

The renormalization group via statistical inference

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 New J. Phys. 17 083005

(<http://iopscience.iop.org/1367-2630/17/8/083005>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 194.95.157.141

This content was downloaded on 02/08/2016 at 12:49

Please note that [terms and conditions apply](#).



PAPER

The renormalization group via statistical inference

OPEN ACCESS

RECEIVED
17 April 2015REVISED
25 June 2015ACCEPTED FOR PUBLICATION
13 July 2015PUBLISHED
5 August 2015

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

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



Cédric Béný and Tobias J Osborne

Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstraße 2, D-30167 Hannover, Germany

E-mail: cedric.beny@gmail.com**Keywords:** quantum information, quantum field theory, distinguishability metrics, Gaussian states**Abstract**

In physics, one attempts to infer the rules governing a system given only the results of imperfect measurements. Hence, microscopic theories may be effectively indistinguishable experimentally. We develop an operationally motivated procedure to identify the corresponding equivalence classes of states, and argue that the renormalization group (RG) arises from the inherent ambiguities associated with the classes: one encounters flow parameters as, e.g., a regulator, a scale, or a measure of precision, which specify representatives in a given equivalence class. This provides a unifying framework and reveals the role played by information in renormalization. We validate this idea by showing that it justifies the use of low-momenta n -point functions as statistically relevant observables around a Gaussian hypothesis. These results enable the calculation of distinguishability in quantum field theory. Our methods also provide a way to extend renormalization techniques to effective models which are not based on the usual quantum-field formalism, and elucidates the relationships between various type of RG.

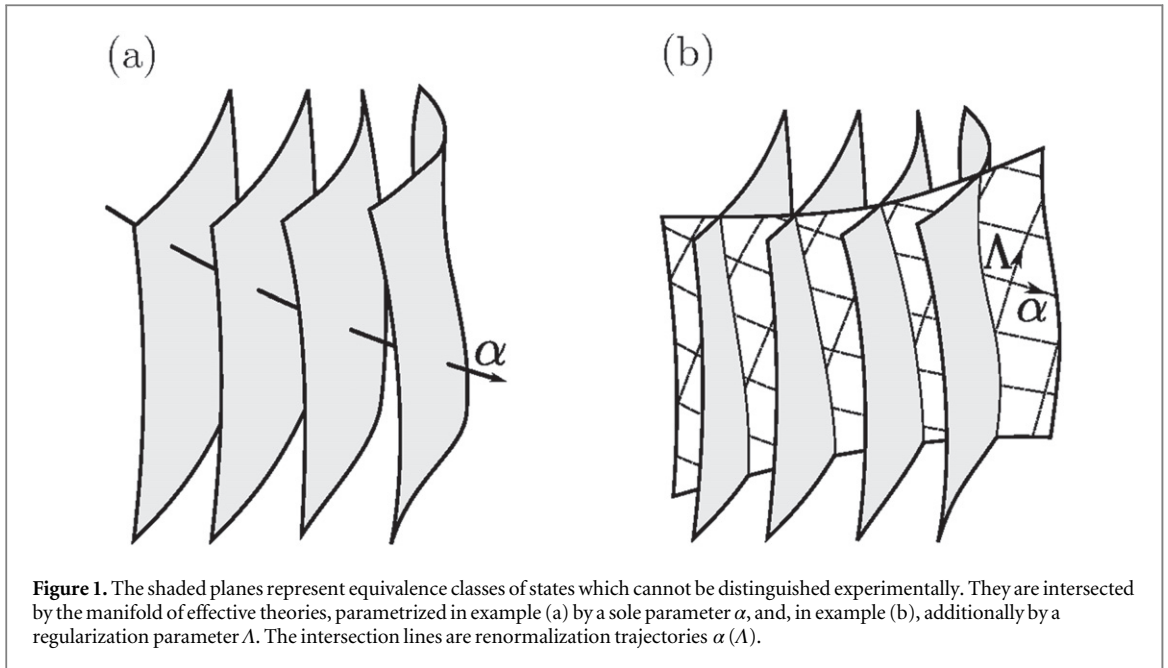
The renormalization group (RG), as conceived by Wilson [1, 2], relies on the idea that it is possible to describe long-distance physics while essentially ignoring short-distance phenomena; Wilson argued that, if we are content with predictions to some specified accuracy, the effects of physics at smaller lengthscales can be absorbed into the values of a few parameters of some *effective theory* for the long-distance degrees of freedom. The RG now underpins much of our understanding of modern theoretical physics and provides the interpretational framework for quantum field theories. It has been applied in a dazzling array of incarnations to study systems from statistical physics [3] to applied mathematics [4].

The general applicability of RG techniques strongly suggests the existence of a deep unifying principle which would make it possible to directly compare different manifestations of the RG and to unlock its full potential. It has been suggested that such a general implementation-independent formulation of the RG is to be found in an information-theoretic approach [5] because the RG works by *ignoring* certain aspects of the system. Although the information-theoretic flavour of the RG is manifest in the case of block-decimation [6–8], it is far less obvious in the context of particle physics from where the terminology of renormalization originates [9]. Previous attempts at tackling this problem (see, e.g., [10–14] for a selection) depend on details of the chosen model or formalism and do not yet offer the truly general unification that one might hope for.

The objective of this paper is to propose an operationally motivated, model-independent, and hence information-theoretic framework for the RG. Our main result is a demonstration that this framework encompasses, as a particular case, the RG implemented with respect to a regulator (as found in QFT).

Our approach is related to that of a recent paper of Machta *et al* [15], who observed that the relevant parameters selected by the RG have the property that they generate perturbations of a statistical state which are distinguishable (in information-theoretic terms) even when the system is coarse-grained. This is why these parameters can be *inferred* experimentally and are useful for predictions.

We first step back, and phrase the inference task as a game played between two players: a passive one, Alice, who simply possesses a quantum or classical system, and Bob, who perceives the system via a known noisy quantum channel \mathcal{E} . (That is, any map linearly taking density matrices to density matrices, even as part of a larger system. Classically, it is any stochastic map.) The channel may for instance represent a coarse-graining.



We think of Alice as possessing the true state of a physical system, while Bob is an experimentalist whose practical limitations are formalised by the channel. When Bob tries to infer the state of Alice's system, he is faced with the ill-posed *inverse problem* of inverting a quantum channel to find the input from the output.

Let us consider first a situation where the channel has a non-trivial kernel. For instance, \mathcal{E} could be the partial trace over all high-momentum modes of a theory. If two states ρ and ρ' are such that $\mathcal{E}(\rho) = \mathcal{E}(\rho')$ then they cannot be distinguished by Bob and hence are both just as good as hypotheses for Alice's state. This indistinguishability results in *equivalence classes* of states: all that Bob can hope to do is to determine in which class the true state is. The classes can be parameterized by a smooth manifold of unique representatives (figure 1(a)). For instance, if \mathcal{E} traces out high-momentum modes, the equivalence classes can be labelled by states whose high momentum modes are in some fiducial product state.

Once the classes of experimentally indistinguishable states are identified, we propose that the various existing types of RG result from an exploration of the freedom available in choosing the representative within a class. For example, when modifying a regularization parameter, as occurs in high-energy physics, or when simplifying the description of the state and isolating the relevant degrees of freedom, as commonly practised in condensed matter theory. Before we describe these two cases in more depth, we need to consider more general, and more realistic experimental limitations. This requires taking *approximate* indistinguishability into account.

1. General framework

A reasonable measure of distinguishability between two states ρ and ρ' to be used in this situation is the relative entropy

$$S(\rho' \parallel \rho) = \text{Tr} \left(\rho \left(\log \rho - \log \rho' \right) \right), \quad (1)$$

which measures the optimal exponential rate of decrease of the probability of mistaking ρ' for ρ as a function of the number of copies available, while still letting the probability of mistaking ρ for ρ' go to zero [16]. This asymmetric scenario is relevant to the situation where one attempts to prove the new hypothesis ρ' against a well established one: ρ . Our framework can also be adapted to different measures, but we use this one here for concreteness. The above interpretation is for an observer able to measure any observable on Alice's system. Bob, however, has a limited access to Alice's state. Since he can only make direct measurements on the states $\mathcal{E}(\rho)$ and $\mathcal{E}(\rho')$, his optimal ability to distinguish between ρ' and ρ according to the above scenario is instead given by the rate $S(\mathcal{E}(\rho') \parallel \mathcal{E}(\rho))$.

The effect of \mathcal{E} can also be thought of as limiting the type of observable that Bob can measure directly on Alice's system, through the *Heisenberg picture* defined via the adjointness relation $\text{tr}(\rho \mathcal{E}^\dagger(A)) = \text{tr}(\mathcal{E}(\rho)A)$: Bob can effectively only measure POVMs on ρ with elements $\mathcal{E}^\dagger(A_i)$, where $0 \leq A_i \leq \mathbf{1}$ and $\sum_i A_i = \mathbf{1}$. His effective distinguishability rate $S(\mathcal{E}(\rho') \parallel \mathcal{E}(\rho))$ is hence smaller than that of an all powerful experimentalist, namely $S(\rho' \parallel \rho)$, because he has access to fewer observables.

Consequently, we could attempt to deem two states ρ and ρ' experimentally equivalent if $S(\mathcal{E}(\rho') \parallel \mathcal{E}(\rho)) < \delta$ for some desired maximal rate δ . However, this does not define an equivalence relation (this relation is not transitive, nor even symmetric). Nevertheless, if δ is sufficiently small, we still expect that the set of states ρ' close to ρ form an approximately linear subspace of matrices, as occurs in the exact case $\delta = 0$. This motivates us to *linearize* the relation around a starting hypothesis ρ .

Let us consider the state $\rho' = \rho + \epsilon X$, where ϵ may be arbitrarily small. We will call the operator X , which must be hermitian and traceless, a *feature*. In terms of the manifold of density matrices, X represents a tangent vector to the point ρ . (It is related to the tangent vector represented as differential operator \hat{X} on scalar functions $f(\rho + \epsilon X) = f(\rho) + \epsilon(\hat{X}f)(\rho) + \mathcal{O}(\epsilon^2)$). Then, to lowest order in ϵ , we have

$$S(\rho + \epsilon X \parallel \rho) \approx \epsilon^2 \text{tr}(X \Omega_\rho^{-1}(X)) \equiv \epsilon^2 \langle X, X \rangle_\rho, \quad (2)$$

where $\Omega_\rho^{-1}(Y) = \frac{d}{dt} \log(\rho + tY)|_{t=0}$ is a non-commutative version of the operation ‘division by ρ ’. The quantity $\langle X, Y \rangle_\rho \equiv \text{tr}(X \Omega_\rho^{-1}(Y))$ is an inner product on operators. Since it is defined at every point of the manifold of states, it is a *metric* in the sense of differential geometry and is one of the many quantum generalizations of the Fisher information metric [17].

In this linear approximation, a state $\rho + X$ is approximately indistinguishable from $\rho + Y$ by Bob if

$$\langle \mathcal{E}(X - Y), \mathcal{E}(X - Y) \rangle_{\mathcal{E}(\rho)} < \delta. \quad (3)$$

The set of states $\rho + X$ satisfying this condition is an ellipsoid within Alice’s state space. If \mathcal{E} is not invertible, the ellipsoid is infinitely wide in the null directions Z with $\mathcal{E}(Z) = 0$. Consequently, in the generic case, we use the following idealized relation: we say that the two states $\rho + X$ and $\rho + Y$ are equivalent if $X - Y$ lies in the span of the ‘largest’ principal directions of the ellipsoid (those that contract ‘the most’).

This idealization removes any trace of the desired precision δ , as we are only talking of the direction of $X - Y$ independently of its magnitude. Instead, Bob must choose the number n of features he deems sufficiently distinguishable. A pertinent way of doing this is to consider the case where the channel \mathcal{E} depends on a parameter σ parameterizing the precision of Bob’s instruments, and to worry about the asymptotic behaviour of the norm $\langle \mathcal{E}(Z), \mathcal{E}(Z) \rangle_{\mathcal{E}(\rho)}$ in the limit of large imprecision σ . The choice of threshold n then amounts to choosing the type of asymptotic behaviour that we deem negligible. This is what happens in the examples presented below.

The principal directions of the ellipsoid are obtained by a singular value decomposition of \mathcal{E} with respect to the scalar product defined by the metric. Let \mathcal{R}_ρ be the *adjoint* of \mathcal{E} defined by $\langle X, \mathcal{R}_\rho(Y) \rangle_\rho = \langle \mathcal{E}(X), Y \rangle_{\mathcal{E}(\rho)}$ for all features X, Y . Explicitly

$$\mathcal{R}_\rho = \Omega_\rho \mathcal{E}^\dagger \Omega_{\mathcal{E}(\rho)}^{-1}. \quad (4)$$

This map generalizes the *transpose channel* [18]. Classically, if $p(y|x)$ are the components of \mathcal{E} , then \mathcal{R}_ρ has for components the conditional probabilities $p(x|y)$ derived from Bayes’ rule with prior ρ . The *principal features* X_j are the solution of

$$\mathcal{R}_\rho \mathcal{E}(X_j) = \eta_j X_j. \quad (5)$$

The eigenvalues η_j are also the singular values of \mathcal{E} , and satisfy $1 \geq \eta_1 \geq \eta_2 \geq \dots \geq 0$. We call η_j the *relevance* of X_j . The linear operator $\mathcal{R}_\rho \mathcal{E}$ is self adjoint in the scalar product $\langle \cdot, \cdot \rangle_\rho$. Therefore, the principal features form an orthogonal basis of the tangent space at ρ .

This concept of relevance is a genuinely coordinate independent version of the *stiffness* defined in [15]. It equals stiffness computed with respect to the special parametrization in which the original metric is given by the identity matrix.

We call a feature X *relevant* if it is in the span of X_1, \dots, X_n , and *irrelevant* if it is orthogonal to those. Our idealized equivalence classes make $\rho + X$ and $\rho + Y$ equivalent from the point of view of Bob if and only if $X - Y$ is irrelevant, or, equivalently, if $\langle X - Y, Z \rangle_\rho = 0$ for all relevant feature Z .

In order to obtain a physically more intuitive condition, let us define the *principal observables* to be the operators $A_j = \Omega_\rho^{-1}(X_j)$, solutions of the dual Heisenberg-picture eigenvalue equation

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(A_j) = \eta_j A_j. \quad (6)$$

Analogously, we say that A is a relevant observable if it is in the span of A_1, \dots, A_n . With this definition, our equivalence condition amounts to considering two effective states ρ' and ρ'' to be equivalent (in the neighbourhood of ρ) when they yield the same expectation values for *all relevant observables*:

$$\rho' \sim_\rho \rho'' \quad \text{iff} \quad \text{tr}(\rho' A_j) = \text{tr}(\rho'' A_j) \quad \text{for } j \leq n. \quad (7)$$

For instance, consider the strictest possible relevance threshold where only features with exactly zero relevance are deemed to be irrelevant. These are the operators X in the kernel of \mathcal{E} . In this case we recover the

exact state-independent equivalence relation which identifies $\rho' \sim \rho''$ if $\mathcal{E}(\rho') = \mathcal{E}(\rho'')$. The corresponding relevant observables are the self-adjoint operators A satisfying $\text{tr}(AX) = 0$ for all X in the kernel of \mathcal{E} , which are precisely those of the form $\mathcal{E}^\dagger(B)$ for some B . In addition, these are all the observables that Bob can ever hope to measure expectation values of, since for all B , $\text{tr}(B\mathcal{E}(\rho)) = \text{tr}(\mathcal{E}^\dagger(B)\rho)$.

2. One classical mode

For a simple but nontrivial example suppose that Alice has a stochastic classical system consisting of a single real variable, e.g., the position x of a particle. The true state to be discovered by Bob is a probability distribution $\rho(x)$. Bob's experimental limitation consists of a finite precision σ at which he can resolve x . This can be modelled by a stochastic map \mathcal{E} whose effect is a convolution of Alice's probability distribution with a Gaussian of width σ :

$$\mathcal{E}(\rho)(x) = \int \rho(y) N_\sigma(x-y) dy \quad (8)$$

here N_σ is the normal distribution with variance σ . Suppose, further, that Bob's initial hypothesis is a simple Gaussian distribution, which we think of as a thermal state $\rho(x) \propto e^{-H(x)}$ for the 'hamiltonian' $H(x) = \frac{x^2}{2\tau^2}$. The action of the operator \mathcal{R}_ρ^\dagger can be written explicitly:

$$\mathcal{R}_\rho^\dagger(f)(x) = \frac{\int dy f(x) \rho(x) N_\sigma(x-y) dy}{\int dy \rho(x) N_\sigma(x-y) dy}. \quad (9)$$

Noting also that $\mathcal{E} = \mathcal{E}^\dagger$, one can check by explicit calculation of the Gaussian integrals that if $G_t(x) = e^{tx/\tau - t^2/2}$, then

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(G_t) = G_{\eta t} \quad (10)$$

with $\eta = \frac{\tau^2}{\sigma^2 + \tau^2}$. Hence, the eigenvalue problem defined in equation (6) is solved by differentiating equation (10) n times with respect to t , evaluated at $t = 0$. Observe that $G_t(x)$ is the generating function for the Hermite polynomials, hence the principal observables are the hermite polynomials $\text{He}_n(x/\tau)$, with respective eigenvalues $\eta_n = \eta^n$, or $\eta_n \approx (\tau/\sigma)^{2n}$ for $\sigma \gg \tau$. Following our criterion this means that, since the first n hermite polynomials span all polynomials of degree n , that for a threshold n , two nearby states are equivalent exactly when they share the same first n moments.

For instance, up to distinguishability of order σ^{-4} ($n = 2$), the effective hamiltonian $H_0(x) = \frac{x^2}{2\tau_0^2} + \lambda x^4$ is equivalent to $H_1(x) = \frac{x^2}{2\tau_1^2}$ provided that τ_1 is 'renormalized' so as to yield the same second moment as H_0 . This simplification from H_0 to H_1 morally corresponds to a step of the type of RG employed in condensed matter theory, where a hamiltonian is simplified in a way that only affects some 'unobservable' small scale features of the systems.

The situation in particle physics is *a priori* quite different. Quantum field theories typically come with an unwanted parameter, a *regulator* Λ , which has no true physical significance, although it often mimics a lattice spacing. Its presence, however, is not a problem if the observable predictions of the theory do not depend on it. This is possible if we assume some reasonable limitation on Bob's measurement abilities, so that any change in Λ can be compensated by a change in the state's other parameters so as to stay within a given equivalence class (figure 1(b)). This dependance of the state's parameters on Λ is the type of RG flow which naturally occurs in QFT.

Using the above toy example, a similar problem could occur for the hamiltonian H_0 if λ were to be experimentally determined to be negative (using a first order approximation in λ for the state). Indeed, the resulting distribution $\rho(x)$ would diverge if calculated non-perturbatively. This can be fixed mathematically by adding a sixth order term x^6/Λ to the effective hamiltonian, which, to distinguishability of order σ^{-4} , can be made to be equivalent to H_0 by adjusting the parameters τ and λ as function of Λ so as to preserve up to the fourth moment.

Those two concepts of renormalization can be made to match in QFT because divergences can be identified as contributions from an infinite number of *irrelevant* features. Hence, the simplification which consists in subtracting them from the state also regularizes the theory.

3. Classical Gaussian states

We solve equation (6) for Gaussian states over arbitrarily many modes, and for a channel \mathcal{E} which is any Gaussian stochastic map. We consider n real random variables ϕ_i . We write $\phi(f) := \sum_i f_i \phi_i$ for any vector f , which corresponds to a 'smeared' field in the continuum limit. A general Gaussian stochastic map \mathcal{E} is defined

by the effect of its transpose to the moment-generating functions:

$$\mathcal{E}^\dagger \left(e^{\phi(f)} \right) = e^{\phi(Xf) + \frac{1}{2}(f, Yf)}, \quad (11)$$

where $(f, g) = \sum_i f_i g_i$, Y and X are real matrices and Y is positive (we give a concrete example in the quantum case). Similarly, an arbitrary (but centred) Gaussian state ρ is defined by

$$\langle e^{\phi(f)} \rangle_\rho = e^{\frac{1}{2}(f, Af)}, \quad (12)$$

where A is real and symmetric. Using the definition of \mathcal{R}_ρ^\dagger as adjoint of \mathcal{E}^\dagger in the dual metric, applied to the generating functions $e^{\phi(f)}$, one can show that the random variables

$$G(f) := e^{-\frac{1}{2}(f, Af) + \phi(f)} \quad (13)$$

satisfy

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(G(f)) = G(Hf) \quad (14)$$

with

$$H = \left(1 + A^{-1} X^{-1} Y X^{-1} \right)^{-1}. \quad (15)$$

(See appendix A.) Note that H is symmetric with respect to the scalar product $(\cdot, A \cdot)$, hence it has a complete orthonormal family of eigenfunctions f_k with eigenvalues η_k . We obtain the eigen-variables of $\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger$ (namely the principal observables) explicitly by differentiating the generating functional G in the directions of the functions f_k any number of times, and evaluating the result at $f=0$.

4. Interactions

The previous result can be used to perturbatively calculate the principal observables around nonGaussian states. In order to do this, we need to work within a representation of the real Hilbert space formed by the principal observables of the Gaussian state, together with the scalar product defined by the metric evaluated at the Gaussian state. This is always a symmetric Fock space, where the vacuum $|0\rangle$ corresponds to the constant random variable $G(0) = 1$ (with relevance 1), and the creation operator a_k^\dagger associated with vector f_k , acting on a principal observable, leads to a new principal observable with relevance multiplied by η_k . The perturbed eigenvalue problem can then be expressed to any order using standard Feynman diagrams. An example is detailed in the appendix B. We show below, however, that the standard RG conditions in QFT can be recovered from the Gaussian results alone.

5. Quantum Gaussian states

A quantum Gaussian state is defined by quantization of a classical phase space. Let f, g denote classical observables which are linear functions in the canonical variables, with some scalar product (f, g) and the symplectic form Δ . Let $\Phi(f)$ denote the quantization of f , such that

$$[\Phi(f), \Phi(g)] = i(f, \Delta g) \mathbf{1}. \quad (16)$$

Any quantum state is uniquely specified by its characteristic function $f \mapsto \langle e^{i\Phi(f)} \rangle_\rho$. For a quantum Gaussian state ρ , this is $e^{-\frac{1}{2}(f, Af)}$, where A is a real symmetric matrix satisfying $A + \frac{i}{2}\Delta \geq 0$. A general Gaussian channel is characterized by its effect on the Weyl operators:

$$\mathcal{E}^\dagger \left(e^{i\Phi(f)} \right) = e^{i\Phi(Xf)} e^{-\frac{1}{2}(f, Yf)}, \quad (17)$$

where X and Y are real matrices such that

$$Y - \frac{i}{2} X^\dagger \Delta X + \frac{i}{2} \Delta \geq 0. \quad (18)$$

One can then verify that the principal observables are polynomials generated by

$$G_A(f) = e^{\frac{1}{2}(f, Af) + i\Phi(f)}. \quad (19)$$

This is done by first noting their orthogonality, and applying the definition of \mathcal{R}_ρ with respect to the generating function as in the classical case (see appendix C).

As an example, we consider a gibbs state of a Klein–Gordon field of mass m at inverse temperature β , with canonical conjugate coordinates $\phi(x)$ and $\pi(x)$. We will need the real fourier components

$$\phi_k = \int dx \left[\cos(kx) \phi(x) - \omega_k^{-1} \sin(kx) \pi(x) \right], \quad (20)$$

$$\pi_k = \int dx \left[\omega_k \sin(kx) \phi(x) + \cos(kx) \pi(x) \right], \quad (21)$$

which are decoupled under the classical dynamics. Because the phase space is infinite-dimensional, the concept of Gaussian state introduced above has to be generalized with some care. Alternatively, one may choose boundary conditions and a momentum cutoff so as to render it finite-dimensional. For our purpose, we define the Gaussian state through the bilinear form that it defines on the space of linear classical observables. In terms of the observables

$$f(\phi, \pi) = \int dk \left(f_k \phi_k + f'_k \pi_k \right), \quad (22)$$

$$g(\phi, \pi) = \int dk \left(g_k \phi_k + g'_k \pi_k \right), \quad (23)$$

the quadratic form is

$$(f, Ag) = \frac{1}{2} \int dk \coth(\beta\omega_k/2) \left(\omega_k f_k g_k + \frac{1}{\omega_k} f'_k g'_k \right). \quad (24)$$

We also consider a Gaussian channel \mathcal{E} . In the continuum, the matrices X and Y become linear functions. We use

$$(Xf)(\phi, \pi) = f(N_\sigma \star \phi, N_\sigma \star \pi), \quad (25)$$

where $N_\sigma \star \cdot$ denotes convolution by a Gaussian of variance σ , and

$$(Yf)(\phi, \pi) = f(y_\phi \phi, y_\pi \pi). \quad (26)$$

The parameter σ characterizes spatial resolution, and y_ϕ and y_π field value resolutions. The condition expressed in equation (18) reduces in this case to the uncertainty relation $y_\phi y_\pi \geq 1$.

For $y_\phi y_\pi \gg 1$, we find that the quantized field observables $\hat{\phi}_k = \Phi(\phi_k)$ and $\hat{\pi}_k = \Phi(\pi_k)$ are principal observables, with respective relevance

$$\eta_k^\phi \simeq \frac{1}{\frac{\beta\omega_k}{2} \coth \frac{\beta\omega_k}{2} + \beta\omega_k^2 y_\phi^2 e^{k^2\sigma^2}} \quad (27)$$

and

$$\eta_k^\pi \simeq \frac{1}{\frac{\beta\omega_k}{2} \coth \frac{\beta\omega_k}{2} + \beta y_\pi^2 e^{k^2\sigma^2}}. \quad (28)$$

Since the channel acts independently on each mode, the products of n copies of such operators with distinct momenta are also eigenvectors with relevance equal to the product of the corresponding values η_k^ϕ or η_k^π . For instance, the n -point functions $\hat{\phi}_{k_1} \cdots \hat{\phi}_{k_n}$ have relevance $\eta_{k_1}^\phi \cdots \eta_{k_n}^\phi$, provided that the momenta k_1, \dots, k_n are all distinct.

6. Renormalization

If we want to recover a RG, we have to pick a threshold on the asymptotic decay of relevance in terms of the three noise parameters σ , y_π and y_ϕ . Since the relevance decays exponentially in the total momentum, the product of field operators can always be considered irrelevant as soon as they involve operators with mode $k \gg 1/\sigma$. At temperatures large compared to m , the relevance of the n -point functions also decays to order $2n$ in y_π and y_ϕ . Hence, in this approximation two states are effectively equivalent if they have the same n -point functions at momenta smaller than $1/\sigma$. Without restriction on n , this is precisely the conditions used in QFT for the RG as function of a regulator Λ (needed as momentum cutoff on divergent integrals resulting from perturbation theory around Gaussians). Instead of viewing the cutoff as an explicit parameter of the state, a change of cutoff from Λ to Λ' can also be absorbed into a *rescaling* of space by a factor $s = \Lambda'/\Lambda$. The condition that the state stays in the same equivalence class independently of s yields the Callan–Symanzik equations.

From the condensed matter point of view, a cutoff Λ is fixed and given by the lattice spacing. The description of the state can be simplified by exploiting the freedom we have in choosing a representative of an equivalence class. We may choose the one closest to ρ in relative entropy: this optimization is well known and yields the Gibbs states with only relevant observables in the hamiltonian perturbations, namely terms with field modes $|k| < 1/\sigma$. The requirement that visible predictions be invariant from σ leads to an RG. This matches the

previous RG in the sense that the procedure is technically equivalent to lowering the cutoff Λ to $1/\sigma$ in perturbative expansions.

But our calculation also tells us that we may neglect features whose distinguishability scales poorly with the field-value precisions y_ϕ and y_π , hence justifying the use of effective hamiltonians with low degree polynomials in the fields. For instance, choosing $n=2$ selects only quadratic terms as relevant observables, and the Gaussians states as a natural family of effective states (this is a very different argument than the one based on renormalizability). We note that here y_ϕ, y_π play very different roles than σ because of the differences in asymptotic relevance behaviour, but we could imagine a different type of experimental limitations where more parameters govern the RG besides σ .

7. Distinguishability in QFT

We can use the solution of equation (6) to compute the effective *distinguishability* $D(A)$ of any perturbation generated by a Hamiltonian term A , defined as the lowest-order approximation of the relative entropy $S(\mathcal{E}(\rho_\epsilon) \parallel \mathcal{E}(\rho_0)) = \epsilon^2 D(A) + \mathcal{O}(\epsilon^3)$, where $\rho_\epsilon \propto e^{-H+\epsilon A}$ are normalized states. Indeed, we have $D(A) = \langle A, \mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(A) \rangle_\rho$, which can be computed by expressing A in the basis of principal observables around ρ . For instance, in the scalar field example, $D(\hat{\phi}_k) = \eta_k^\phi \langle \hat{\phi}_k, \hat{\phi}_k \rangle_\rho$, where η_k^ϕ is given above, and $\langle \hat{\phi}_k, \hat{\phi}_k \rangle_\rho = 1/\beta\omega_k^2$. With the standard tools of perturbative QFT, this can be generalized to higher-order expansions of the exponential (while keeping with the lowest order approximation of the relative entropy).

For non-local terms, D has to be made into a density. It may be argued that the unit of volume used to define the density ought to explicitly *scale* with σ , leading to the distinguishability density $d_\sigma(A) = \sigma^d \lim_{\Sigma} D_\sigma(A_\Sigma)/|\Sigma|$, where d is the dimension of space and A_Σ a restriction of A to a region Σ of volume $|\Sigma|$. A Hamiltonian term may then be said to be *relevant* in information-theoretic terms if d_σ scales as a positive power of σ (for a fixed state). Preliminary calculations indicate that the result is compatible with the Wilsonian analysis classically, but may differ in important ways in the genuinely quantum analysis. This will be analysed in further work.

8. Concluding remarks

We introduced a framework which allows for the definition of effective theories in very general terms, taking into account any measure of distinguishability and any model of experimental limitations. We demonstrated the pertinence of this approach by showing that it naturally contains, as a particular case, the concept of effective theory as defined by the RG of quantum field theory. Further work will explore how varying the assumptions lead to effective theories which differ from the standard QFT framework. For instance, taking the field-value resolutions into account in the interacting context leads to a concept of dressed effective field which depends in principle on the detail of the coarse-graining channel and distinguishability metric.

Most interestingly, the fact that this approach is not at all tied to the standard QFT formalism means that it can in principle be applied to completely different types of theories, as well as very different models of experimental limitation (not necessarily related to scale). For instance, the case of loop quantum gravity [19], which proposes a class of background-free quantum field theories, could provide interesting applications.

This approach can also be naturally applied to spin lattice systems, so as to derive effective field theories describing their large scale properties. In standard approaches, a spins system is connected to an effective QFT through symmetry arguments (observables in the discrete and continuous descriptions are paired by identifying the group transformations they generate). Our approach provides a more bottom-up approach, where the effective QFT can in principle be derived through the mechanism by which it emerges, i.e., through the identification of the degrees of freedom which are effectively ignored. This can be performed numerically by solving equation (6) using techniques such as matrix product states.

Finally, the framing of effective QFT in this fundamentally information-theoretical approach elucidates precisely what information is being destroyed when the theory is renormalized. A concrete way of quantifying this is proposed in the last section. This may provide a first step towards generalizing Zamolodchikov's ctheorem [9, 20], which could in turn provide new techniques for the general classification of effective QFTs.

Acknowledgments

Helpful discussions with numerous people are most gratefully acknowledged: a partial list includes John DeBrot, Andrew Doherty, Jens Eisert, Steve Flammia, Jutho Haegeman, Gerard Milburn, Terry Rudolph, Tom Stace, Frank Verstraete, and Reinhard Werner. This work was supported by the ERC grants QFTCMPS and

SIQS and by the cluster of excellence EXC 201 Quantum Engineering and Space-Time Research. We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Leibniz Universität Hannover.

Appendix A. Classical Gaussian states

We want to find the action of \mathcal{R}_ρ^\dagger on $G(f)$. By directly using its definition as adjoint of \mathcal{E}^\dagger , we find that for any linear classical observables f and g

$$\begin{aligned} \langle \mathcal{R}_\rho^\dagger(e^{\phi(f)}), e^{\phi(g)} \rangle_{\mathcal{E}(\rho')} &= \langle e^{\phi(f)}, \mathcal{E}^\dagger(e^{\phi(g)}) \rangle_{\rho'} \\ &= e^{\frac{1}{2}(g, Yg)} \langle e^{\phi(f+Xg)} \rangle_{\rho'} \\ &= e^{\frac{1}{2}(g, Yg)} + \frac{1}{2}(f+Xg, A(f+Xg)) \\ &= e^{\frac{1}{2}(f, Af) + \frac{1}{2}(g, (XAX+Y)g) + (g, XAf)}. \end{aligned}$$

But note that, using

$$j := (XAX + Y)^{-1}XAf, \quad (29)$$

we have

$$\begin{aligned} \langle e^{\phi(j)}, e^{\phi(g)} \rangle_{\mathcal{E}(\rho)} &= e^{\frac{1}{2}(j, Yj) + \frac{1}{2}(g, (XAX+Y)g) + (g, XAf)} \\ &\quad \times e^{\frac{1}{2}(j, XAXj)}. \end{aligned}$$

By comparing the two expressions, we obtain

$$\begin{aligned} \langle \mathcal{R}_\rho^\dagger(e^{\phi(f)}), e^{\phi(g)} \rangle_{\mathcal{E}(\rho)} &= e^{\frac{1}{2}(f, Af) - \frac{1}{2}(j, Yj) - \frac{1}{2}(j, XAXj)} \\ &\quad \times \langle e^{\phi(j)}, e^{\phi(g)} \rangle_{\mathcal{E}(\rho)}. \end{aligned}$$

Since this is true for all g , then

$$\mathcal{R}_\rho^\dagger(e^{\phi(f)}) = e^{\frac{1}{2}(f, Af) - \frac{1}{2}(j, (XAX+Y)j)} e^{\phi(j)}$$

or

$$e^{-\frac{1}{2}(f, Af)} \mathcal{R}_\rho^\dagger(e^{\phi(f)}) = e^{-\frac{1}{2}(j, (XAX+Y)j)} e^{\phi(j)} \quad (30)$$

and hence

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(e^{-\frac{1}{2}(f, Af) + \phi(f)}) = e^{-\frac{1}{2}(Xj, Xj)} e^{\phi(Xj)}. \quad (31)$$

Note that

$$Xj = (1 + A^{-1}X^{-1}YX^{-1})^{-1}f. \quad (32)$$

Hence, defining $H = (1 + A^{-1}X^{-1}YX^{-1})^{-1}$, we conclude that

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger(G(f)) = G(Hf). \quad (33)$$

Observe that $AH = H^T A$. Hence H is symmetric in the scalar product $(\cdot, A \cdot)$. Let f_k be an orthonormal basis of eigenfunction of H :

$$H(f_k) = \eta_k f_k \quad \text{and} \quad (f_k, A f_l) = \delta_{kl}. \quad (34)$$

(Note that it may be convenient to consider a complex eigenbasis. Hence we complexify this real Hilbert space in the obvious way.)

We obtain the eigenfunctions of $\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger$ by differentiating $G(f)$ with respect to the basis functions f_k . Indeed, let δ_k denote the functional derivative in the direction of f_k , i.e., for any functional $Z(f)$,

$$\delta_k Z(f) := \frac{\partial}{\partial t} Z(f + t f_k) \Big|_{t=0}, \quad (35)$$

then we obtain

$$\mathcal{E}^\dagger \mathcal{R}_\rho^\dagger \left((\delta_{k_1} \cdots \delta_{k_n} G)(0) \right) = \eta_{k_1} \cdots \eta_{k_n} (\delta_{k_1} \cdots \delta_{k_n} G)(0). \quad (36)$$

We have

$$\begin{aligned} & \left\langle \left(\delta_{k_1} \cdots \delta_{k_n} G \right) (0), \left(\delta_{k_1} \cdots \delta_{k_m} G \right) (0) \right\rangle_{\rho} \\ &= \delta_{k_1} \cdots \delta_{k_n} \delta'_{k_1} \cdots \delta'_{k_m} e^{(f, Af')} |_{f=f'=0}. \end{aligned}$$

where the primed derivatives are relative to f .

The functions $(\delta_{k_1} \cdots \delta_{k_n} G)(0)$ form an orthogonal basis of a representation of the symmetric Fock space \mathcal{F} built from the test functions, with scalar product $\langle \cdot, \cdot \rangle_{\rho}$. One can think of $G(0) = 1 \equiv |0\rangle$ as the vacuum. The other eigenfunctions are obtain by acting on it with the creation operators a_k^{\dagger} for the ‘mode’ f_k . The commutation relations are

$$\left[a_k, a_l^{\dagger} \right] = \delta_{kl} \mathbf{1}. \quad (37)$$

Explicitly

$$a_{k_n}^{\dagger} \cdots a_{k_1}^{\dagger} |0\rangle = \delta_{k_1} \cdots \delta_{k_n} G(f) |_{f=0}. \quad (38)$$

Appendix B. Perturbation theory

This pictures allows one to find the principal observables around non-Gaussian states by using perturbation theory. The trick is to express the information metric with respect to the perturbed state through its kernel K expressed in that Fock space. This allows one to write also the map $\mathcal{R}_{\rho'}^{\dagger}$, for the perturbed state ρ' , also perturbatively as an operator in Fock space. The eigenvalue problem can then be formulated and computed to any degree using standard Feynman diagram techniques.

Let us defined the generating functions

$$K(f, g) = \langle G(f), G(g) \rangle_{\rho'}. \quad (39)$$

Differentiating this functional yields the components in the Fock basis of an operator K that is the kernel of the metric defined by ρ' with respect to the unperturbed metric ρ :

$$\begin{aligned} & \langle 0 | a_{k_1} \cdots a_{k_n} K a_{l_m}^{\dagger} \cdots a_{l_1}^{\dagger} | 0 \rangle \\ &= \delta_{k_1} \cdots \delta_{k_n} \delta'_{l_1} \cdots \delta'_{l_m} K(f, f') |_{f=f'=0}, \end{aligned}$$

where δ'_k denotes derivation with respect to f in the direction f_k . Similarly,

$$I(f, g) = \langle G(f), G(g) \rangle_{\rho} = e^{(f, Ag)} \quad (40)$$

is the generating function of the identity operator.

It will be convenient also to consider the Fock space defined from the metric at point $\mathcal{E}(\rho)$, with the same vacuum; but with creation operators b_k^{\dagger} defined by

$$b_{k_n}^{\dagger} \cdots b_{k_1}^{\dagger} |0\rangle \equiv \delta_{k_1} \cdots \delta_{k_n} Q(f) |_{f=0}, \quad (41)$$

where

$$Q(f) = \mathcal{R}_{\rho}^{\dagger} \left(G \left(H^{-\frac{1}{2}} f \right) \right). \quad (42)$$

These are indeed orthogonal since

$$\langle Q(f), Q(g) \rangle_{\mathcal{E}(\rho)} = e^{\left(H^{-\frac{1}{2}} f, A H^{\frac{1}{2}} g \right)} = e^{(f, Ag)}, \quad (43)$$

where we used the fact that H is positive in terms of the scalar product $(\cdot, A \cdot)$.

In this basis we express

$$L(f, g) = \langle Q(f), Q(g) \rangle_{\mathcal{E}(\rho')}. \quad (44)$$

We can compute from our previous results that

$$\langle \mathcal{R}_{\rho}^{\dagger}(G(f)), \mathcal{R}_{\rho}^{\dagger}(G(g)) \rangle_{\mathcal{E}(\rho')} = \langle G(Hf), G(Hg) \rangle_{\rho'} e^{h^2 (Hf, X^{-2} Hg)}. \quad (45)$$

Hence we find that K and J are related by

$$L(f, g) = K \left(H^{\frac{1}{2}} f, H^{\frac{1}{2}} g \right) e^{h^2 \left(H^{\frac{1}{2}} f, X^{-2} H^{\frac{1}{2}} g \right)}. \quad (46)$$

Also, the channel \mathcal{E}^\dagger naturally maps between the two Fock spaces, represented as the operator E generated by

$$E(f, g) = \langle G(f), \mathcal{E}^\dagger(Q(g)) \rangle_\rho = \langle G(f), G(Hg) \rangle_\rho = e^{(f, AH\frac{1}{2}g)} = e^{(H\frac{1}{2}f, Ag)}.$$

Finally, the unknown is the operator R representing $\mathcal{R}_{\rho'}^\dagger$ as

$$R(f, g) = \langle Q(f), \mathcal{R}_{\rho'}^\dagger(G(g)) \rangle_{\mathcal{E}(\rho)}. \tag{47}$$

It is defined by the relation

$$\langle \mathcal{R}_{\rho'}^\dagger(G(f)), Q(g) \rangle_{\mathcal{E}(\rho')} = \langle G(f), \mathcal{E}^\dagger(Q(g)) \rangle_{\rho'} \tag{48}$$

for all f and g . Expanding this relation in the respective Fock basis, we obtain

$$R^T L = KE \tag{49}$$

or

$$R = L^{-1}EK. \tag{50}$$

If we expand

$$\rho' = \rho \left(\mathbf{1} + \lambda X_1 + \mathcal{O}(\lambda^2) \right), \tag{51}$$

we have

$$K = \mathbf{1} + \lambda K_1 + \mathcal{O}(\lambda^2), \tag{52}$$

and

$$L = \mathbf{1} + \lambda L_1 + \mathcal{O}(\lambda^2). \tag{53}$$

Expanding ER , we obtain

$$ER = E \left(I + \lambda L_1 + \dots \right)^{-1} E \left(I + \lambda K_1 + \dots \right) = E^2 + \lambda \left(E^2 K_1 - E L_1 E \right) + \dots$$

In order to compute the first order corrections to the unperturbed eigenvalue problem, we need the generating function of the perturbation

$$V_1 = E^2 K_1 - E L_1 E. \tag{54}$$

Since E is diagonal in the Fock basis, we only need to worry about K_1 and L_1 directly.

The generating function $K_1(f, f')$ of the operator K_1 is

$$K_1(f, f') = \left\langle X_1 G(f) G(f') \right\rangle_\rho. \tag{55}$$

We will consider an interaction of the form

$$H_I = \frac{1}{4!} \sum_x \phi(f_x)^4 = \frac{1}{4!} \sum_x \frac{\delta^4}{\delta f_x^4} e^{\phi(f)} |_{f=0}, \tag{56}$$

where the functions f_x possibly form a different basis than f_k . This generates the state $\rho' = \rho(\mathbf{1} + \lambda X_1 + \dots)$ where

$$X_1 = H_I - \langle H_I \rangle_\rho \mathbf{1}. \tag{57}$$

We have

$$K_1(f, f') = \frac{1}{4!} \sum_x \frac{\delta^4}{\delta f_x^4} Z(f, f', f'') |_{f''=0} - \langle H_I \rangle_\rho I(f, f'), \tag{58}$$

where

$$\begin{aligned} Z(f, f', f'') &= \left\langle e^{\phi(f'')} G(f) G(f') \right\rangle_\rho \\ &= \left\langle e^{\phi(f''+f'+f)} \right\rangle_\rho e^{-\frac{1}{2}(f, Af) - \frac{1}{2}(f', Af')} \\ &= e^{\frac{1}{2}(f''+f'+f, A(f''+f'+f)) - \frac{1}{2}(f, Af) - \frac{1}{2}(f', Af')}. \end{aligned}$$

When differentiated, this free partition function yields Feynman diagrams with no propagation between the vertices associated with f or f' respectively.

As an example, we performed this calculation for the state corresponding to the euclidean form of the Klein–Gordon scalar field theory with ϕ^4 interaction (in an arbitrary number of spatial dimensions). The channel defined an operator X which performs a Gaussian convolution over scale σ as in the quantum example in the article. We also use $Y = y^2 1$.

The free theory yields the Gaussian state defined by the operator A , inverse of $(A^{-1}f)(x) = \beta [\sum_i \partial_i^2 + m^2]f(x)$. Given that it commutes with the operator X defined by

$$(Xf)(x) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \int e^{-\frac{(x-y)^2}{2\sigma^2}} f(y) dy, \quad (59)$$

with d the number of dimensions, H is codiagonal with X and H which are all self-adjoint in the $L^2(\mathbb{R})$ scalar product.

Using the eigenfunctions of A :

$$f_k(x) = \omega_k e^{ikx} \quad (60)$$

with

$$\omega_k^2 = k^2 + m^2, \quad (61)$$

we obtain the eigenvalues of H :

$$\eta_k = \frac{1}{1 + y^2 \omega_k^2 e^{\sigma^2 k^2}}. \quad (62)$$

The normalized unperturbed ‘one-particle’ principal observables are $a_k^\dagger |0\rangle \equiv \phi(f_k)$. Note that we used complex eigenfunctions because it makes the calculations much simpler. The degeneracy between the k and $-k$ eigenfunctions allows one to recover the real eigenfunctions by linear composition of the complex ones.

The interaction is defined as above using the improper basis

$$f_x(y) = \delta(x - y). \quad (63)$$

Hence, besides $A_{kl} = \delta(k - l)$, we find

$$A_{xy} := (f_x, Af_y) = \int dk \frac{e^{ik(y-x)}}{\omega_k^2} \quad (64)$$

and

$$A_{kx} := (f_k, Af_y) = \frac{f_k(x)}{\omega_k^2}. \quad (65)$$

The principal observables around ρ' , obtained by perturbation from the one-particle observables for the Gaussian state ρ , are $|\psi\rangle_k = a_k^\dagger |0\rangle + \lambda |\psi_k^1\rangle + \mathcal{O}(\lambda^2)$ where the only non-zero components of the first order correction $|\psi_k^1\rangle$ are

$$\begin{aligned} \langle 0 | a_{l_1} a_{l_2} a_{l_3} | \psi_k^1 \rangle &= \frac{\delta(k - l_1 - l_2 - l_3) (1 + \eta_k) \eta_{l_1} \eta_{l_2} \eta_{l_3}}{\omega_k \omega_{l_1} \omega_{l_2} \omega_{l_3} (\eta_k - \eta_{l_1} \eta_{l_2} \eta_{l_3})} \\ &+ \frac{\delta(k - l_1) \delta(l_2 + l_3)}{\omega_{l_2} \omega_{l_3}} \frac{\eta_{l_2} \eta_{l_3}}{(1 - \eta_{l_2} \eta_{l_3})} \int dk' \frac{1}{\omega_{k'}^2} \\ &+ \dots, \end{aligned}$$

where we omitted the next two terms which are obtained by rotating l_1, l_2 and l_3 .

Appendix C. Quantum Gaussian states

We use the notation introduced in the paper. In order to compute the metric explicitly, we need the commutation relation

$$e^{-i\Phi(\bar{f})} e^{i\Phi(g)} = e^{\frac{i}{2}(f, \Delta g)} e^{i\Phi(g - \bar{f})}, \quad (66)$$

where we are now working in a complexified phase-space so as to accommodate imaginary time evolutions. Indeed, we need the group of complex matrices $s \mapsto R_s^A$ associated with the Gaussian state ρ such that

$$\rho^s \Phi(f) \rho^{-s} = \Phi(R_s^A f). \quad (67)$$

The metric (in the Heisenberg picture) is $\langle A, B \rangle_\rho = \int_0^1 \langle A \rho^s B \rho^{-s} \rangle_\rho$. This group is symplectic: $R_s^T \Delta R_s = \Delta$ and leaves the state ρ invariant: $R_s^T A R_s = A$.

Using these properties, we find that the polynomials generated by $G_A(f) = e^{\frac{1}{2}(f, Af) + i\Phi(f)}$ are orthogonal with respect to the metric when they are of different degrees in the canonical observables. This follows from the fact that $\langle G_A(f), G_A(g) \rangle_\rho = \int_0^1 e^{(f, K_s g)} ds$ where $K_s = R_s^\dagger (A + \frac{i}{2}\Delta) R_s$. Indeed, the derivatives of the integrand evaluates to zero at $f = g = 0$ whenever the number of differentiations with respect to f is not equal to the number of differentiations with respect to g .

Moreover, since the channel maps ρ to a Gaussian state defined by the new matrix $B = X^T A X + Y$, we find that $\mathcal{E}^\dagger(G_B(f)) = G_A(Xf)$. Finally, using the definition of \mathcal{R}_ρ^\dagger as adjoint of \mathcal{E}^\dagger , we obtain

$$\langle \mathcal{R}_\rho^\dagger(G_A(f)), G_B(g) \rangle_{\mathcal{E}(\rho)} = \langle G_A(f), G_A(Xg) \rangle_\rho. \quad (68)$$

This imply that a polynomial of order n generated by G_A is mapped by \mathcal{R}_ρ^\dagger to a polynomial of order n generated by G_B , which is then mapped back to a polynomial of order n generated by G_A .

Therefore, we conclude that the principal observables are polynomials generated by G_A . Finding the exact polynomials of a given degree can be done for each order independently, which is a finite-dimensional problem.

This statement is in fact true for all quantum generalizations of the Fisher information metric. Classically, the Fisher information metric is characterized as the only metric on the manifold of probability distributions which contracts under the action of any stochastic map. In the quantum case, Petz and Sudár [21] characterized all contractive metrics. They are defined by an *operator monotone* function $\theta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $\theta(t) = t\theta(t^{-1})$ for all $t > 0$. An operator monotone function has the property that, when applied to operators via functional calculus, $\theta(A) \leq \theta(B)$ whenever $A < B$ (i.e., $B - A$ is positive). The function θ defines the kernel Ω_ρ^{-1} via its inverse Ω_ρ as follows:

$$\Omega_\rho = R_\rho^{\frac{1}{2}} \theta \left(L_\rho R_\rho^{-1} \right) R_\rho^{\frac{1}{2}}, \quad (69)$$

where $R_\rho(A) := A\rho$ and $L_\rho(A) := \rho A$ for any matrix A . It is straightforward to adapt the above argument to this general form.

References

- [1] Wilson K G and Kogut J B 1974 The renormalization group and the epsilon expansion *Phys. Rep.* **12** 75–200
- [2] Wilson K G 1975 The renormalization group: critical phenomena and the Kondo problem *Rev. Mod. Phys.* **47** 773–840
- [3] Fisher M E 1998 Renormalization group theory: its basis and formulation in statistical physics *Rev. Modern Phys.* **70** 653–81
- [4] Barenblatt G I 1996 *Scaling, Self-Similarity, and Intermediate Asymptotics: Dimensional Analysis and Intermediate Asymptotics* vol 14 (Cambridge: Cambridge University Press)
- [5] Preskill J 2000 Quantum information and physics: some future directions *J. Mod. Opt.* **47** 127–37
- [6] Kadanoff L P 1976 Notes on Migdal's recursion formulas *Ann. Phys.* **100** 359–94
- [7] Kadanoff L P 1966 Scaling laws for Ising models near T_c *Physics* **2** 263–72
- [8] Kadanoff L P 1977 The application of renormalization group techniques to quarks and strings *Rev. Mod. Phys.* **49** 267–96
- [9] Zamolodchikov A B 1986 'Irreversibility' of the flux of the renormalization group in a 2D field theory *JETP Lett.* **43** 730–2
Zamolodchikov A B 1986 *Pis'ma Zh. Èksper Teoret. Fiz.* **43** 565–7 (Russian transl.)
- [10] Apenko S M 2012 Information theory and renormalization group flows *Physica A* **391** 62–77
- [11] Brody D C and Ritz A 1998 On the symmetry of real-space renormalization *Nucl. Phys. B* **522** 588–604
- [12] Casini H and Huerta M 2007 A c-theorem for entanglement entropy *J. Phys. A* **40** 7031–6
- [13] Gaiete J and O'Connor D 1996 Field theory entropy, the H theorem, and the renormalization group *Phys. Rev. D* **54** 5163–73
- [14] Heckman J J 2013 Statistical inference and string theory (arXiv:1305.3621)
- [15] Machta B B, Chachra R, Transtrum M K and Sethna J P 2013 Parameter space compression underlies emergent theories and predictive models *Science* **342** 604–7
- [16] Hiai F and Petz D 1991 The proper formula for relative entropy and its asymptotics in quantum probability *Commun. Math. Phys.* **143** 99–114
- [17] Petz D 1996 Monotone metrics on matrix spaces *Linear Algebr. Appl.* **244** 81–96
- [18] Ohya M and Petz D 2004 *Quantum Entropy and its Use* (Berlin: Springer)
- [19] Rovelli C and Vidotto F 2014 *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory* (Cambridge: Cambridge University Press)
- [20] Bény C and Osborne T J 2012 Information geometric approach to the renormalisation group (arXiv:1206.7004)
- [21] Petz D and Sudár C 1996 Geometries of quantum states *J. Math. Phys.* **37** 2662–73