\rho\,(F^0_\beta)^\dagger + \sum_{\substack{\alpha,\beta\\ \omega\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\beta(\omega)|\beta}\, F^\omega_\alpha\,\rho\,(F^\omega_\beta)^\dagger = \tag{A.141}$$

$$= \sum_{i,j} D^0_{i,j}\,\Delta_i\,\rho\,(\Delta_j)^\dagger + \sum_{\substack{\alpha,\beta,\\ \omega\neq 0}} D^\omega_{\alpha,\beta}\,F^\omega_\alpha\,\rho\,(F^\omega_\beta)^\dagger\,, \tag{A.142}$$

where we introduced the coefficients $D^{\omega}_{\alpha,\beta} := (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\beta(\omega)|\beta}$ and

$$D^0_{i,j} = \sum_{\alpha,\beta} (\mathcal{L}_{\mathcal{D}})^{\alpha|\alpha}_{\beta|\beta}\, \bar{U}_{\alpha,i}\, U_{j,\beta}\,, \tag{A.143}$$

where U is the unitary defined by $\Delta_i := U_{i,\alpha} F_\alpha$. In order to make the dissipator explicitly trace preserving, we highlight the terms containing the identity operator. In analogy to Sec. 1.5.1 this leads to the form:

$$\mathcal{L}_{\mathcal{D}_1}(\rho) = \{K, \rho\} + \sum_{\substack{i,j\neq(1,1)\\ \omega}} D^{\omega}_{i,j}\; X^{\omega}_i\, \rho\, (X^{\omega}_j)^\dagger\,, \tag{A.144}$$

and $K := \frac{D^0_{1,1}}{2d}\mathbb{1} + \frac{1}{\sqrt{d}}\sum_i D^0_{i,1}\Delta_i$. The absence of the Hamiltonian contribution follows from the fact that we are looking at the dissipator alone. Equivalently, it is not hard to show that all the $D^0_{i,j}$ are indeed real.

It should be noticed that $\mathcal{L}_{\mathcal{D}_2}$ only gives contributions out of the diagonal, so one can impose trace preservation on $\mathcal{L}_{\mathcal{D}_1}$ alone, i.e., $\mathcal{L}^\dagger_{\mathcal{D}_1}(\mathbb{1}) = 0$. This directly implies that $K = -\frac{1}{2}\left(\sum_{i,j\neq(1,1)} D^{\omega}_{i,j}(X^{\omega}_j)^\dagger X^{\omega}_i\right)$, so the dissipator can be rewritten in Lindblad form:

$$\mathcal{L}_{\mathcal{D}_1}(\rho) = \sum_{\substack{i,j\neq(1,1)\\ \omega}} D^{\omega}_{i,j}\; \left(X^{\omega}_i\, \rho\, (X^{\omega}_j)^\dagger - \frac{1}{2}\left\{(X^{\omega}_j)^\dagger X^{\omega}_i, \rho\right\}\right)\,. \tag{A.145}$$

The property of the dissipator of being adjoint preserving implies that the matrix $D^{\omega}_{i,j}$ is Hermitian. Then, there exists a unitary matrix V, such that $D^{\omega}_{i,j} = V^{\omega}_{i,m}(\lambda^{\omega}_m \delta^m_n)(V^{\omega})^\dagger_{n,j}$, where we sum over repeated indexes. Hence, define the jump operators $A^{\omega}_m := X^{\omega}_i V^{\omega}_{i,m}$. Then, Eq. (A.145) becomes:

$$\mathcal{L}_{\mathcal{D}_1}(\rho) = \sum_{\substack{i,j\neq(1,1)\\ \omega}} \lambda^{\omega}_m\; V^{\omega}_{i,m}(V^{\omega})^\dagger_{m,j}\left(X^{\omega}_i\, \rho\, (X^{\omega}_j)^\dagger - \frac{1}{2}\left\{(X^{\omega}_j)^\dagger X^{\omega}_i, \rho\right\}\right) = \tag{A.146}$$

$$= \sum_{m,\omega} \lambda^{\omega}_m\; \left(A^{\omega}_m\, \rho\, (A^{\omega}_m)^\dagger - \frac{1}{2}\left\{(A^{\omega}_m)^\dagger A^{\omega}_m, \rho\right\}\right)\,. \tag{A.147}$$

We can now characterise the properties of the jump operators and of the rates in the previous equation. First, it should be noticed that, since X^{ω}_i are eigenoperators of the auto-modular operator $\mathbb{L}_\pi \mathbb{R}^{-1}_\pi$, the same

holds for A^ω_m, as the unitary does not mix X^ω_is with different ωs. Moreover, Eq. (A.140) implies that $D^\omega_{\alpha,\beta} = e^\omega\, D^{-\omega}_{\delta_\beta(\omega),\gamma_\alpha(\omega)}$, where we used the same indices as in the equation. This relation shows that $D^{-\omega}_{\delta_\beta(\omega),\gamma_\alpha(\omega)}$ can be diagonalised as $e^{-\omega}\lambda^\omega_m \delta^m_n = (V^\omega)^\dagger_{m,\alpha} D^{-\omega}_{\delta_\beta(\omega),\gamma_\alpha(\omega)} (V^\omega)_{\beta,n}$. We can deduce the following two facts: first, the spectrum of $D^\omega_{\alpha,\beta}$ satisfies $\lambda^\omega_i = e^\omega\, \lambda^{-\omega}_i$; second, since $(X^\omega_\alpha)^\dagger = X^{-\omega}_{\gamma_\alpha(\omega)}$, it also holds that $(A^\omega_i)^\dagger = X^{-\omega}_{\gamma_\alpha(\omega)} (V^\omega_{\alpha,i})^\dagger = X^{-\omega}_{\gamma_\alpha(\omega)} V^{-\omega}_{\gamma_\alpha(\omega),i} = A^{-\omega}_i$. Hence, we see that $\mathcal{L}_{\mathcal{D}_1}$ satisfies the same conditions of Def. 2, namely:

1. $(A^\omega_i)^\dagger = A^{-\omega}_i$;
2. $\pi\, A^\omega_i\, \pi^{-1} = e^\omega\, A^\omega_i$;
3. $\lambda^\omega_i = e^\omega\, \lambda^{-\omega}_i$.

Since $\mathcal{L}_{\mathcal{D}_1}$ is the only component of the dissipator if one uses Def. 1, this proves the equivalence between this notion of detailed balance and the structural characterisation in Def. 2.

We can now pass to characterise $\mathcal{L}_{\mathcal{D}_2}$. Thanks to Eq. (A.138) we can rewrite it as:

$$\mathcal{L}_{\mathcal{D}_2}(\rho) = \sum_{\substack{\alpha,\omega,\\ \omega_1\neq 0}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega+\omega_1)|\alpha}_{\delta_\alpha(\omega)|\beta_\alpha(\omega_1)} \, (F^\omega_\alpha\, \rho\, (F^\omega_{\beta_\alpha(\omega_1)})^\dagger)^T = \tag{A.148}$$

$$= \sum_{\substack{\alpha\neq\beta,\\ \omega}} (\mathcal{L}_{\mathcal{D}})^{\gamma_\beta(\omega)|\alpha}_{\delta_\alpha(\omega)|\beta} \, F^\omega_\beta\, \rho^T\, (F^\omega_\alpha)^\dagger = \tag{A.149}$$

$$= \sum_{\substack{\alpha\neq\beta,\\ \omega}} (\mathcal{L}_{\mathcal{D}})^{\delta_\beta(\omega)|\alpha}_{\gamma_\alpha(\omega)|\beta} \, F^\omega_\beta\, \rho^T\, (F^\omega_\alpha)^\dagger\,, \tag{A.150}$$

where we eliminated the dependence on ω_1 by enforcing the constraint $\alpha \neq \beta$. Finally, in the last line we changed exchanged the dummy indexes γ and δ to highlight the analogy with the other part of the Lindbladian. Indeed, define the matrix $\widetilde{D^\omega_{\alpha,\beta}} := (\mathcal{L}_{\mathcal{D}})^{\delta_\beta(\omega)|\alpha}_{\gamma_\alpha(\omega)|\beta}$ for any $\alpha \neq \beta$ and zero on the diagonal. It is interesting to compare it to the off-diagonal elements of $D^\omega_{\alpha,\beta} = (\mathcal{L}_{\mathcal{D}})^{\gamma_\alpha(\omega)|\alpha}_{\delta_\beta(\omega)|\beta}$: as it can be seen, the two are related by an exchange $\gamma_\alpha(\omega) \leftrightarrow \delta_\beta(\omega)$. Moreover, thanks to Eq. (A.123) it also holds that:

$$\widetilde{D^\omega_{\alpha,\beta}} = e^\omega\, \widetilde{D}^{-\omega}_{\delta_\beta(\omega),\gamma_\alpha(\omega)}\,, \tag{A.151}$$

which shows the analogy with $D^{\omega}_{\alpha,\beta}$ even further. Using Eq. (A.118) it also follows that $\widetilde{D^{\omega}_{\alpha,\beta}}$ is hermitian, so there exist a unitary matrix W such that $\widetilde{D^{\omega}_{\alpha,\beta}} = W^{\omega}_{\alpha,m}(\mu^{\omega}_{m}\delta^{m}_{n})(W^{\omega})^{\dagger}_{n,\beta}$. We also define the jump operators $B^{\omega}_{m} := (W^{\omega}_{m,\alpha})^{\dagger}F^{\omega}_{\alpha}$. Then, one can rewrite $\mathcal{L}_{\mathcal{D}_2}$ as:

$$\mathcal{L}_{\mathcal{D}_2}(\rho) = \sum_{\substack{\alpha\neq\beta,\\ \omega}} W^{\omega}_{\alpha,m}(\mu^{\omega}_{m}\delta^{m}_{n})(W^{\omega})^{\dagger}_{n,\beta}\, F^{\omega}_{\beta}\,\rho^{T}\,(F^{\omega}_{\alpha})^{\dagger} = \tag{A.152}$$

$$= \sum_{m} \mu^{\omega}_{m}\, B^{\omega}_{m}\,\rho^{T}(B^{\omega}_{m})^{\dagger}\,. \tag{A.153}$$

In analogy with the previous case it also holds that $\mu^{\omega}_{i} = e^{\omega}\mu_{i}^{-\omega}$ and $(B^{\omega}_{m})^{\dagger} = B^{-\omega}_{m}$, together with the fact that B^{ω}_{m} is an eigenoperator of $\mathbb{L}_{\pi}\mathbb{R}^{-1}_{\pi}$. Still, there is one crucial difference with $\mathcal{L}_{\mathcal{D}_1}$: since all the diagonal elements of $\widetilde{D^{\omega}_{\alpha,\beta}}$ are zero, this matrix is traceless, meaning that the sum of the eigenvalues is zero also zero. Hence, whereas $\lambda^{\omega}_{i} \geq 0$ for all i and ω, we have the extra constraint $\sum_i \mu^{\omega}_{i} = 0$, implying the negativity of some of the μ^{ω}_{i}. Putting everything together, we finally obtain the characterisation in Eq. (1.343).

Appendix B

Generalised contextuality

The reality described by quantum mechanics appears to be radically different from the classical world, but it is not so obvious how to pin-point what causes this divergence in phenomena: features as exotic and iconically quantum as wave-particle duality and quantum tunneling can appear in classical systems, together with quantisation of energy and bound states (as shown by the hydrodynamical analogues [104]), whereas more technical but equally important properties as the Heisenberg uncertainty principle, the no-cloning and no-broadcasting theorem can be reproduced simply by imposing epistemic restrictions on a carefully designed classical stochastic system [105]. The latter model is even able to show entanglement, arguably the most renowned quantum effect.

For this reason, in order to set a clear distinction between what classical and quantum systems can do, one has to take into account all classical hidden-variable models at once and explore their limitations with respect to quantum systems. Historically, this approach was initiated by Bell, who found the first genuinely quantum signature, i.e., non-locality. This was formulated in terms of inequalities that any hidden-variable classical model needs to satisfy, but that are violated by quantum mechanics (the famous Bell's inequalities).

Another very powerful concept that was later introduced as a genuine quantum feature is the one of contextuality. Differently from non-locality, that applies to multi-partite scenarios, contextuality is well suited to explore the non-classicality of single systems and, for this reason, it gives a powerful tool to certify the quantumness of phenomena appearing in quantum thermodynamics (see, e.g., Sec. 2.3.2).

B.1 Definition

The framework in which generalised contextuality is formulated is the one of ontological models [106]. These are a theoretical representation of the most general way in which a physical system can give rise to certain statistics. In particular, taking an operational stance, it can be argued that the only experimentally accessible aspect of reality are given by correlations of the form $p(k|\,\mathcal{P},\mathcal{T},\mathcal{M})$, i.e., the probability of obtaining an output k given a preparation $\mathcal{P}$, a transformation $\mathcal{T}$ and a measurement $\mathcal{M}$. It should be noticed that in this context $\mathcal{P}$, $\mathcal{T}$ and $\mathcal{M}$ do not refer to any underlying physical system, but rather describe an operational procedure carried out in the experiment.

At this point, one assumes the existence of a physical system subject to the experiment, whose state-space is usually referred by Λ and called the space of ontic states. Its elements provide an ultimate description of the physical state of the system, that can be classical (as in the case of hidden variables models), quantum or even post-quantum. Then, one assumes that to each preparation $\mathcal{P}$ one can associate a probability density $\mu_{\mathcal{P}}(\lambda)$ on the ontic space. In the same way, one associate to $\mathcal{T}$ and $\mathcal{M}$ two functions describing how the actions of the experimentalist affect the system: namely a transition matrix $\Gamma_{\mathcal{T}}(\lambda',\lambda)$ and an indicator function $\xi_{\mathcal{M}}(k|\lambda)$, where k is the output of the experiment. These elements should be compatible with the statistics $p(k|\,\mathcal{P},\mathcal{T},\mathcal{M})$, as encoded in the equation:

$$p(k|\,\mathcal{P},\mathcal{T},\mathcal{M}) = \int_{\Lambda} \mathrm{d}\lambda\, \mu_{\mathcal{P}}(\lambda)\, \Gamma_{\mathcal{T}}(\lambda',\lambda)\, \xi_{\mathcal{M}}(k|\lambda')\,. \tag{B.1}$$

In order to simplify the treatment and without loss of generality, it is customary to include the effects of the transformation in the measurement procedure, and for this reason we will neglect it in the following.

There are some properties of the probability density and of the indicator functions that need to be specified: first, the indicator function is required to satisfy $\sum_k \xi_{\mathcal{M}}(k|\lambda) = 1$, meaning that all the ontic states give some output with probability one. Then, we require $\mu_{\mathcal{P}}(\lambda)$ and $\xi_{\mathcal{M}}(k|\lambda)$ to be compatible with convex mixing: in particular, if one choses $\mathcal{P}_1$ or $\mathcal{P}_2$ conditioned to a coin toss with bias q, it is imposed that the corresponding probability density is given by $q\,\mu_{\mathcal{P}_1}(\lambda) + (1-q)\mu_{\mathcal{P}_2}(\lambda)$ (and similarly for $\xi_{\mathcal{M}}(k|\lambda)$). Finally, if a notion of coarse-graining is present in the theory, we also impose that the indicator functions are compatible with it. That is, if a measurement $\widetilde{\mathcal{M}}$ can be performed by summing over

some partition of the measurement $\mathcal{M}$, then the corresponding indicator function should satisfy $\xi_{\widetilde{\mathcal{M}}}(k|\lambda) = \sum_{k_{\mathcal{M}}} \xi_{\mathcal{M}}(k_{\mathcal{M}}|\lambda)$.

In order to exemplify how the various ingredients can be used in practice, it is interesting to examine the Beltrametti-Bugajski model for quantum mechanics [107]: in this case Λ is simply the set of all wave functions, the preparation of a pure state $|\psi\rangle$ is associated to the δ-function:

$$\mu_\psi(\lambda) = \delta(\lambda - \psi)\,, \tag{B.2}$$

while the indicator function associated to a set of POVM $\{M_i\}_{i\in\{1,2..,n\}}$ is given by:

$$\xi_{\{M_i\}}(k|\lambda) = \mathrm{Tr}\left[M_k\ |\lambda\rangle\langle\lambda|\right]\,. \tag{B.3}$$

In this way, it is obvious that the statistics of quantum mechanics can be completely recovered. In fact, given a density matrix $\rho = \sum_i p_i\ |\psi_i\rangle\langle\psi_i|$, and a POVM $\{M_i\}_{i\in\{1,...,n\}}$, the corresponding response function is given by:

$$p(k|\rho, \{M_i\}_{i\in\{1,2..,n\}}) = \sum_i p_i \int_\Lambda \mathrm{d}\lambda\, \delta(\lambda - \psi_i)\, \mathrm{Tr}\left[M_k\ |\lambda\rangle\langle\lambda|\right] = \tag{B.4}$$

$$= \sum_i p_i\, \mathrm{Tr}\left[M_k\ |\psi_i\rangle\langle\psi_i|\right] = \mathrm{Tr}\left[M_k\,\rho\right]\,, \tag{B.5}$$

i.e., one recovers the usual Born rule for density matrices.

A central concept in this framework is the on of operational indistinguishability. Two preparations $\mathcal{P}_1$ and $\mathcal{P}_2$ are operationally indistinguishable if for any measurements $\mathcal{M}$ they satisfy:

$$p(k|\,\mathcal{P}_1, \mathcal{M}) = p(k|\,\mathcal{P}_2, \mathcal{M})\,. \tag{B.6}$$

In the same way, for measurements we say that $\mathcal{M}_1$ and $\mathcal{M}_2$ are operationally indistinguishable if for any preparations they satisfy:

$$p(k|\,\mathcal{P}, \mathcal{M}_1) = p(k|\,\mathcal{P}, \mathcal{M}_2)\,. \tag{B.7}$$

Now, a system is called non-contextual if operational indistinguishability is reflected at the level of the ontological model, i.e., if for any operationally equivalent preparation one assigns the same probability distribution $\mu_{\mathcal{P}}(\lambda)$ (and similarly for the measurement procedure). Interestingly, classical systems always admit a non-contextual ontological model.

On the other hand, we can see that the most obvious model for quantum mechanics, the Beltrametti-Bugajski one, is contextual: given two different decomposition of the state $\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i| = \sum_j q_j \, |\phi_j\rangle\langle\phi_j|$, one obtains two different representations on the ontic state

$$\sum_i p_i \, \mu_{\psi_i}(\lambda) \neq \sum_j q_j \, \mu_{\phi_j}(\lambda) \, . \tag{B.8}$$

Somehow surprisingly this is a common feature for quantum mechanics: it was shown in [106] that there is no non-contextual ontological model for quantum theory. Hence, if one can certify contextuality, then it also implies that the phenomenon studied does not admit a classical hidden-variable explanation.

B.2 Relation to other notions of non-classicality

The textbook definition for contextuality could seem to differ from the one presented here. In particular, one says that a measurement is Kochen-Specker non-contextual if one can specify the statistics of any elements of a POVM without having to refer to the other elements that are not being measured [108]. In this sense, the output of a measurement does not depend on the context in which it is being measured. Quantum mechanics is contextual according to Kochen-Specker for Hilbert space of dimension strictly higher than 2.

It is straightforward to see how this definition compares to the one of non-contextuality given above. To this end, consider an element M_k of a generic measurement. This can be prepared in at least two ways: either as part of the full POVM $\{M_i\}_{i\in\{1,\ldots,n\}}$, or as part of the reduced set $\{M_k, \mathbb{1} - \sum_{i\neq k} M_i\}$. Since for non-contextual models the indicator function does not depend on the way in which a POVM is prepared, the statistics for M_k needs to coincide in the two cases. Hence, non-contextuality implies Kochen-Specker non-contextuality. It should be noticed, though, that whereas Kochen-Specker contextuality was proven in dimension 3 or higher, the generalised version holds for any quantum system.

Another important criterion for non-classicality is the presence of negativities in the Wigner function of a state. Interestingly, these are strictly related to contextuality. In order to specify exactly how, we need to introduce the framework of quasi-probability representations: in this context,

to any operational theory one assigns a quasi-measure $\mu_\rho(\lambda)$ to each operationally equivalent preparation and a set of real functions $\xi_k(\lambda)$ that sum to unity for each operationally equivalent measurement. Then, a non-negative representation is one in which quasi-probabilities can be substituted with probabilities, that is, all the quantities are strictly positive. These are equivalent to non-contextuality [109]: in fact, if one has a non-negative representation, this provides a non-contextual ontological model, as $\mu_\rho(\lambda)$ and $\xi_k(\lambda)$ do not depend on the specific preparation; on the other hand, a non-contextual model provides the required non-negative representation. Hence, the contextuality of quantum mechanics implies that one cannot construct a quasi-probability representation which does not show negativities.

B.3 Witnessing contextuality through weak measurements

Contextuality is a rather delicate property, so the back-action caused by strong measurement can disturb the system enough to completely wash it away. For this reason, one has to rely on weak measurements to witness it (Thm. 25).

These are minimally invasive measurements which are usually realised as follows. Suppose one wants to estimate the observable $\mathcal{O} := \sum_i o_i \, \Pi^{(i)}$, where $\Pi^{(i)}$ are orthonormal projectors. First, one prepares an ancillary system called the pointer in some specific state. One widely used choice is to set the pointer state to be Gaussian, that is of the form:

$$\big|\psi_\nearrow\big\rangle = \frac{1}{\sqrt[4]{\pi s^2}} \int \mathrm{d}x \; e^{-\frac{x^2}{2s^2}} \; |x\rangle \; . \tag{B.9}$$

Then, the pointer interacts with the system for a time τ_0 through the Hamiltonian $H_{\text{int}} = g(t) \; \Pi^{(i)} \otimes P$, where P is the momentum operator, while the units of time are chosen so that $\int_0^\tau \mathrm{d}t \;\; g(t) = 1$; in this step information about $\rho \, \Pi^{(i)}$ is transferred to the pointer (where ρ is the initial state of the sytem). The final state of the system and pointer is given by:

$$\begin{aligned} U_{\text{int}} \rho \otimes \big|\psi_\nearrow\big\rangle\big\langle\psi_\nearrow\big| \, U_{\text{int}}^\dagger = \\ = \frac{1}{\sqrt{\pi s^2}} \sum_{k,k' \in \{0,1\}} \rho^{kk'} \otimes \int \mathrm{d}x \mathrm{d}y \; e^{-\frac{(x-k)^2}{2s^2}} e^{-\frac{(y-k')^2}{2s^2}} \; |x\rangle\langle y| \; , \end{aligned} \tag{B.10}$$

where it is useful to decompose the initial state of the system as:

$$\rho^{11} = \Pi^{(i)}\,\rho\,\Pi^{(i)}\,, \qquad \rho^{01} = (\Pi^{(i)})^{\perp}\,\rho\,\Pi^{(i)}\,, \tag{B.11}$$

$$\rho^{10} = \Pi^{(i)}\,\rho\,(\Pi^{(i)})^{\perp}\,, \quad \rho^{00} = (\Pi^{(i)})^{\perp}\,\rho\,(\Pi^{(i)})^{\perp}\,. \tag{B.12}$$

At this point, the pointer state is projectively measured in the basis $|x\rangle\langle x|$, inducing a POVM on the system denoted by $\{M_s^x\}_{x\in\mathbb{R}}$. Finally, one projectively measures a second observable $\tilde{\mathcal{O}} := \sum_j o_j\,\tilde{\Pi}^{(j)}$ and post-selects some value j. After this step, the position of the pointer is measured, while $\tilde{\Pi}_j$ is post-selected on the system. Then, the average position of the pointer is given by:

$$\langle X\rangle_{|j} = \mathrm{Tr}\left[(\tilde{\Pi}^{(j)}\otimes X)\,U_{\mathrm{int}}\,\rho\otimes\left|\psi_{\nearrow}\right\rangle\!\left\langle\psi_{\nearrow}\right|U_{\mathrm{int}}^{\dagger}\right]/q_j\,, \tag{B.13}$$

where q_j is the probability of success of the post-selection, that is:

$$q_j = \mathrm{Tr}\left[(\tilde{\Pi}^{(j)}\otimes \mathbb{1})U_{\mathrm{int}}\,\rho\otimes\left|\psi_{\nearrow}\right\rangle\!\left\langle\psi_{\nearrow}\right|U_{\mathrm{int}}^{\dagger}\right]. \tag{B.14}$$

The average position of the pointer encodes information about the initial value of $\Pi^{(i)}$. In particular, in the following we show how $\langle X\rangle_{|j}$ is connected to the concept of generalised weak value, which are defined as:

$$\frac{\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\,\Pi^{(i)}\,\rho\right]}{\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\,\rho\right]}\,, \tag{B.15}$$

which will be the central object of Thm. 25.

We can now estimate the value of $\langle X\rangle_{|j}$. Thanks to the Gaussian structure of the pointer state, it is straightforward to carry out the computation of q_j and $\langle X\rangle_{|j}$. In particular, the success probability is given by:

$$\begin{aligned} q_j = {} & \\ & \left(\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{00}\right] + \mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{11}\right]\right)\frac{1}{\sqrt{\pi s^2}}\int \mathrm{d}x\mathrm{d}y\; e^{-\frac{x^2}{2s^2}}e^{-\frac{y^2}{2s^2}}\,\delta(x-y) + \\ & + \left(\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{01}\right] + \mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{10}\right]\right)\frac{1}{\sqrt{\pi s^2}}\int \mathrm{d}x\mathrm{d}y\; e^{-\frac{x^2}{2s^2}}e^{-\frac{(y-1)^2}{2s^2}}\,\delta(x-y) = \end{aligned} \tag{B.16}$$

$$= \mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{00}\right] + \mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{11}\right] + 2\,e^{-1/4s^2}\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{01}\right]\Big] = \tag{B.17}$$

$$= \mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho\right] - 2\,(1-e^{-1/4s^2})\,\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\rho^{01}\right]\Big]\,, \tag{B.18}$$

where in Eq. (B.16) we grouped together ρ^{00} and ρ^{11}, since changing the average of the Gaussian does not affect the normalisation, in Eq. (B.17) we used the fact that $(\rho^{01})^\dagger = \rho^{10}$ to take the real part, and in the last line we added and subtracted $(\rho^{01} + \rho^{10})$.

In the same way, we can compute the numerator of $\langle X \rangle_{|j}$. In this case, we need the following three integrals:

$$\frac{1}{\sqrt{\pi s^2}} \int \mathrm{d}x \; x \, e^{-\frac{x^2}{s^2}} = 0 \,; \qquad \frac{1}{\sqrt{\pi s^2}} \int \mathrm{d}x \; x \, e^{-\frac{(x-1)^2}{s^2}} = 1 \,; \tag{B.19}$$

$$\frac{1}{\sqrt{\pi s^2}} \int \mathrm{d}x \; x \, e^{-\frac{(x-1)^2 + x^2}{2s^2}} = \frac{e^{-1/4s^2}}{2} \,. \tag{B.20}$$

It is a matter of simple algebra to show that:

$$\langle X \rangle_{|j} = \frac{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{11}\right] + e^{-1/4s^2} \mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{01}\right]\Big]}{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho\right] - 2\,(1 - e^{-1/4s^2})\,\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{01}\right]\Big]} \,. \tag{B.21}$$

Interestingly, in the limit $s \to 0$ the pointer state becomes a δ-function, and the measurement becomes a normal projective measurement. Indeed, we have that:

$$\lim_{s \to 0} \langle X \rangle_{|j} = \frac{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{11}\right]}{q_j} = \frac{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \Pi^{(i)} \rho \Pi^{(i)}\right]}{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} (\rho^{00} + \rho^{11})\right]} \,, \tag{B.22}$$

i.e, the probability of measuring the output i for the observable $\mathcal{O}$, followed by the measurement of the output j for the observable $\tilde{\mathcal{O}}$. Notice that in this case one has to renormalise with respect to the two possible output of the PVM $\Pi^{(i)}$, namely, 0 or 1.

In the opposite limit one obtains the generalised weak value from Eq. (B.15). Indeed, for $s \to \infty$, (which corresponds to a weak measurement, as the pointer is completely delocalised), the average position becomes:

$$\lim_{s \to \infty} \langle X \rangle_{|j} = \frac{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{11}\right] + \mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho^{01}\right]\Big]}{\mathrm{Tr}\left[\tilde{\Pi}^{(j)} \rho\right]} = \frac{\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\, \Pi^{(i)}\, \rho\right]\Big]}{\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\, \rho\right]} \,, \tag{B.23}$$

where in the second equation one implicitly uses that $\Pi^{(i)} + (\Pi^{(i)})^\perp = \mathbb{1}$. Notice that the measurement is said to be weak because, as the spread

in the position basis is infinite ($s \to \infty$), the momentum of the particle is zero, which causes almost no disturbance to be introduced in the system. The construction just presented provides an operational procedure of accessing the generalised weak values.

A weak value is called anomalous if it is not contained in the spectrum of the measured operator. It was shown in [110] for pure states, and later generalised to density matrices in [52], that anomalous weak values are a proof of contextuality. This is done in two steps: first, it is shown that a set of requirements cannot be satisfied by any non-contextual ontological model (Thm. 25). Then, we show that in the presence of anomalous weak values these requirements can be satisfied (Thm. 26). The proofs presented follow [52].

Theorem 25. *Given a POVM* $\{M_s^x\}_{x\in\mathbb{R}}$ *and a post-selection* $\tilde{\Pi}^{(j)}$ *satisfying the following three properties:*

1. *every element of the POVM can be decomposed as* $M_s^x = p^s(x-1)\Pi^{(i)} + p^s(x)(\Pi^{(i)})^\perp$*, where* p^s *is a probability distribution with median* $x = 0$*;*

2. *Disregarding the outcome of the initial weak measurement leads to a disturbance* $p_d^s \leq \frac{1}{2}$*, i.e.:*

$$S_s = \int \mathrm{d}x\ (N_s^x)^\dagger\, \tilde{\Pi}^{(j)}\, N_s^x = (1-p_d^s)\,\tilde{\Pi}^{(j)} + p_d^s\, E^d\,, \tag{B.24}$$

   where N_s^x *satisfies the property* $M_s^x := (N_s^x)^\dagger N_s^x$*, and* E^d *is an element of a POVM;*

3. *The values of* x *under the pre- and post-selection have a negative bias which outweighs* p_d^s *for* s *large enough. That is:*

$$p_-^s := \frac{1}{p_{\tilde{\Pi}^{(j)}}} \int_{-\infty}^{0} \mathrm{d}x\ \mathrm{Tr}\left[(N_s^x)^\dagger \tilde{\Pi}^{(j)} N_s^x\, \rho\right] > \frac{1}{2} + \frac{p_d^s}{p_{\tilde{\Pi}^{(j)}}}\,, \tag{B.25}$$

   where $p_{\tilde{\Pi}^{(j)}}$ *is the probability of measuring* $\tilde{\Pi}^{(j)}$ *in the state* ρ*.*

Then, there exists no non-contextual, outcome deterministic ontological model reproducing the statistics of the experiment.

Proof. The proof proceeds by showing that these three properties cannot be satisfied by any measurement non-contextual outcome deterministic ontological model.

Before starting, let us define the POVM $\{S_s^x\}_{x\in\mathbb{R}}\cup\{F_s^x\}_{x\in\mathbb{R}}$ describing the overall effect of the initial POVM and the final post-selection, and corresponding to whether this was successful or not. In particular, these operators are given by $S_s^x := (N_s^x)^\dagger\tilde{\Pi}^{(j)}N_s^x$ and $F_s^x := (N_s^x)^\dagger(\mathbb{1}-\tilde{\Pi}^{(j)})N_s^x$.

Now, consider the probability of obtaining the outcome x in the POVM $\{M_s^x\}_{x\in\mathbb{R}}$, given the preparation corresponding to a state ρ. In the ontological model this is expressed by the response function:

$$p(x|\,\rho,\{M_s^x\}_{x\in\mathbb{R}}) = \int_\Lambda \mathrm{d}\lambda\,\mu_\rho(\lambda)\;\xi_{\{M_s^x\}}(x|\lambda)\,. \tag{B.26}$$

If the model is non-contextual equivalent preparations are represented in the same way at the level of the ontological model. In particular, consider the two possibilities for measuring x: in one case, the measurement $\{S_s^x\}\cup\{F_s^x\}$ is performed, and one disregards the output of the projective measurement; in the other case, one measures $\{\Pi^{(i)}\}\cup\{(\Pi^{(i)})^\perp\}$ and classically samples from $p^s(x)$. Thanks to condition 1 this reconstructs the POVM M_s^x. Non-contextuality means that these two procedures are equivalent at the level of the ontological model, i.e.,:

$$\xi(M_s^x|\lambda) = \xi(S_s^x|\lambda)+\xi(F_s^x|\lambda) \tag{B.27}$$

$$= p^s(x-1)\,\xi(\Pi^{(i)}|\lambda)+p^s(x)\,\xi((\Pi^{(i)})^\perp|\lambda)\,. \tag{B.28}$$

Notice that, thanks to non-contextuality, we can drop the dependence of the indicator function ξ on which POVM is being performed. Eq. (B.28) allows us to prove the inequality:

$$\int_{-\infty}^0 \mathrm{d}x\;\xi(S_s^x|\lambda) \le \int_{-\infty}^0 \mathrm{d}x\;\left(\xi(S_s^x|\lambda)+\xi(F_s^x|\lambda)\right) = \tag{B.29}$$

$$= \int_{-\infty}^0 \mathrm{d}x\;\xi(M_s^x|\lambda) = \tag{B.30}$$

$$= \int_{-\infty}^0 \mathrm{d}x\;\left(p^s(x-1)\,\xi(\Pi^{(i)}|\lambda)+p^s(x)\,\xi((\Pi^{(i)})^\perp|\lambda)\right) \le \tag{B.31}$$

$$\le \frac{1}{2}\left(\xi(\Pi^{(i)}|\lambda)+\xi((\Pi^{(i)})^\perp|\lambda)\right) = \frac{1}{2}\,, \tag{B.32}$$

where the last inequality comes from the fact that $p^s(x)$ has median in $x = 0$, and we finally used the fact that the probability of two incompatible and exhaustive events is 1.

A second ingredient we need is the probability of obtaining a successful post-selection. Again there are two ways to compute it: first,

just by considering the probability of obtaining $\{S_s^x\} \cup \{F_s^x\}$, and then disregarding the value of x; secondly, by using condition 2, one can perform the measurement $\{\tilde{\Pi}^{(j)}\} \cup \{\mathbb{1} - \tilde{\Pi}^{(j)}\}$ with probability $(1 - p_d^s)$, or $\{E^d\} \cup \{\mathbb{1} - E^d\}$ with probability p_d^s. For this last procedure, one post-selects the first outcome in both cases. Then, we can rewrite the indicator function as:

$$\xi(S_s|\lambda) = \int \mathrm{d}x\, \xi(S_s^x|\lambda) = (1 - p_d^s)\, \xi(\tilde{\Pi}^{(j)}|\lambda) + p_d^s\, \xi(E^d|\lambda). \tag{B.33}$$

Since $\tilde{\Pi}^{(j)}$ is a projective measurement and the ontological model is assumed to be outcome deterministic, Λ can be partitioned as $\Lambda_0 \cup \Lambda_1$, with the property that $\lambda \in \Lambda_i \implies p(\tilde{\Pi}^{(j)}|\lambda) = \delta_{i,1}$. All the ingredients to estimate p_-^s are now ready. Its formulation in the ontological model is given by:

$$p_-^s = \frac{1}{p_{\tilde{\Pi}^{(j)}}} \int_{-\infty}^{0} \mathrm{d}x\, \mathrm{Tr}\left[(N_s^x)^\dagger \tilde{\Pi}^{(j)} N_s^x\, \rho\right] = \tag{B.34}$$

$$= \frac{1}{p_{\tilde{\Pi}^{(j)}}} \int_{-\infty}^{0} \mathrm{d}x \int_{\Lambda} \mathrm{d}\lambda\, \mu_\rho(\lambda)\; \xi(S_s^x|\lambda) = \tag{B.35}$$

$$= \frac{1}{p_{\tilde{\Pi}_\tau^{(j)}}} \left(\int_{-\infty}^{0} \mathrm{d}x \int_{\Lambda_0} \mathrm{d}\lambda\, \mu_\rho(\lambda)\; \xi(S_s^x|\lambda) + \int_{-\infty}^{0} \mathrm{d}x \int_{\Lambda_1} \mathrm{d}\lambda\, \mu_\rho(\lambda)\; \xi(S_s^x|\lambda) \right). \tag{B.36}$$

The first integral can be bound by p_d: in fact, decomposing it through Eq. (B.33), one can use the fact that on Λ_0 one has $\xi(\tilde{\Pi}^{(j)}|\lambda) = 0$, so the value of the first integral can be bound by p_d^s. On the other hand, thanks to Eq. (B.29), the second integral can be bounded by $1/2$, times the area of Λ_1, which by construction is $p_{\tilde{\Pi}_\tau^{(j)}}$. This means that for a non-contextual model $p_-^s \leq (p_{\tilde{\Pi}_\tau^{(j)}}/2 + p_d^s)/p_{\tilde{\Pi}_\tau^{(j)}}$. But this is in contradiction with condition 3. Hence, there is no non-contextual ontological model reproducing the three conditions of the theorem. □

Theorem 25 gives a way to certify the genuine quantumness of a phenomena. In particular, we show now that anomalous weak values are able to satisfy all the conditions of Thm. 25 whenever they show some negativity:

Theorem 26. *If the real part of the weak value in Eq.* (B.15) *is negative for some choice of indexes i and j, then it always exists a pointer*

state with s large enough so that there is no measurement non-contextual, outcome deterministic ontological model reproducing the statistics of the weak measurement.

Proof. Consider a weak measurement scheme as explained above. It should be noticed that the POVM induced by the measurement of the pointer state can be expressed as:

$$N_s^x = \frac{1}{\sqrt[4]{\pi s^2}}\, e^{-\frac{(x-1)^2}{2s^2}}\, \Pi^{(i)} + \frac{1}{\sqrt[4]{\pi s^2}}\, e^{-\frac{x^2}{2s^2}}\, (\Pi^{(i)})^{\perp}\,; \tag{B.37}$$

$$M_s^x = \frac{1}{\sqrt{\pi s^2}}\, e^{-\frac{(x-1)^2}{s^2}}\, \Pi^{(i)} + \frac{1}{\sqrt{\pi s^2}}\, e^{-\frac{x^2}{s^2}}\, (\Pi^{(i)})^{\perp}\,, \tag{B.38}$$

where N_s^x satisfies $M_s^x := (N_s^x)^\dagger N_s^x$. Indeed, one can immediately deduce M_s^x from Eq. (B.10), while for N_s^x one needs to use the condition $\Pi^{(i)}(\Pi^{(i)})^\perp = (\Pi^{(i)})^\perp \Pi^{(i)} = 0$. Then condition 1 is satisfied with $p^s(x) = \frac{1}{\sqrt{\pi s^2}}\, e^{-\frac{x^2}{s^2}}$.

We can now evaluate the operator:

$$S_s = \int \mathrm{d}x\, (N_s^x)^\dagger\, \tilde{\Pi}^{(j)}\, N_s^x\,, \tag{B.39}$$

Again, the Gaussian structure of the pointer state allows for an analytical solution (we omit the lengthy but straightforward computations):

$$\begin{aligned} S_s &= \\ &= \frac{1+e^{-\frac{1}{4s^2}}}{2}\, \tilde{\Pi}^{(j)} + \frac{1-e^{-\frac{1}{4s^2}}}{2}\, \left(\Pi^{(i)} - (\Pi^{(i)})^{\perp}\right) \tilde{\Pi}^{(j)} \left(\Pi^{(i)} - (\Pi^{(i)})^{\perp}\right). \end{aligned} \tag{B.40}$$

One can define $E^d := \left(\Pi^{(i)} - (\Pi^{(i)})^{\perp}\right) \tilde{\Pi}^{(j)} \left(\Pi^{(i)} - (\Pi^{(i)})^{\perp}\right)$ to be part of the POVM $\{E^d, \mathbb{1} - E^d\}$ (notice in fact that it is not only positive, but also a projector, as it satisfies $(E^d)^2 = E^d$), and set the probability of disturbance to $p_d^s := \frac{1-e^{-\frac{1}{4s^2}}}{2}$. With these choices condition 2 is also satisfied.

It remains to show that for s large enough the negative bias outweighs p_d^s, i.e.,

$$p_-^s := \frac{1}{p_{\tilde{\Pi}_\tau^{(j)}}} \int_{-\infty}^{0} \mathrm{d}x\, \mathrm{Tr}\left[(N_s^x)^\dagger \tilde{\Pi}_\tau^{(j)} N_s^x \rho\right] > \frac{1}{2} + \frac{p_d^s}{p_{\tilde{\Pi}_\tau^{(j)}}}\,. \tag{B.41}$$

To this end, it is useful to keep in mind the following Gaussian integrals:

$$\frac{1}{\sqrt{\pi s^2}}\int_{-\infty}^{0} \mathrm{d}x\; e^{-\frac{(x-1)^2}{s^2}} = \frac{1}{2}\mathrm{erfc}\left(\frac{1}{\mathrm{s}}\right); \tag{B.42}$$

$$\frac{1}{\sqrt{\pi s^2}}\int_{-\infty}^{0} \mathrm{d}x\; e^{-\frac{x^2}{s^2}} = \frac{1}{2}; \tag{B.43}$$

$$\frac{1}{\sqrt{\pi s^2}}\int_{-\infty}^{0} \mathrm{d}x\; e^{-\frac{x^2}{2s^2}} e^{-\frac{(x-1)^2}{2s^2}} = \frac{e^{-\frac{1}{4s^2}}}{2}\mathrm{erfc}\left(\frac{1}{2\mathrm{s}}\right), \tag{B.44}$$

where erfc(x) is the complementary error function. Notably, in the limit of $s \to \infty$ we can expand it as:

$$\frac{1}{2}\mathrm{erfc}\left(\frac{1}{\mathrm{s}}\right) = \frac{1}{2} - \frac{1}{\sqrt{\pi s^2}} + \mathcal{O}\left(\frac{1}{s^2}\right) \tag{B.45}$$

At this point one can plug in the explicit expression of N_s^x in Eq. (B.41), and it is a matter of computations to show that:

$$p_-^s = \frac{1}{2} - \frac{1}{s\sqrt{\pi\, p^2_{\tilde{\Pi}_\tau^{(j)}}}}\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}^{(j)}\,\Pi^{(i)}\,\rho\right]\Big] + \mathcal{O}\left(\frac{1}{s^2}\right). \tag{B.46}$$

On the other hand, the probability of disturbance can be expanded at infinity as:

$$p_d^s = \frac{1 - e^{-\frac{1}{4s^2}}}{2} = \frac{1}{8s^2} + \mathcal{O}\left(1/s^3\right). \tag{B.47}$$

By assumption, there exists i and j such that the weak value in Eq. (B.15) has negative real part. it should be noticed that the negativity must arise from the numerator, as the denominator is just the probability of measuring the output j. This means that:

$$\mathrm{Re}\Big[\mathrm{Tr}\left[\tilde{\Pi}_\tau^{(j)}\,\Pi_0^{(i)}\,\rho\right]\Big] < 0\,, \tag{B.48}$$

so that also condition 3 can be satisfied for s large enough. This shows that there cannot be any non-contextual explanations for negativities in the real part of the weak values. □

It should be noticed that the negativity which is treated here are not equivalent to the one in the Wigner function. In the latter case, one wants to provide a quasi-probabilistic representation of all quantum mechanics, and the failing to do so is a witness for the contextuality of the theory. In this context, instead, the negativity is just associated to an anomalous weak value, and for this reason it is not directly related to a quasi-probabilistic representation.

Appendix C

Outperforming Bayes' retrodiction

In this appendix we present the results obtained in [2] on the characterisation of state retrieval maps. Our aim is to find a definition of a reverse map that retrieves the original states as well as possible. Somewhat surprisingly, this does not coincide with the one obtained from Bayes' theorem. In particular, we show that in some cases taking Bayes' as the reverse map results in a further deterioration of the information contained in the state. Finally, we give arguments showing that an extra axiom could in fact isolate Bayes' reverse map and Petz recovery map as the only inversion.

Given a channel Φ and a prior state π, we define a state retrieval map to be one satisfying the following four conditions:

1. $\tilde{\Phi}$ is physically implementable;

2. if Φ^{-1} is physically implementable then $\tilde{\Phi} \equiv \Phi^{-1}$;

3. $\tilde{\Phi}\Phi$ is detailed balance with respect to the prior π;

4. $\tilde{\Phi}\Phi$ has only positive eigenvalues.

The first three were motivated in the main text, while an explanation of the fourth will be given in Sec. C.2. Before showing how one can define a criteria to isolate an optimal state retrieval, we present a useful parametrisation of physical maps which have a specific transition $\pi \to \sigma$.

C.1 Parametrisation of maps for a given transition

In this section we show how to parametrise all the maps Ψ having a definite transition $\pi \to \sigma$, i.e., satisfying $\Psi(\pi) = \sigma$. We will first present how this can be done for classical stochastic maps, and then we will pass to analyse the case for quantum channels.

Suppose, then, to have a map Ψ having the transition $\pi \to \sigma$. It is useful to rewrite it as follows:

$$\Psi = L^{\Psi} \mathcal{J}_{\pi}^{-1}, \tag{C.1}$$

where $\mathcal{J}_{\pi}$ is a diagonal matrix with entries $(\mathcal{J}_{\pi})_{i,i} := (\pi)_i$, and L^{Ψ} is implicitly defined by the equation $L^{\Psi} := \Psi\, \mathcal{J}_{\pi}$. Since $\mathcal{J}_{\pi}$ has positive entries, the same holds for L^{Ψ} . Moreover, it satisfies the two conditions:

$$\sum_i L^{\Psi}_{i,j} = \sum_i \Psi_{i,j}\, (\mathcal{J}_{\pi})_{j,j} = \pi_j\,; \tag{C.2}$$

$$\sum_j L^{\Psi}_{i,j} = \sum_j L^{\Psi}_{i,j}\, (\mathcal{J}_{\pi}^{-1})_{j,j} (\pi)_j = \sigma_i\,, \tag{C.3}$$

where in Eq. (C.2) we used the fact that stochastic matrices satisfy $\sum_i \Psi_{i,j} = 1$, while in Eq. (C.3) that $L^{\Psi} \mathcal{J}_{\pi}^{-1}(\pi) = \Psi(\pi) = \sigma$. In this way, we are able to uniquely associate to each Ψ a member L^{Ψ} of $\mathcal{U}(\sigma, \pi)$, the space of matrices with non-negative entries, with columns summing to σ and rows summing to π. This space was studied in [111] where it was shown that it is a convex polytope with a finite number of vertices, denoted by $V^{(k)}_{\sigma|\pi}$. What is also interesting is that since the two conditions in Eq. (C.2)-(C.3) are connected by a matrix transpose, there is a one to one relation between the vertices of $\mathcal{U}(\sigma, \pi)$ and the one of $\mathcal{U}(\pi, \sigma)$ given by $(V^{(k)}_{\sigma|\pi})^T = V^{(k)}_{\pi|\sigma}$.

Then, for every map Ψ, there exists at least one probability vector $\{\lambda^{(\Psi)}_k\}$ such that:

$$\Psi = \sum_k \lambda^{(\Psi)}_k\, V^{(k)}_{\sigma|\pi} \mathcal{J}_{\pi}^{-1}\,. \tag{C.4}$$

Since $\mathcal{U}(\sigma, \pi)$ is a convex polytope, but not a simplex, the coefficients $\{\lambda^{(\Psi)}_k\}$ are not necessarily unique.

A similar construction can be made for quantum channels. In this case, for a given map Ψ we can decompose it as:

$$\Psi = \Lambda^{\Psi}\, \mathbb{J}_{\sqrt{x}}^{-1}\big|_{\pi}\,, \tag{C.5}$$

where $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$ acts as $\mathbb{J}_{\sqrt{x}}\big|_{\pi}(\rho) := \sqrt{\pi}\rho\sqrt{\pi}$ (see Sec. 1.4.8), and Λ^{Ψ} is given by $\Lambda^{\Psi} = \Psi\, \mathbb{J}_{\sqrt{x}}\big|_{\pi}$. Since both Ψ and $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$ are CP, Λ^{Ψ} is CP as well. Analogously to the conditions in Eq. (C.2)-(C.3), we can characterise Λ^{Ψ} by its action on the identity operator:

$$(\Lambda^{\Psi})^{\dagger}[\mathbb{1}] = \mathbb{J}_{\sqrt{x}}\big|_{\pi}\, \Psi^{\dagger}[\mathbb{1}] = \pi\,; \tag{C.6}$$

$$(\Lambda^{\Psi})[\mathbb{1}] = \Psi\, \mathbb{J}_{\sqrt{x}}\big|_{\pi}\, [\mathbb{1}] = \sigma\,, \tag{C.7}$$

where in Eq. (C.6) we used the fact that Ψ is trace preserving together with the self-adjointness of $\mathbb{J}_{\sqrt{x}}\big|_{\pi}$, while the condition in Eq. (C.7) follows from $\mathbb{J}_{\sqrt{x}}\big|_{\pi}\,[\mathbb{1}] = \pi$. In this way, similarly to what happens for the classical case, a quantum channel is uniquely identified by a map $\Lambda^{\Psi} \in \mathcal{U}_Q(\sigma, \pi)$, the space of CP transformations that map the identity to σ, and whose adjoints map the identity to π. Even in this case $\mathcal{U}_Q(\sigma, \pi)$ is convex, and it has been characterised in terms of the Kraus operators corresponding to each vertex [36]. In particular, a map $V[\rho] := \sum_i V_i\, \rho\, V_i^{\dagger}$ is a vertex of $\mathcal{U}_Q(\sigma, \pi)$ if the following conditions hold:

1. $\sum_i V_i V_i^{\dagger} = \sigma$;
2. $\sum_i V_i^{\dagger} V_i = \pi$;
3. $(V_i V_j^{\dagger})_{i,j}$ and $(V_j^{\dagger} V_i)_{i,j}$ are jointly linear independent.

One important difference with the classical case is that $\mathcal{U}_Q(\sigma, \pi)$ contains a non-trivial symmetry (i.e., one that is not reducible to a relabeling): consider the two unitary maps U_{π} and V_{σ}, defined by $U_{\pi}[\rho] := U\rho U^{\dagger}$, and satisfying $U_{\pi}[\pi] = \pi$ (and analogously for V_{σ}, with $V_{\sigma}[\sigma] = \sigma$). Then, the conditions in Eq. (C.6)-(C.7) are invariant under the transformation:

$$\Lambda^{\Psi} \to V_{\sigma}\, \Lambda^{\Psi}\, U_{\pi}\,. \tag{C.8}$$

This means that every Λ^{Ψ} contained in $\mathcal{U}_Q(\sigma, \pi)$ is part of an invariant family connected by the unitary transformations defined in Eq. (C.8). Finally, notice that also in this case the elements of $\mathcal{U}_Q(\sigma, \pi)$ and $\mathcal{U}_Q(\pi, \sigma)$ are in one-to-one correspondence through the adjoint transformation.

Another possible characterisation of the space $\mathcal{U}_Q(\sigma,\pi)$ is in terms of a marginal problem for the corresponding unnormalised Choi state. This is defined by the formula $\mathcal{C}^\Psi := d\,(\mathbb{I}_A \otimes \Psi)[\,|\Omega\rangle\,\langle\Omega\,|]$, where $|\Omega\rangle$ is the maximally entangled state (see Eq. (1.255)). Then, the application of Ψ to a state ρ can be equivalently expressed as $\Psi[\rho] = \mathrm{Tr}_A\left[(\rho^T \otimes \mathbb{1})\,\mathcal{C}^\Psi\right]$. Moreover, it also holds that $\Psi^\dagger[\rho] = (\mathrm{Tr}_B\left[(\mathbb{1}\otimes\rho)\,\mathcal{C}^\Psi\right])^T$. This implies that we can translate the conditions on Λ^Ψ to constraints on the Choi matrix, namely:

$$\mathrm{Tr}_B\left[\mathcal{C}^{\Lambda^\Psi}\right] = \mathrm{Tr}_B\left[(\mathbb{1}\otimes\mathbb{1})\,\mathcal{C}^{\Lambda^\Psi}\right] = ((\Lambda^\Psi)^\dagger[\mathbb{1}])^T = (\pi)^T\,; \tag{C.9}$$

$$\mathrm{Tr}_A\left[\mathcal{C}^{\Lambda^\Psi}\right] = \mathrm{Tr}_A\left[(\mathbb{1}\otimes\mathbb{1})\,\mathcal{C}^{\Lambda^\Psi}\right] = (\Lambda^\Psi)[\mathbb{1}] = \sigma\,. \tag{C.10}$$

Thanks to the Choi-Jamiołkowski isomorphism we know that $\mathcal{C}^{\Lambda^\Psi}$ is positive semidefinite, and it follows from the two equations above that $\mathrm{Tr}\left[\mathcal{C}^{\Lambda^\Psi}\right] = 1$. Hence, $\mathcal{U}_Q(\sigma,\pi)$ can be parametrised as the set of all maps whose Choi matrix corresponds to a bipartite quantum state ρ_{AB} compatible with the two marginals $\rho_A = \sigma$ and $\rho_B = (\pi)^T$. This identification allows to constrain the spectrum of the Choi states in $\mathcal{U}_Q(\sigma,\pi)$. In fact, one can construct a system of linear inequalities depending on the spectrum of ρ_A and ρ_B to constrain the spectrum of ρ_{AB} [112, 113].

We can now formalise the concept of map reversion: this is a transformation $\mathcal{R}$ from $\mathcal{U}(\Phi\pi,\pi)$ to $\mathcal{U}(\pi,\Phi\pi)$ such that the conditions (1-4) are satisfied ($\mathcal{U}_Q(\Phi\pi,\pi)$ for quantum transformations). To be more precise, in order to revert a map Φ, we first need to define the corresponding $L^\Phi = \Phi\,\mathcal{J}_\pi$. Then, the reverse map $\tilde{\Phi}$ is given by $\tilde{\Phi} := \mathcal{R}(L^\Phi)\,\mathcal{J}^{-1}_{\Phi(\pi)}$, and we impose compatibility with conditions (1-4).

C.1.1 Bayes inspired reverse and Petz' recovery map

An interesting feature of this parametrisation is that the transformation $\mathcal{R}$ corresponding to Bayes' reverse map or to the Petz recovery map is particularly natural. In fact, if one writes Bayes' theorem in matrix form one obtains:

$$\tilde{\Phi}_B = \sum_{i,j} \frac{\pi_i\,\Phi_{j,i}}{(\Phi(\pi))_j}\;|i\rangle\langle j\,| = \mathcal{J}_\pi\,(\Phi)^T\mathcal{J}^{-1}_{\Phi(\pi)} = \mathcal{J}_\pi\,(L^\Phi\mathcal{J}_\pi^{-1})^T\mathcal{J}^{-1}_{\Phi(\pi)} = \tag{C.11}$$

$$= (L^\Phi)^T\mathcal{J}^{-1}_{\Phi(\pi)}\,. \tag{C.12}$$

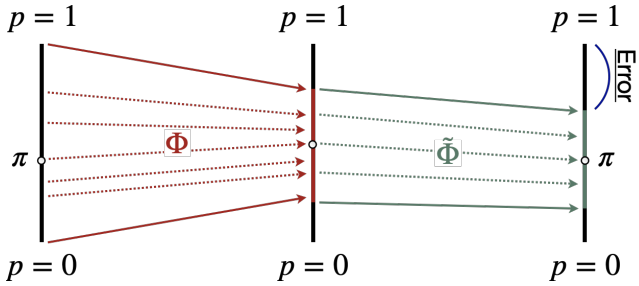


Figure C.1: Depiction of the action of a stochastic map Φ on a two levels classical system. Any physical state retrieval $\tilde{\Phi}$ must be contracting, implying that many states will not be in its domain. For this reason, in order to get as less error as possible, one requires the image of $\tilde{\Phi}\Phi$ to be as big as possible. A natural quantifier of the volume of the image of a linear map is its determinant. Moreover, it should be noticed that any negative eigenvalues in $\tilde{\Phi}\Phi$ corresponds to a flip in the image, which will significantly increase the error. For this reason, imposing condition 4 prevents this pathological situation from happening.

This shows that in this case $\mathcal{R}$ is simply the transpose map. In the same way, in the case of the Petz recovery map it is straightforward to show that:

$$\tilde{\Phi}_P = \mathbb{J}_{\sqrt{x}}\big|_{\pi}\, \Phi^{\dagger}\, \mathbb{J}^{-1}_{\sqrt{x}}\big|_{\Phi(\pi)} = \mathbb{J}_{\sqrt{x}}\big|_{\pi}\, (\Lambda^{\Phi}\, \mathbb{J}^{-1}_{\sqrt{x}}\big|_{\pi})^{\dagger}\, \mathbb{J}^{-1}_{\sqrt{x}}\big|_{\Phi(\pi)} = \tag{C.13}$$

$$= (\Lambda^{\Phi})^{\dagger}\, \mathbb{J}^{-1}_{\sqrt{x}}\big|_{\Phi(\pi)}\,. \tag{C.14}$$

Again, $\mathcal{R}$ is simply given by the adjoint map $\Lambda^{\Phi} \to (\Lambda^{\Phi})^{\dagger}$.

C.2 The max-det principle

Conditions (1-4) do not select a unique retrieval map, but rather a whole family of transformations. For this reason, we motivate here a maximisation principle that singles out a unique optimal state retrieval map $\tilde{\Phi}_O$.

To this end, consider a stochastic map from a space into itself (see Fig. C.1). This type of maps are contracting: the volume of their image will be smaller than the one of their domain. The composite transformation $\tilde{\Phi}\Phi$ falls into this class. In order to preserve as many states as possible, a minimal requirement is that the image of $\tilde{\Phi}\Phi$ is not too small.

A standard measure of the volume of the image of a linear map is its determinant, and for this reason we demand that:

Principle. *Optimal retrieval maps maximise the determinant of* $\tilde{\Phi}_O\,\Phi$ *under the constraints (1-4).*

Interestingly, this minimal requirement is enough to single out a unique $\tilde{\Phi}_O$. Before, moving on to provide some analytical insights about why we expect this to be a good definition, it is worth to point out that we imposed condition 4 following the same kind of geometrical considerations. In fact, notice that any negative or complex eigenvalue in the spectrum of $\tilde{\Phi}\Phi$ corresponds to a reflection or a rotation of the domain, which would increase the statistical distance between a state and its evolved version. For this reason, it is sufficient to explore state retrievals corresponding to $\tilde{\Phi}\Phi$ with positive spectrum. It should also be noticed that, indeed, both Bayes' and Petz maps satisfy this requirement. In fact, by performing a similarity transformation one can see that:

$$\mathcal{J}_\pi^{-1/2}\left(\tilde{\Phi}_B\Phi\right)\mathcal{J}_\pi^{1/2} = \mathcal{J}_\pi^{1/2}\left(\Phi\right)^T\mathcal{J}_{\Phi(\pi)}^{-1}\Phi\mathcal{J}_\pi^{1/2} = \tag{C.15}$$

$$= \left(\mathcal{J}_{\Phi(\pi)}^{-1/2}\Phi\mathcal{J}_\pi^{1/2}\right)^T\left(\mathcal{J}_{\Phi(\pi)}^{-1/2}\Phi\mathcal{J}_\pi^{1/2}\right), \tag{C.16}$$

which implies that their spectrum is also positive (the computations for Petz recovery map are completely analogous, substituting $\mathcal{J}$ with $\mathbb{J}$).

C.2.1 Analytical insights on the max-det principles

Beyond the intuitive necessity of having the image of the retrieval map as big as possible, the principle of the maximisation of the determinant can be justified more rigorously. We provide here some analytical insights on why this should in fact be a good requirement.

First, it should be noticed that in order to optimise the quality of the retrieval we have to make $\tilde{\Phi}\Phi$ as similar as possible to the identity transformation. Since both $\tilde{\Phi}\Phi$ and $\mathbb{I}$ are positive semidefinite matrices, the relative entropy between the two is well defined and takes the form:

$$D(\mathbb{I}||\tilde{\Phi}\Phi) = \mathrm{Tr}\left[\mathbb{I}\left(\log\mathbb{I} - \log\tilde{\Phi}\Phi\right)\right] = \tag{C.17}$$

$$= -\mathrm{Tr}\left[\log\tilde{\Phi}\Phi\right] = \log\det(\tilde{\Phi}\Phi)^{-1}, \tag{C.18}$$

where one uses the well known matrix identity $\mathrm{Tr}\left[\log A\right] = \log\det A$. Minimising this relative entropy is then equivalent to the maximisation

of the determinant of $\tilde{\Phi}\Phi$. Even if the relative entropy between two superoperators is not a standard object, Eq. (C.18) suggests that our principle indeed identifies the right statistical property to optimise.

A more standard approach, even if it provides only bounds and not equalities, is to study the action of $\tilde{\Phi}\Phi$ through the use of contrast functions. We remind the reader that for classical states Csiszár contrast functions are defined as (see Sec. 1.1):

$$H_g(\rho||\sigma) := \sum_i \rho_i \, g\left(\frac{\sigma_i}{\rho_i}\right) , \tag{C.19}$$

where g is an arbitrary positive convex function such that $g(0) = 0$. Moreover, $H_g(\rho||\sigma)$ are connected to the classical Fisher information by:

$$H_g(\rho||\rho + \delta\rho) \simeq \frac{1}{2}\mathrm{Tr}\left[\delta\rho \, \mathcal{J}_\rho^{-1}[\delta\rho]\right] = \sum_i \frac{\delta\rho_i^2}{\rho_i} , \tag{C.20}$$

where $|\delta\rho| \ll 1$. On the other hand, from the family of quantum contrast functions we consider here only the one associated to the square root (see Sec. 1.4.8 for more details about it):

$$H_{\sqrt{x}}(\rho||\sigma) = \frac{1}{2}\mathrm{Tr}\left[\sqrt{\rho}(\rho - \sigma)\sqrt{\sigma^{-1}}\right] , \tag{C.21}$$

which gives the expansion:

$$H_{\sqrt{x}}(\rho||\rho + \delta\rho) \simeq \frac{1}{2}\mathrm{Tr}\left[\delta\rho \, \mathbb{J}_{\sqrt{x}}^{-1}\big|_\rho[\delta\rho]\right] . \tag{C.22}$$

Thanks to condition 3, $\tilde{\Phi}\Phi$ is self-adjoint with respect to the scalar product induced by $\mathbb{J}_{\sqrt{x}}\big|_\pi$. In fact, one has:

$$\mathrm{Tr}\left[A \, \mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[\tilde{\Phi}\Phi(B)]\right] = \mathrm{Tr}\left[\mathbb{J}_{\sqrt{x}}\big|_\pi[(\tilde{\Phi}\Phi)^\dagger(\mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[A])] \, \mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[B]\right] = \tag{C.23}$$

$$= \mathrm{Tr}\left[\tilde{\Phi}\Phi(A) \, \mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[B]\right] , \tag{C.24}$$

and the same of course holds for the classical case by substituting $\mathbb{J}_{\sqrt{x}}\big|_\pi$ with $\mathcal{J}_\pi$ (in fact, this holds for all the calculations in this section). Indeed, as it was proved in Sec. 1.5.5 the condition of being detailed balance is equivalent to being self-adjoint with respect to the Fisher scalar product.

Then, thanks to the self-adjointness of $\tilde{\Phi}\Phi$ we can find an orthonormal basis $\{E_i\}$ such that $\tilde{\Phi}\Phi[E_i] := \varphi_i E_i$ and $\mathrm{Tr}\left[E_i \mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[E_j]\right] = \delta_{i,j}$. Moreover, due to condition 4, all the φ_i are positive and less then one (since $\tilde{\Phi}\Phi$ is a CP-map) meaning that we can express them as $\varphi_i = e^{-\lambda_i}$, where λ_i are all positive. Notice that the λs are also connected to the determinant of the forth-and-back map by the relation:

$$\log\det\tilde{\Phi}\Phi = \mathrm{Tr}\left[\log\tilde{\Phi}\Phi\right] = -\sum_i \lambda_i\,. \tag{C.25}$$

Consider now the modified contraction rate close to the prior state π:

$$\mu_g^F(\pi,\delta\rho) := \frac{H_g(\pi||\pi+\delta\rho) - H_g(\tilde{\Phi}\Phi(\pi)||\tilde{\Phi}\Phi(\pi+\delta\rho))}{H_g(\pi||\pi+\delta\rho)}, \tag{C.26}$$

where $H_g(\rho||\sigma)$ is a generic contrast function in the classical case, or the one in Eq. (C.21) in the quantum one. Thanks to the expansion in Eq. (C.22) and the self-adjointness of $\tilde{\Phi}\Phi$, we can rewrite the $\mu_g^F(\pi,\delta\rho)$ as:

$$\mu_g^F(\pi,\delta\rho) = \frac{\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{\sqrt{x}}\big|_\pi^{-1}[\delta\rho]\right] - \mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[(\tilde{\Phi}\Phi)^2(\delta\rho)]\right]}{\mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[\delta\rho]\right]} = \tag{C.27}$$

$$= \frac{\sum_i |\delta\rho_i|^2(1-e^{-2\lambda_i})}{\sum_i |\delta\rho_i|^2}, \tag{C.28}$$

where $\delta\rho_i := \mathrm{Tr}\left[\delta\rho\,\mathbb{J}_{\sqrt{x}}\big|_\pi^{-1}[E_i]\right]$ are the component of $\delta\rho$ in the eigenbasis of $\tilde{\Phi}\Phi$.

Using the inequality $1-e^{-x} \leq x$ holding for positive x (with equality only for $x \equiv 0$), we can bound the modified contraction rate by:

$$\mu_g^F(\pi,\delta\rho) \leq 2\,\frac{\sum_i |\delta\rho_i|^2\lambda_i}{\sum_i |\delta\rho_i|^2} \leq \tag{C.29}$$

$$\leq 2\sum_i \lambda_i = 2\log\det(\tilde{\Phi}\Phi)^{-1}\,. \tag{C.30}$$

Define now the two rates:

$$\mu_g^{RE} := \inf_{\rho,\sigma}\ \frac{H_g(\rho||\sigma) - H_g(\tilde{\Phi}\Phi(\rho)||\tilde{\Phi}\Phi(\sigma))}{H_g(\rho||\sigma)}\,; \tag{C.31}$$

$$\mu_g^F := \inf_{\rho,\delta\rho}\ \frac{H_g(\rho||\rho+\delta\rho) - H_g(\tilde{\Phi}\Phi(\rho)||\tilde{\Phi}\Phi(\rho+\delta\rho))}{H_g(\rho||\rho+\delta\rho)}\,, \tag{C.32}$$

which are related to the one discussed in Sec. 1.5.4 by $\mu_g = 1 - \eta_g$. Then, since it was argued in the main text that $\eta_g^{RE} \geq \eta_g^F$, it directly follows that $\mu_g^{RE} \leq \mu_g^F$. Using the standard rewriting:

$$\mu_g^F \equiv \inf_{\rho,\delta\rho} \mu_g^F(\rho, \delta\rho) \leq \mu_g^F(\pi, \delta\rho)\,, \tag{C.33}$$

we obtain the following chain of inequalities

$$\mu_g^{RE} \leq \mu_g^F \leq \mu_g^F(\pi, \delta\rho) \leq 2\log\det(\tilde{\Phi}\Phi)^{-1}\,. \tag{C.34}$$

Writing explicitly μ_g^{RE}, we finally get:

$$\inf_{\rho,\sigma} \frac{H_g(\rho||\sigma) - H_g(\tilde{\Phi}\Phi(\rho)||\tilde{\Phi}\Phi(\sigma))}{H_g(\rho||\sigma)} \leq 2\log\det(\tilde{\Phi}\Phi)^{-1}\,. \tag{C.35}$$

Hence, the inverse of the determinant bounds the maximum rate at which any two states become indistinguishable (also see Sec. 1.5.4 for discussions of similar quantities). Since this quantifies the velocity at which information is lost, maximising the determinant slows down the irreversible flow of information out of the system.

Another interesting inequality can be obtained with analogous methods: consider, in fact, the ability of $\tilde{\Phi}\Phi$ to retrieve states close to the prior as quantified by the contrast functions H_g. Then, we can rewrite this quantity as:

$$H_g(\pi + \delta\rho||\tilde{\Phi}\Phi(\pi + \delta\rho)) = \frac{1}{2}\mathrm{Tr}\left[\delta\rho\, \mathbb{J}_{\sqrt{x}}^{-1}\big|_\pi[(\mathbb{1} - \tilde{\Phi}\Phi)^2(\delta\rho)]\right] = \tag{C.36}$$

$$= \frac{1}{2}\sum_i |\delta\rho_i|^2(1 - e^{-\lambda_i})^2\,. \tag{C.37}$$

Thanks to the inequality $(1 - e^{-x})^2 \leq x/2$ holding for positive x, it directly follows that:

$$H_g(\pi + \delta\rho||\tilde{\Phi}\Phi(\pi + \delta\rho)) \leq \frac{1}{4}\sum_i |\delta\rho_i|^2\lambda_i \leq \frac{H_g(\pi||\pi + \delta\rho)}{2}\log\det(\tilde{\Phi}\Phi)^{-1}\,. \tag{C.38}$$

This tells us that the determinant bounds the ability of retrieving states close to the prior. Arguably this is the most relevant situation, for example, in error correction, in which one wants to correct the effect on the noise especially for states close to the fiducial state π.

We stress again that the computations leading to Eq. (C.35) and Eq. (C.38) only hold in the quantum case for the contrast function in Eq. (C.21). Still, if the condition 3 gets modified with the requirement that $\tilde{\Phi}\Phi$ satisfies the canonical definition of detailed balance (see Sec. 1.5.5) all the steps can be generalised to any quantum contrast function. Despite this, since all the other discussions hold also for the weaker definition of detailed balance (the one with respect to $\mathbb{J}_{\sqrt{x}}$ only) we don't have strong enough reasons to make condition 3 more stringent.

In this section we have presented some analytical insights justifying the max-det principle (namely Eq. (C.18), (C.35) and Eq. (C.38)). To further motivate this choice, in the next section we compare its performance in state retrieving with the Bayes' and Petz maps.

C.2.2 Quality of the retrieval

Before studying two particular cases, we give an emblematic example of how the Petz recovery map (or the Bayes' reverse channel) seems to deviate from the intuitive notion of state retrieval map. In particular, we have the following:

Theorem 27. *Whenever a transformation Φ is detailed balanced with respect to the prior state (i.e, $\Phi\,\mathbb{J}_{\sqrt{x}}\big|_{\pi} = \mathbb{J}_{\sqrt{x}}\big|_{\pi}\,\Phi^{\dagger}$) and has positive spectrum, we have the following:*

1. *the Petz recovery map coincides with Φ;*
2. *the optimal state retrieval map is the identity operator $\mathbb{I}$.*

The proof of this fact is straightforward. In fact, from the definition of the Petz recovery map we obtain:

$$\tilde{\Phi}_P = \mathbb{J}_{\sqrt{x}}\big|_{\pi}\,\Phi^{\dagger}\,\mathbb{J}^{-1}_{\sqrt{x}}\big|_{\pi} = \Phi\,, \tag{C.39}$$

(notice that we used the fact that π is a fixed point of the dynamics), while for the second half of the statement it is sufficient to notice that:

$$\det(\tilde{\Phi}\Phi) \leq \det(\Phi)\,, \tag{C.40}$$

with equality if and only if $\tilde{\Phi}$ is a unitary transformation. Thanks to the assumptions in the Theorem, condition 3 and 4 are automatically satisfied for $\tilde{\Phi} = \mathbb{I}$, and thanks to the inequality in Eq. (C.40) this choice is indeed optimal.

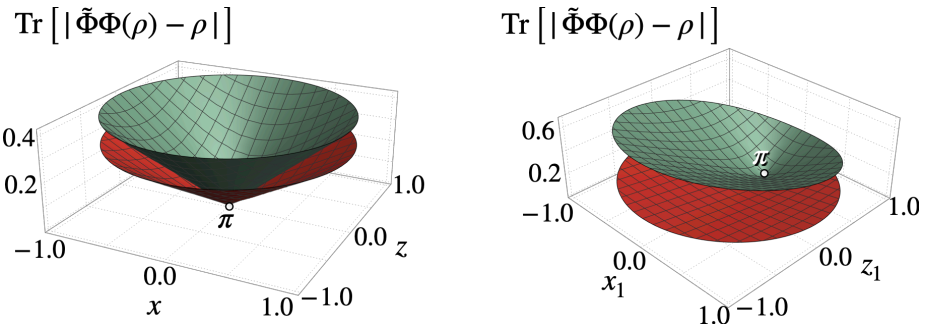


Figure C.2: Comparison between the quality of retrieval using the Petz recovery channel (in green) and the optimal state retrieval (in red) as measured by the trace distance. On the left, we choose a depolarising channel Δ_η acting on a qubit and x and z are the coordinates in the Bloch sphere. On the right, we use the channel defined in Eq. (C.42), where x_1 and z_1 are the Bloch coordinates on the first qubit, while the second is set to $\pi_\beta(H)$. The choice of the trace distance as a quantifier is simply aesthetic, and we observed similar behaviour using other statistical divergences.

We exemplify this result on the left panel of Fig. C.2. We considered the depolarising channel acting on a qubit:

$$\Delta_\eta(\rho) = (1-\eta)\,\rho + \eta\,\frac{\mathbb{1}}{2}\,, \tag{C.41}$$

where η is a scalar parameter in $[0, 4/3]$, and we chose $\mathbb{1}/2$ as the prior state. Then, it is easy to verify that all of the assumptions of the Thm. 27 are satisfied. Hence, in this case, the optimal strategy is to leave the system unperturbed. On the other hand, using the Petz recovery map corresponds to another application of Δ_η, leading to a further deterioration of the information about the initial state. This shows how our definition of optimal retrieval is more suited in the task of recovering a state after a transformation.

The right panel of Fig. C.2 shows the difference in performance of the optimal map and the Petz' one for the two-qubit channel defined by:

$$\Phi(\rho_A \otimes \rho_B) = \theta_{\lambda_1}(\rho_B) \otimes \theta_{\lambda_2}(\rho_A), \tag{C.42}$$

where θ_λ is the thermalising channel on a qubit defined by:

$$\theta_\lambda(\rho) = (1-\lambda)\rho + \lambda \mathrm{Tr}\,[\rho]\,\pi_\beta(H) \tag{C.43}$$

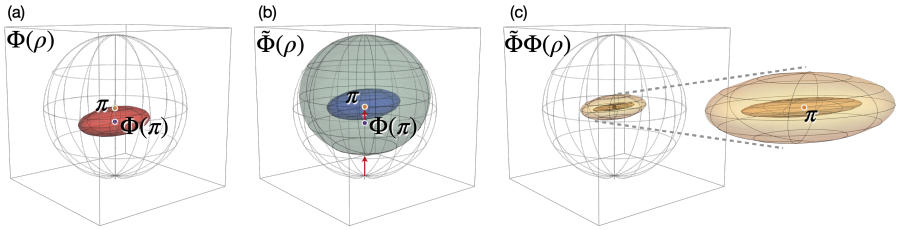


Figure C.3: Geometrical representation of the action of optimal maps for qubits. In panel (a) the action of the map Φ defined as a contraction of the Bloch sphere, composed with a translation. In panel (b) in blue we represent the Petz recovery map, while in light green the optimal state retrieval. As it can be noticed, the latter corresponds to a uniform translation bringing $\Phi(\pi)$ to π, together with the minimal contraction that allows to perform this operation physically. Finally, in panel (c) in light yellow the image of $\tilde{\Phi}_O\Phi$ and in orange the one of $\tilde{\Phi}_P\Phi$. As it can be noticed, there is a clear gap in the volume of states that are covered by the two different maps.

and $\pi_\beta(H)$ is the thermal state associated to the Hamiltonian $H := \epsilon\,|1\rangle\langle 1|$. Taking the prior state to be $\pi := \pi_\beta(H)\otimes\pi_\beta(H)$ allows to solve the problem analytically. A simple calculations shows that, even in this case, the Petz map coincides with the original channel, i.e., $\tilde{\Phi}_P = \Phi$. On the other hand, the optimal state retrieval is given by $\tilde{\Phi}_O = \text{SWAP}$, the swap operator. Again we see that the action of the optimal state retrieval is intuitively what we expect: since the dynamics is completely depolarising, the maximum one can achieve is obtained by undoing the swap between the two systems.

So far the optimal maps have been always unital. In Fig. C.3 we present an example in which this is not the case. Consider a map on the Bloch sphere which acts as the composition of a contraction and a translation. We choose the prior state to be given by $\pi = \frac{\mathbb{1}}{2}$, so that $\tilde{\Phi}\Phi$ is unital even if Φ is not (this technical point is needed for the numerical optimisation we implemented, which made use of the parametrisation given in [114]). Then, the action of the Petz map and of the optimal state retrieval are presented in the Fig. C.3.(b). This offers a neat geometrical interpretation of Φ_O: this is given by the composition of a compression and a translation specified as follows: the translation is the one recovering the desired prior π (the red arrow in the picture) and the compression is the minimal one making $\tilde{\Phi}_O$ physical (i.e., so that the image of $\tilde{\Phi}_O$ is

contained in the Bloch sphere). As one can see from Fig. C.3.(c) indeed the image of $\tilde{\Phi}_O\Phi$ is considerably bigger than the one for $\tilde{\Phi}_P\Phi$.

C.3 Isolation of Bayes' inverse

In this section we discuss the evidences suggesting that the extra condition:

5. the reversal is an involutive operations, meaning that $\tilde{\tilde{\Phi}} = \Phi$,

together with the other four, allows to single out the Bayes' one as unique inverse map. In particular, we assume that the action of $\mathcal{R}$ to be specified only by its action on the vertices $V^{(k)}_{\Phi\pi|\pi}$ of the convex set $\mathcal{U}(\Phi\pi, \pi)$. Notice that in this context $\Phi\pi$ denotes a vector, and has little to do with the original map.

Then, since $\mathcal{U}(\Phi\pi, \pi)$ and $\mathcal{U}(\pi, \Phi\pi)$ are connected by transposition, we will identify $V^{(k)}_{\Phi\pi|\pi}$ with $V^{(\tilde{k})}_{\pi|\Phi\pi} = (V^{(k)}_{\Phi\pi|\pi})^T$ so that $\mathcal{R}$ is a map from a space into itself. Moreover, doing so shows that the $\mathcal{R}$ corresponding to Bayes' inversion is the identity matrix, since, as it was noticed in Sec. C.1.1, Bayes' inversion corresponds to a transposition on the space $\mathcal{U}(\Phi\pi, \pi)$.

Then, we can divide the vertices of $\mathcal{U}(\Phi\pi, \pi)$ in two groups: the ones corresponding to permutations (in the sense that $V^{(k)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-1}$ is a permutation), and all the others. Thanks to condition 2, on the vertices of the first kind any reverse map gives the identity, as:

$$(V^{(k)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-1})^{-1} = (V^{(k)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-1})^T = (\mathcal{J}_\pi^{-1}(V^{(k)}_{\Phi\pi|\pi})^T\mathcal{J}_{\Phi(\pi)})\mathcal{J}^{-1}_{\Phi(\pi)} = \quad \text{(C.44)}$$

$$= ((V^{(k)}_{\Phi\pi|\pi})^T\mathcal{J}^{-1}_{\Phi(\pi)}\mathcal{J}_{\Phi(\pi)})\mathcal{J}^{-1}_{\Phi(\pi)} = (V^{(k)}_{\Phi\pi|\pi})^T\mathcal{J}^{-1}_{\Phi(\pi)}\,. \quad \text{(C.45)}$$

Then, it follows that $\mathcal{R}$ is the direct sum of the identity matrix acting on the first ℓ sites and a permutation matrix with cycles of maximal length 2 acting on sites $\ell+1, \ldots, n$. Indeed, thanks to the structure of $\mathcal{U}(\Phi\pi, \pi)$, one can interpret the coefficients $\{\lambda^{(\Phi)}_k\}$ as a probability vector. Thus $\mathcal{R}$ must map probability distributions into probability distributions, meaning that $\mathcal{R}$ is a stochastic matrix. Moreover, axiom 5 implies $\mathcal{R}^2 = \mathbb{1}$, meaning that $\mathcal{R}$ is invertible and coincides with its inverse. It should be noticed that all the invertible stochastic matrices are permutations. The involutive condition then also implies that it must be a permutation of cycle at most 2. We can now focus on the action of $\mathcal{R}$ on the first ℓ

indices. From condition 2 we know that permutations must be mapped into their inverse, that is $U \to U^T$. Thanks to the relation between the vertices of $\mathcal{U}(\Phi\pi, \pi)$ and $\mathcal{U}(\pi, \Phi\pi)$ this corresponds to $\mathcal{R}$ acting as the identity on the first ℓ elements of $\{\lambda_k^{(\Phi)}\}$.

Since it is a stochastic matrix, in order to study $\mathcal{R}$, it is sufficient to study its action on the vertices of the simplex of the probability vectors $\{\lambda_k^{(\Phi)}\}$. In particular we need to check if there is any permutation of two vertices of this simplex that is admissible other than the identity.

We focus on the action of $\mathcal{R}$ on single vertices. Consider in particular the case in which $\Phi := V^{(i)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-1}$. Since $\mathcal{R}$ is a permutation, there exists a vertex $V^{(j)}_{\pi|\Phi\pi}$ which satisfies $\tilde{\Phi} \equiv V^{(j)}_{\pi|\Phi\pi}\mathcal{J}^{-1}_{\Phi(\pi)} = (V^{(j)}_{\Phi\pi|\pi})^T\mathcal{J}^{-1}_{\Phi(\pi)}$. From principles (3 and 4 the following matrix

$$X_{i,j} = \mathcal{J}_\pi^{-\frac{1}{2}}(V^{(j)}_{\Phi\pi|\pi})^T\mathcal{J}^{-1}_{\Phi(\pi)}V^{(i)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-\frac{1}{2}} \tag{C.46}$$

is positive semidefinite.

At the same time, due to principle 5, if the vertex $(V^{(j)}_{\Phi\pi|\pi})^T$ corresponds to the inverse of $V^{(i)}_{\Phi\pi|\pi}$, then $V^{(i)}_{\Phi\pi|\pi}$ must be the inverse of $(V^{(j)}_{\Phi\pi|\pi})^T$. This consideration, together with conditions 3 and 4, then also implies that the matrix

$$Y_{i,j} = \mathcal{J}^{-\frac{1}{2}}_{\Phi(\pi)}V^{(i)}_{\Phi\pi|\pi}\mathcal{J}_\pi^{-1}(V^{(j)}_{\Phi\pi|\pi})^T\mathcal{J}^{-\frac{1}{2}}_{\Phi(\pi)} \tag{C.47}$$

is positive semidefinite.

Since the number of vertices is finite, it is easy to explicitly verify for which set of indices Eq. (C.46) and Eq. (C.47) are positive semidefinite. We verified this for many possible families of stochastic maps that the only admissible $\mathcal{R}$ is the identity, meaning that axiom 5 seems to be enough to single out the Bayes reversion (see Fig. C.4). Despite this promising result, an analytical proof of this fact is still missing. In fact we miss a characterisation of the properties of the vertices for generic $\mathcal{U}(\Phi\pi, \pi)$. To the best of our knowledge, for an arbitrary pair $(\Phi\pi, \pi)$, it is not even possible to know the precise number of vertices of the set $\mathcal{U}(\Phi\pi, \pi)$ without first mechanically constructing them using the algorithm of Jurkat and Ryser [111].

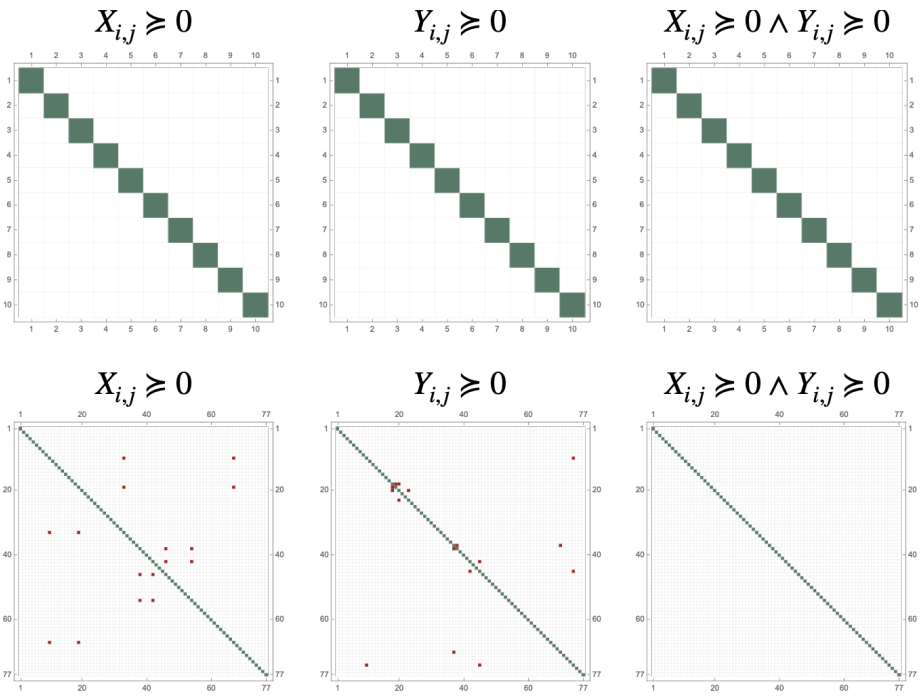


Figure C.4: In order to check if $\mathcal{R}$ can be any a permutation different from the identity one checks which of the matrices $X_{i,j}$ and $Y_{i,j}$ are Positive SemiDefinite (PSD). If, for a fixed value of i and j, both the matrix $X_{i,j}$ and $Y_{i,j}$ are PSD, then a $\mathcal{R}$ permuting i and j is allowed. In the first line we choose $\pi = (0.1, 0.2, 0.7)$ and $\Phi\pi = (0.3, 0.6, 0.1)$, while in the second line $\pi = (0.1, 0.6, 0.1, 0.2)$ and $\Phi\pi = (0.1, 0.2, 0.3, 0.4)$. The elements in the grid are green if the corresponding matrix is PSD and on the diagonal, red if it PSD but out of the diagonal, and white otherwise. As it can be seen, in the first line only requiring that $X_{i,j}$ or $Y_{i,j}$ are positive semidefinite is enough to single out the diagonal, corresponding to the identity transformation, that is, as it was explained in the text, to the Bayes inversion. In the second line we see that one actually needs to ask that both $X_{i,j}$ and $Y_{i,j}$ are PSD, because there are cases in which one is PSD but the other is not. When one considers the two conditions together, again the only possibility left is the one of Bayes inversion.

Appendix D

Elements of Kolmogorov complexity

Kolmogorov's most famous contribution to modern mathematics is the formalisation of probability theory in terms of an axiomatic construction. It might be surprising, then, that among the many other results and theories that he produced, mostly in the context of stochastic processes and dynamical systems, one can find his name in a small fringe of mathematics, the theory of algorithmic complexity. Indeed, in this context the Kolmogorov complexity of a string is defined as the shortest program that can generate it. This quantity is apparently so far away from the other areas that he had explored, that it appears more as a mathematical curiosity than a serious pursuit. This could not be further from the truth: in fact, rather than being the result of an erratic exploration, the central aim motivating his study of algorithmic complexity is to justify the applicability of probability theory to the real world. To say it with his words (as cited in [66]):

> *This theory was so successful, that the problem of finding the basis of real applications of the results of the mathematical theory of probability became rather secondary to many investigators. [...However,] the basis for the applicability of the results of the mathematical theory of probability to real "random phenomena" must depend in some form on the frequency concept of probability, the unavoidable nature of which has been established by von Mises in a spirited manner.*

The main foundational problem here can be explained as follows: say

one tosses 100 times a fair coin, obtaining 100 times head; this is a perfectly normal output in probability theory, albeit of exponentially small probability. Still, one would be quite surprised to see something like this happening, and would rather think that the original coin is not fair, than to accept this string as a mere improbable event. Hence, there must be some notion of randomness that applies to single sequences (and not to sets), to which we intuitively refer to when talking of probability.

The key aspect captured by Kolmogorov complexity is the difficulty of finding any regularity in random sequences. Indeed, whereas a uniform sequence of heads and tails seems perfectly normal, having instances in which only one outcome is repeated many times (as in the example above) seem rather suspicious. What makes the definition of Kolmogorov quite remarkable, is the intuition connecting the presence of regularities to the existence of a short description: indeed, while exponentially improbable outcomes can be easily characterised (e.g., 100 times head), for more probable strings one does not have any possibility of describing it other than writing it down explicitly digit by digit.

This idea was made precise in the work of Martin-Löf, and we refer to [66] for an in-depth investigation of the matter. In the following, we will just briefly provide the most technical aspects of Kolmogorov complexity that are used in Sec. 3.1.2 and Sec. 3.2.4. Still, even if we do not dwell too much on the connection between randomness, probability theory and algorithmic complexity, one should keep it in mind as the fundamental raison d'être of the subject.

D.1 Main definitions and properties

We will assume the reader to be familiar with the notion of Turing machine (TM) and universal Turing machine (UTM). These are formalisations of the intuitive notion of program, i.e., a computation following a finite set of instructions. In this context, with universality one refers to the ability of UTMs of simulating any other Turing machine [115]. One can also think of TMs as the program that one uses to define them on a standard UTM, for example their expression in Python for modern computers. This allows to give an enumeration of all TMs: in this example, by interpreting the programs in Python as integers (by associating them to their expression in binary code). Having an enumeration of all possible TMs shows that not all questions that can be formulated, can also be answered in finite time. Indeed, the first example of this kind was the

famous halting problem, that is the question about whether an arbitrary TM will halt or not. For some TMs, e.g., the ones that won't compile, this decision problem is straightforward. Still, there is no universal method deciding all the instances of this question. Indeed, suppose there were such a program, call it H, that outputs 1 if the TM on which it is run halts, and 0 otherwise. Then one could construct a TM $\tilde{T}$ that loops for ever if $H(\tilde{T}) = 1$ and halts if $H(\tilde{T}) = 0$. This is clearly a contradiction. Hence, H cannot give a solution in finite time on every input.

This leads to the definition of recursive enumerability: a set is recursive enumerable if its member are outputted by some Turing Machine in infinite time. One can associate to each recursive set a binary function deciding the membership of a given input to the set. In this case, the function either halts or runs for ever, and there is no way of deciding which one of the two cases it is. On the other hand, a function is called recursive if it halts on all inputs.

Before passing to the definition of algorithmic complexity, it is useful to introduce some more notation. First, it should be noticed that the set of strings has a natural order, the lexicographic partial order, which we denote by $<_\ell$ and is defined as:

$$y <_\ell x \qquad \Longleftrightarrow \qquad \exists z \mid x = yz\,. \tag{D.1}$$

A set of strings S is called prefix free, if for any two $x, y \in S$, $y <_\ell x$ if and only if $y = x$. We choose a standard encoding E from the set of binary strings to some infinite prefix free set, and denote by $\bar{x} := E(x)$. We also require that in order for E to be a proper encoding, it should satisfy $\ell^*(x) := \ell(\bar{x}) \leq \ell(x) + 2\log \ell(x)$, where $\ell(x)$ is the length of the string. Finally, we will also use the notation $T(x) < \infty$ to indicate that the Turing machine T halts on input x.

In the following, rather than normal TMs, we will focus on Chaitin computers. These are defined as follows: consider a UTM U with two inputs, one indicating the program p that it should use, the other the input x on which to run p (i.e, the UTM acts as $U(p, x) = p(x)$). Then, a universal Turing machine C is a Chaitin computer if the domain of definition of the TM defined as $C_x(p) := C(p, x)$ is prefix free [116]. This restriction implies that the programs read by C_x are self-delimiting: in fact, for every p such that $C_x(p) < \infty$, if $p <_\ell q$ and $C_x(q) < \infty$, this implies that $p \equiv q$, because otherwise the domain of C_x would not be prefix free. Equivalently, the information about when to stop reading p is already contained in itself, without the need for C_x to scan any additional

end-markers.

We fix a universal Chaitin computer C to be the reference machine with respect to which all the following definitions are given. Then, we can define the Kolmolgorov prefix complexity as [66]:

$$K(x|y) := \min \{\ell^*(p)|\, C(p,y) = x\} \ . \tag{D.2}$$

For each string x there exists an uncomputable string $x^*_{|y}$ defined to be the shortest program to output x on input y, chosen in lexicographic order. Then, $K(x|y) = \ell(x^*_{|y})$. We also define $K(x) := K(x|\varepsilon)$, where ε is the empty string, and $x^* := x^*_{|\varepsilon}$.

It is easy to see that $K(x|y)$ or, equivalently, $x^*_{|y}$ are not recursive. Indeed, suppose the contrary, i.e., that there were a program $F_{|y}$ associating to each string its algorithmic complexity. Then, consider the following program $\tilde{F}_{|y}$: on input a natural number, it outputs the first string (in lexicographic order) with that conditional complexity. Obviously $K(\tilde{F}_{|y}(n)|y) = n$. But, then, for any $n = K(\tilde{F}_{|y}(n)|y) > \ell^*(\tilde{F}_{|y})$, $\tilde{F}_{|y}$ would encode the number $\tilde{F}_{|y}(n)$ better than the optimal program, generating a contradiction. Hence, $\tilde{F}_{|y}$ is not computable[1].

It should be noticed that changing the reference machine C to C' changes $K(x|y)$ by a constant. Indeed, given the optimal program $x^*_{|y}$ for C, C' can simulate it with a overhead of $\ell^*(C)$. The same argument can be made swapping C and C', so one writes:

$$K_C(x|y) = K_{C'}(x|y) + \mathcal{O}\,(1) \ , \tag{D.3}$$

where we highlighted the dependency of the algorithmic complexity on the defining Chaitin computer. Eq. (D.3) takes the name of invariance theorem, which shows that one can refer to $K(x|y)$ as a well-defined quantity, as it does not depend drastically on its defining machine. This comes at the cost that all the equality that we will write in the following are actually to be understood up to a constant factor (i.e., as in Eq. (D.3)).

There are a number of interesting properties that the Kolmogorov complexity satisfies. Consider for example the quantity $K(x, f(x))$, i.e., of the vector $\{x, f(x)\}$, where we assume that f is a computable function. Then, it is easy to modify the reference machine as: $\tilde{C}(p,y) :=$

[1]The very same concept is at the root of Berry paradox. This can be expressed as following: "consider the smallest integer not definable in less than 12 words". It is easy to see that no such integer can exists, making the notion of integers definable in less than n words problematic.

$\{p(y), f(p(y))\}$. Then, thanks to the invariance theorem, this means that:

$$K(x, f(x)) = K(x)\,, \tag{D.4}$$

where we remind the reader that this equality is defined up to a constant additive term. Moreover, for the same reason if $p(y) = x$ then it also holds that:

$$K(p, y) = K(x, y)\,, \tag{D.5}$$

since in this case one can redefine the reference machine to first run on some input to generate $\{p, y\}$, and then it applies the first argument to the second, i.e., $\{p(y), y\} = \{x, y\}$.

Interestingly, the same holds also for $K(x, K(x))$, even if K is not a computable function. In this case the reason for this follows from the minimality of $K(x)$. Indeed, it is easy to show that $K(x, y) \geq K(x)$, as one can give a program for x starting from one for $\{x, y\}$ with just a constant overhead, but not the other way around. Moreover, one can generate $\{x, K(x)\}$ from the minimal program x^* by changing the reference machine to $\tilde{C}(p, y) := \{p(y), \ell(p)\}$. This shows that $K(x, K(x)) \leq K(x)$ which, together with the inequality above gives:

$$K(x, K(x)) = K(x)\,. \tag{D.6}$$

Considering the conditional Kolmogorov complexity, it is straightforward to verify that:

$$K(x|y) \leq K(x)\,, \tag{D.7}$$

since the set of programs on which one takes the minimum in the first case is at least as large as the one for the unconditional case (as one can construct a $\tilde{C}$ that always disregard any extra inputs). Similarly $K(x|y_1, \ldots, y_n) \leq K(x|y_1, \ldots, y_{n-1})$.

An interesting property of the Kolmogorov prefix complexity which does not hold if one drops the request that the domain of C is prefix free is the subadditivity, i.e.,

$$K(x, y) \leq K(x) + K(y)\,. \tag{D.8}$$

The proof of this result can be found in [66]. This property is in common with the Shannon entropy, but it is not the only connection between

the two. Indeed, one can prove that given a computable probability distribution $p(x)$ such that $H(p) < \infty$, then it holds that [66]:

$$0 \leq \sum_x p(x) \left(K(x) - H(p)\right) \leq c_p \,, \tag{D.9}$$

where $c_p = K(p) + \mathcal{O}(1)$ only depends on p. It is also interesting to point out that part of the proof is based on the noiseless coding theorem. Moreover, the relation with Shannon theory is even stronger as it can be verified considering the following scenario: consider an ensemble of strings $\{s_i\}$ distributed according to $\{p(s_i)\}$, then it holds that [66]:

$$H(\{p(s_i)\}) \leq \sum_i p(s_i) K(s_i) \leq H(\{p(s_i)\}) + K(\{s_i, p(s_i)\}) \,, \tag{D.10}$$

where $K(\{s_i, p(s_i)\})$ is the Kolmogorov complexity of the ensemble.

The similarity between Kolmogorov complexity and Shannon theory motivated Chaitin to his study of the algorithmic information theory [117]. Still, an important property of the Shannon entropy, namely the symmetry of information does not hold for Kolmogorov complexity. Indeed, one has that:

$$K(x, y) = K(x|y) + K(y) \tag{D.11}$$

only holds up to logarithmic corrections in $K(y)$. Still, using the coding theorem (Thm. 28) it is possible to show that:

$$K(x, y) = K(x|y^*) + K(y) = K(y|x^*) + K(x) \,, \tag{D.12}$$

within an additive constant. It should be noticed that we use the notation $K(x|y^*) := K(x|y, y^*)$, from which it trivially follows that $K(x|y^*) \leq K(x|y)$ (see Eq. (D.7)).

Finally, we hint at the definition of randomness in this context. A string is considered random, or incompressible, if $K(x) \geq \ell(x)$. It should be noticed that any minimal programs x^* are random, as otherwise $K(K(x)) \leq \ell^*(x^*) = K(x)$, and one could use this more concise encoding to simulate x with a program shorter than $K(x)$. This also means that $x^{**} = x^*$, i.e., the best description of the minimal program is the minimal program itself (run on a machine that just returns the input). This argument was implicitly used in the analysis of the algorithmic Maxwell's demon in Sec. 3.1.2.

D.2 Coding theorem

Another important property that makes the Kolmogorov prefix complexity particularly desirable compared to other definition of algorithmic complexity is that it can be used to define a prefix free code. Indeed, one can associate to each program the corresponding output. For any universal Chaitin computer U, this is indeed a prefix free code, so it follows from the Kraft inequality that:

$$\Omega_C := \sum_{p\,|U(p)<\infty} 2^{-\ell^*(p)} < 1\,, \tag{D.13}$$

It should be noticed that Ω_C, called Chaitin's constant, is less than one since there are programs that do not halt.

This discussion suggests the definition of the following semimeasure: suppose one has a fair coin, that sequentially flips to generate a program p. Running it on a standard Chaitin computer on input x, one obtains a string $p(x)$ (or it loops for ever). Then, one can define the transition probability of starting from x and ending in y as:

$$Q_U(x \to y) := \sum_{p|\,U(p,x)=y} 2^{-\ell^*(p)}\,. \tag{D.14}$$

Thanks to Kraft inequality Q_U is indeed a semimeasure. This quantity is connected to the Kolmogorov complexity by one of the most important results in the subject, the coding theorem:

Theorem 28. *There is a constant c_U that only depends on the choice of universal Chaitin computer U such that:*

$$K(y|x) = -\log Q_U(x \to y) + c_U\,. \tag{D.15}$$

The theorem above tells us that the probability of randomly obtaining y starting from x is within a multiplicative constant of $2^{-K(y|x)}$. This remarkable result was used in Sec. 3.2.4 to define algorithmic fluctuation relations.

This concludes the brief summary of the essential ingredients of algorithmic complexity used in the main text. Once again we refer the interested reader to [66] to explore this vast and interesting subject.

Bibliography

[1] M. Scandi, P. Abiuso, D. De Santis, and J. Surace, "Quantum Fisher information and its dynamical nature," In preparation, (2023).

[2] J. Surace and M. Scandi, "State retrieval beyond Bayes' retrodiction," Quantum **7**, 990 (2023).

[3] M. Scandi, D. Barker, S. Lehmann, K. A. Dick, V. F. Maisi, and M. Perarnau-Llobet, "Minimally dissipative information erasure in a quantum dot via thermodynamic length," Physical Review Letters **129**, 270601 (2022).

[4] P. Abiuso, M. Scandi, D. De Santis, and J. Surace, "Characterizing (non-)Markovianity through Fisher information," (2022), 10.48550/arXiv.2204.04072.

[5] D. Barker, M. Scandi, S. Lehmann, C. Thelander, K. A. Dick, M. Perarnau-Llobet, and V. F. Maisi, "Experimental verification of the work fluctuation-dissipation relation for information-to-work conversion," Physical Review Letters **128**, 040602 (2022).

[6] P. Abiuso, H. J. D. Miller, M. Perarnau-Llobet, and M. Scandi, "Geometric optimisation of quantum thermodynamic processes," Entropy **22**, 1076 (2020).

[7] M. Scandi, H. J. D. Miller, J. Anders, and M. Perarnau-Llobet, "Quantum work statistics close to equilibrium," Physical Review Research **2**, 023377 (2020).

[8] M. Scandi, H. J. D. Miller, J. Anders, and M. Perarnau-Llobet, "Work fluctuations in slow processes: quantum signatures and optimal control," Physical Review Letters **123**, 230603 (2019).

[9] M. Scandi and M. Perarnau-Llobet, "Thermodynamic length in open quantum systems," Quantum **3**, 197 (2019).

[10] M. Scandi and J. Surace, "Undecidability in resource theory: Can you tell resource theories apart?" Physical Review Letters **127**, 270501 (2021).

[11] N. Pancotti, M. Scandi, M. T. Mitchison, and M. Perarnau-Llobet, "Speed-ups to isothermality: Enhanced quantum thermal machines through control of the system-bath coupling," Physical Review X **10**, 031015 (2020).

[12] I. Csiszár and P. C. Shields, *Information theory and statistics: a tutorial*, Vol. 1 (Now Publishers, Inc., 2004) pp. 417–528.

[13] I. Bengtsson and K. Zyczkowski, *Geometry of quantum states: An Introduction to quantum entanglement*, 2nd ed. (Cambridge University Press, Cambridge, 2017).

[14] N. N. Chentsov, *Statistical decision rules and optimal inference* (American Mathematical Soc., 2000).

[15] M. Bauer, P. W. Michor, and O. Müller, "Riemannian geometry of the space of volume preserving immersions," Differential Geometry and its Applications **49**, 23–42 (2016).

[16] A. Lesniewski and M. B. Ruskai, "Monotone Riemannian metrics and relative entropy on non-commutative probability spaces," Journal of Mathematical Physics **40**, 5702–5724 (1999).

[17] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (American Association of Physics Teachers, 2002).

[18] F. Hiai and D. Petz, *Introduction to matrix analysis and applications*, Universitext (Springer, 2014).

[19] D. Petz, "Monotone metrics on matrix spaces," Linear Algebra and its Applications **244**, 81–96 (1996).

[20] F. Hansen, "Metric adjusted skew information," Proceedings of the National Academy of Sciences **105**, 9909–9916 (2008).

[21] F. Hiai, H. Kosaki, D. Petz, and M. B. Ruskai, "Families of completely positive maps associated with monotone metrics," Linear Algebra and its Applications **439**, 1749–1791 (2013).

[22] M. Reed and B. Simon, *Functional analysis II: Fourier analysis, self-adjointness*, Methods of Modern Mathematical Physics (Elsevier Science, 1975).

[23] F. Hiai and H. Kosaki, "Means for matrices and comparison of their norms," Indiana University Mathematics Journal , 899–936 (1999).

[24] A. Uhlmann, "Geometric phases and related structures," Reports on Mathematical Physics **36**, 461–481 (1995).

[25] A. Uhlmann, "Spheres and hemispheres as quantum state spaces," Journal of Geometry and Physics **18**, 76–92 (1996).

[26] D. Petz and C. Ghinea, "Introduction to quantum Fisher information," Quantum Probability and Related Topics , 261–281 (2011), 1008.2417 .

[27] E. P. Wigner and M. M. Yanase, "Information contents of distributions," Proceedings of the National Academy of Sciences of the United States of America **49**, 910–918 (1963).

[28] A. Besenyei and D. Petz, "Completely positive mappings and mean matrices," Linear algebra and its applications **435**, 984–997 (2011).

[29] P. Gibilisco and T. Isola, "Wigner–Yanase information on quantum state space: The geometric approach," Journal of Mathematical Physics **44**, 3752–3762 (2003).

[30] S. Friedland and W. So, "On the product of matrix exponentials," Linear algebra and its applications **196**, 193–205 (1994).

[31] K. M. Audenaert, J. Calsamiglia, R. Munoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating states: The quantum Chernoff bound," Physical Review Letters **98**, 160501 (2007).

[32] F. Hiai and M. B. Ruskai, "Contraction coefficients for noisy quantum channels," Journal of Mathematical Physics **57**, 015211 (2016).

[33] D. Petz and G. Toth, "The Bogoliubov inner product in quantum statistics: Dedicated to J. Merza on his 60th birthday," Letters in Mathematical Physics **27**, 205–216 (1993).

[34] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," IEEE Transactions on Information Theory **59**, 7693–7710 (2013).

[35] Á. Rivas, S. F. Huelga, and M. B. Plenio, "Quantum non-Markovianity: characterization, quantification and detection," Reports on Progress in Physics **77**, 094001 (2014).

[36] M.-D. Choi, "Completely positive linear maps on complex matrices," Linear Algebra and its Applications **10**, 285–290 (1975).

[37] E. A. Carlen and J. Maas, "Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance," Journal of Functional Analysis **273**, 1810–1869 (2017).

[38] X.-M. Lu, X. Wang, and C. P. Sun, "Quantum Fisher information flow and non-Markovian processes of open systems," Physical Review A **82**, 042103 (2010).

[39] B. Bylicka, M. Johansson, and A. Acín, "Constructive method for detecting the information backflow of non-Markovian dynamics," Physical Review Letters **118**, 120501 (2017).

[40] B. O'Neill, *Semi-Riemannian geometry with applications to relativity*, Pure and Applied Mathematics (Elsevier Science, 1983).

[41] R. Alicki, "On the detailed balance condition for non-Hamiltonian systems," Reports on Mathematical Physics **10**, 249–258 (1976).

[42] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete, "The χ^2-divergence and mixing times of quantum Markov processes," Journal of Mathematical Physics **51**, 122201 (2010).

[43] A. Ferraro, A. García-Saez, and A. Acín, "Intensive temperature and quantum correlations for refined quantum measurements," Europhysics Letters **98**, 10009 (2012).

[44] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, "Locality of temperature," Physical Review X **4**, 031019 (2014).

[45] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert, "Strong coupling corrections in quantum thermodynamics," Physical Review Letters **120** (2018).

[46] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," Reports on Progress in Physics **75**, 126001 (2012).

[47] A. E. Allahverdyan and T. M. Nieuwenhuizen, "Fluctuations of work from quantum subensembles: The case against quantum work-fluctuation theorems," Physical Review E **71**, 066102 (2005).

[48] R. Resta, "Manifestations of Berry phase in molecules and condensed matter," Journal of Physics: Condensed Matter **12**, R107–R143 (2000).

[49] P. Talkner, E. Lutz, and P. Hänggi, "Fluctuation theorems: Work is not an observable," Physical Review E **75** (2007), 10.1103/PhysRevE.75.050102.

[50] C. Jarzynski, H. T. Quan, and S. Rahav, "Quantum-classical correspondence principle for work distributions," Physical Review X **5**, 031038 (2015).

[51] M. Perarnau-Llobet, E. Bäumer, K. V. Hovhannisyan, M. Huber, and A. Acin, "No-Go theorem for the characterization of work fluctuations in coherent quantum systems," Physical Review Letters **118**, 070601 (2017).

[52] M. Lostaglio, "Quantum fluctuation theorems, contextuality, and work quasiprobabilities," Physical Review Letters **120**, 040602 (2018).

[53] A. E. Allahverdyan, "Nonequilibrium quantum fluctuations of work," Physical Review E **90**, 032137 (2014).

[54] A. Eddington, *The nature of the physical world* (Literary Licensing, LLC, 2013).

[55] M. J. Klein, "Maxwell, his demon, and the second law of thermodynamics," American Scientist **58**, 84–97 (1970).

[56] W. Thomson, "The kinetic theory of the dissipation of energy," Proceedings of the Royal Society of Edinburgh **8**, 325–334 (1875).

[57] L. Brillouin, "Maxwell's demon cannot operate: Information and entropy. I," Journal of Applied Physics **22**, 334–337 (1951).

[58] C. H. Bennett, "The thermodynamics of computation—a review," International Journal of Theoretical Physics **21**, 905–940 (1982).

[59] C. H. Bennett, "Logical reversibility of computation," IBM journal of Research and Development **17**, 525–532 (1973).

[60] E. Fredkin and T. Toffoli, "Conservative logic," International Journal of theoretical physics **21**, 219–253 (1982).

[61] R. Landauer, "Irreversibility and heat generation in the computing process," IBM journal of research and development **5**, 183–191 (1961).

[62] D. Mandal and C. Jarzynski, "Work and information processing in a solvable model of Maxwell's demon," Proceedings of the National Academy of Sciences **109**, 11641–11645 (2012).

[63] P. Strasberg, G. Schaller, T. Brandes, and C. Jarzynski, "Second laws for an information driven current through a spin valve," Physical Review E **90**, 062107 (2014).

[64] W. H. Zurek, "Thermodynamic cost of computation, algorithmic complexity and the information metric," Nature **341**, 119–124 (1989).

[65] W. H. Zurek, "Algorithmic information content, Church—Turing thesis, physical entropy, and Maxwell's demon," in *Information Dynamics* (Springer, 1991) pp. 245–259.

[66] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Texts in Computer Science (Springer International Publishing, Cham, 2019).

[67] L. A. Levin, "Universal sequential search problems," Problems of Information Transmission **9**, 265–266 (1973).

[68] D. Reeb and M. M. Wolf, "An improved Landauer principle with finite-size corrections," New Journal of Physics **16**, 103011 (2014).

[69] C. Jarzynski, "Nonequilibrium equality for free energy differences," Physical Review Letters **78**, 2690 (1997).

[70] C. Jarzynski, "Nonequilibrium work relations: Foundations and applications," The European Physical Journal B **64**, 331–340 (2008).

[71] C. Jarzynski, "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale," Annual Review of Condensed Matter Physics **2**, 329–351 (2011).

[72] G. E. Crooks, "Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems," Journal of Statistical Physics **90**, 1481–1487 (1998).

[73] E. H. Feng and G. E. Crooks, "Length of time's arrow," Physical review letters **101**, 090602 (2008).

[74] K. V. Hovhannisyan and A. Imparato, "Energy conservation and fluctuation theorem are incompatible for quantum work," (2021), 10.48550/arXiv.2104.09364.

[75] C. C. Aw, F. Buscemi, and V. Scarani, "Fluctuation theorems with retrodiction rather than reverse processes," AVS Quantum Science **3**, 045601 (2021).

[76] S. Watanabe, "Conditional probability in physics," Progress of Theoretical Physics Supplement **65**, 135–160 (1965).

[77] G. Manzano, J. M. Horowitz, and J. M. Parrondo, "Nonequilibrium potential and fluctuation theorems for quantum maps," Physical Review E **92**, 032129 (2015).

[78] G. Manzano, J. M. Horowitz, and J. M. Parrondo, "Quantum fluctuation theorems for arbitrary environments: adiabatic and nonadiabatic entropy production," Physical Review X **8**, 031037 (2018).

[79] G. E. Crooks, "Quantum operation time reversal," Physical Review A **77**, 034101 (2008).

[80] P. Salamon and R. S. Berry, "Thermodynamic length and dissipated availability," Physical Review Letters **51**, 1127–1130 (1983).

[81] F. Weinhold, "Metric geometry of equilibrium thermodynamics," The Journal of Chemical Physics **63**, 2479 (1975).

[82] B. Andresen, R. S. Berry, R. Gilmore, E. Ihrig, and P. Salamon, "Thermodynamic geometry and the metrics of Weinhold and Gilmore," Physical Review A **37**, 845–848 (1988).

[83] J. Nulton, P. Salamon, B. Andresen, and Q. Anmin, "Quasistatic processes as step equilibrations," The Journal of Chemical Physics **83**, 334 (1985).

[84] G. E. Crooks, "Measuring thermodynamic length," Physical Review Letters **99**, 100602 (2007).

[85] D. A. Sivak and G. E. Crooks, "Thermodynamic metrics and optimal paths," Physical Review Letters **108**, 190602 (2012) (2012).

[86] M. F. Ludovico, F. Battista, F. von Oppen, and L. Arrachea, "Adiabatic response and quantum thermoelectrics for *ac*-driven quantum systems," Physical Review B **93**, 075136 (2016).

[87] M. Lostaglio, K. Korzekwa, and A. Milne, "Markovian evolution of quantum coherence under symmetric dynamics," Physical Review A **96**, 032109 (2017).

[88] M. Lostaglio, "An introductory review of the resource theory approach to thermodynamics," Reports on Progress in Physics **82**, 114001 (2019).

[89] E. H. Lieb and J. Yngvason, "The physics and mathematics of the second law of thermodynamics," Physics Reports **310**, 1–96 (1999).

[90] F. Brandao, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, "The second laws of quantum thermodynamics," Proceedings of the National Academy of Sciences **112**, 3275–3279 (2015).

[91] M. Lostaglio, D. Jennings, and T. Rudolph, "Description of quantum coherence in thermodynamic processes requires constraints beyond free energy," Nature Communications **6**, 6383 (2015).

[92] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, "Quantum majorization and a complete set of entropic conditions for quantum thermodynamics," Nature Communications **9**, 5352 (2018).

[93] A. D. Varizi, M. A. Cipolla, M. Perarnau-Llobet, R. C. Drumond, and G. T. Landi, "Contributions from populations and coherences in non-equilibrium entropy production," New Journal of Physics **23**, 063027 (2021).

[94] I. Marvian Mashhad, *Symmetry, asymmetry and quantum information* (University of Waterloo, 2012).

[95] J. Åberg, "Catalytic coherence," Physical Review Letters **113**, 150402 (2014).

[96] D. J. Evans and D. J. Searles, "The fluctuation theorem," Advances in Physics **51**, 1529–1585 (2002).

[97] P. Salamon, J. Nulton, G. Siragusa, T. Andersen, and A. Limon, "Principles of control thermodynamics," Energy **26**, 307–319 (2001).

[98] P. Abiuso and M. Perarnau-Llobet, "Optimal cycles for low-dissipation heat engines," Physical Review Letters **124**, 110606 (2020).

[99] M. Lostaglio, "Certifying quantum signatures in thermodynamics and metrology via contextuality of quantum linear response," Physical Review Letters **125**, 230603 (2020).

[100] R. Kunjwal, M. Lostaglio, and M. F. Pusey, "Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts," Physical Review A **100**, 042116 (2019).

[101] M. Müller, "Probabilistic theories and reconstructions of quantum theory," SciPost Physics Lecture Notes , 028 (2021).

[102] G. Ruppeiner, "Thermodynamic curvature measures interactions," American Journal of Physics **78**, 1170–1180 (2010).

[103] M. Lostaglio and K. Korzekwa, "Continuous thermomajorization and a complete set of laws for Markovian thermal processes," Physical Review A **106**, 012426 (2022).

[104] J. W. Bush and A. U. Oza, "Hydrodynamic quantum analogs," Reports on Progress in Physics **84**, 017001 (2020).

[105] R. W. Spekkens, "Evidence for the epistemic view of quantum states: A toy theory," Physical Review A **75**, 032110 (2007).

[106] R. W. Spekkens, "Contextuality for preparations, transformations, and unsharp measurements," Physical Review A **71**, 052108 (2005), quant-ph/0406166 .

[107] E. Beltrametti and S. Bugajski, "A classical extension of quantum mechanics," Journal of Physics A: Mathematical and General **28**, 3329 (1995).

[108] S. Kochen and E. P. Specker, "The problem of hidden variables in quantum mechanics," in *The logico-algebraic approach to quantum mechanics* (Springer, 1975) pp. 293–328.

[109] R. W. Spekkens, "Negativity and contextuality are equivalent notions of nonclassicality," Physical Review Letters **101**, 020401 (2008).

[110] M. F. Pusey, "Anomalous weak values are proofs of contextuality," Physical Review Letters **113**, 200401 (2014).

[111] W. Jurkat and H. J. Ryser, "Term ranks and permanents of nonnegative matrices," Journal of Algebra **5**, 342–357 (1967).

[112] A. Klyachko, "Quantum marginal problem and representations of the symmetric group," (2004), arXiv:quant-ph/0409113.

[113] S. Bravyi, "Compatibility between local and multipartite states," Quantum Information and Computation **4**, 012–026 (2004).

[114] M. B. Ruskai, S. Szarek, and E. Werner, "An analysis of completely-positive trace-preserving maps on M_2," Linear Algebra and its Applications **347**, 159–187 (2002).

[115] A. Turing, "On computable numbers, with an application to the Entscheidungs problem," Proceedings of the London Mathematical Society **s2-42**, 230–265.

[116] C. Calude, *Information and randomness: An algorithmic perspective*, 2nd ed., rev. and extended ed., Texts in Theoretical Computer Science (Springer, Berlin ; New York, 2002).

[117] G. J. Chaitin, "Algorithmic information theory," IBM journal of research and development **21**, 350–359 (1977).