symmetry

MDPI

Article

Signal Detection in Nearly Continuous Spectra and $\mathbb{Z}_2$-Symmetry Breaking

Vincent Lahoche [1], Dine Ousmane Samary [1,2,*] and Mohamed Tamaazousti [1]

[1] Université Paris-Saclay, CEA, List, F-91120 Palaiseau, France; vincent.lahoche@cea.fr (V.L.); mohamed.tamaazousti@cea.fr (M.T.)
[2] International Chair in Mathematical Physics and Applications (ICMPA-UNESCO Chair), Cotonou 072B.P.50, Benin
* Correspondence: dine.ousmanesamary@cipma.uac.bj

Abstract: The large scale behavior of systems having a large number of interacting degrees of freedom is suitably described using the renormalization group from non-Gaussian distributions. Renormalization group techniques used in physics are then expected to provide a complementary point of view on standard methods used in data science, especially for open issues. Signal detection and recognition for covariance matrices having nearly continuous spectra is currently an open issue in data science and machine learning. Using the field theoretical embedding introduced in Entropy, 23(9), 1132 to reproduce experimental correlations, we show in this paper that the presence of a signal may be characterized by a phase transition with $\mathbb{Z}_2$-symmetry breaking. For our investigations, we use the nonperturbative renormalization group formalism, using a local potential approximation to construct an approximate solution of the flow. Moreover, we focus on the nearly continuous signal build as a perturbation of the Marchenko-Pastur law with many discrete spikes.

Keywords: renormalization group; field theory; symmetry breaking; data analysis; signal detection

PACS: 05.10.Cc; 05.40.-a; 07.05.Kf

check for updates

Citation: Lahoche, V.; Ousmane Samary, D.; Tamaazousti, M. Signal Detection in Nearly Continuous Spectra and $\mathbb{Z}_2$-Symmetry Breaking. *Symmetry* **2022**, *14*, 486. https://doi.org/10.3390/sym14030486

Academic Editor: Axel Pelster

Received: 23 November 2021
Accepted: 10 February 2022
Published: 28 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The renormalization group (RG) is one of the most important discoveries of the twentieth century in physics. It is a more general idea rather than a specific law of nature, aiming to extract relevant features of statistical or quantum states in a modern conception due to [1,2]. Introduced in the area of statistical physics, it is, in particular, the most powerful concept at explaining the universality of large distance physics for systems involving a very large number of interacting degrees of freedom without requiring a complete description of these fundamental degrees of freedom. The RG explains the universality and efficiency of effective descriptions of physical laws through a progressive dilution of information with *coarse-graining*, which is absorbed into the *running* parameters defining effective theory [3,4]. The most universal formalization of the RG [5–13] is based on the existence of an intrinsic hierarchy of degrees of freedom in such a way that we can progressively ignore some of them, and "integrated" in a less fundamental effective description for the remaining ones. For this reason, the RG is particularly relevant in many-body physics, for all problems involving a very large number of interacting degrees of freedom. In physics, this hierarchy is intrinsically related to the notion of scale; and RG aims to construct large scale effective theories integrating out microscopic degrees of freedom in such a way to preserve long-distance physics (see [5–13]). More generally, the Kadanoff and Wilson idea is the statement that the best way to study a sub-number of degrees of freedom in a large system is to integrate out the remaining degrees of freedom. Standard incarnations of the RG take the form of a flow in the formal space of Hamiltonian's

(log-likelihood in probability theory), describing a sequence of distributions having the same long-distance physics.

Data analysis and machine learning both aim to extract relevant features from among data sets of very large dimensions. This is in particular the case within the big data paradigm. Principal component analysis (PCA) (see [14–28]) looks for a linear projection into a lower-dimensional space, keeping only the relevant features, which is exactly what the RG aims to do. For this reason, the RG is expected to be a relevant and competitive approach to standard PCA, and perhaps a solid alternative in context where PCA fails. Standard PCA work well for spectra involving a small number of discrete spikes. In that way, one expect the existence of a "gap" in eigenvalues, meaning that only a small number of eigenvalues capture a large fraction of the total covariance matrix. In this paper, we focus especially on the nearly continuous spectra issue. For such a spectrum the gap goes to zero, and standard PCA fails to provide a clean separation between "what is relevant" and "what we can ignore". In a computational point of view, the origin of the problem can be traced from the intrinsic computational hardness of finding optimal *k-means clustering* (the planar k-means problem being, for instance, NP-Hard), see [29]. For this reason, RG is expected to be a relevant and competitive approach to standard PCA. This can be achieved through a field theoretical embedding, as considered in [15,16], from an analogy with what happens in standard field theory. This field theory can be viewed as describing some unconventional "matter" filling an abstract space of dimension 1. For ordinary field theories, the number of relevant terms in the Hamiltonian spanning the distinguishable distributions at large scales depend on the dimension of space d. However, a moment of reflection shows that it is indeed a property of the momentum distribution $\rho(p^2) = (p^2)^{d/2-1}$ involved in loop integrals. From this elementary observation, it seems reasonable to investigate the RG flow associated with the eigenvalue distribution of the covariance matrix through a suitable field theoretical embedding that is able to reproduce (at least partially) the data correlations and to extract the relevant features of the distributions. Note that such a strategy follows the current point of view of field theory, understood as effective descriptions at a large scale of some partially understood microscopic physics [1,2,30]. In this way, a signal could be differentiated from noise by simple comparisons of the universality classes generated by the relevant couplings. Note, moreover, that such a strategy does not allow *à priori* to inferring of effective properties of data. In this paper, we only claim to build an effective theory, in the same class of long-range equivalence as the "true" theory (see Figure 1).

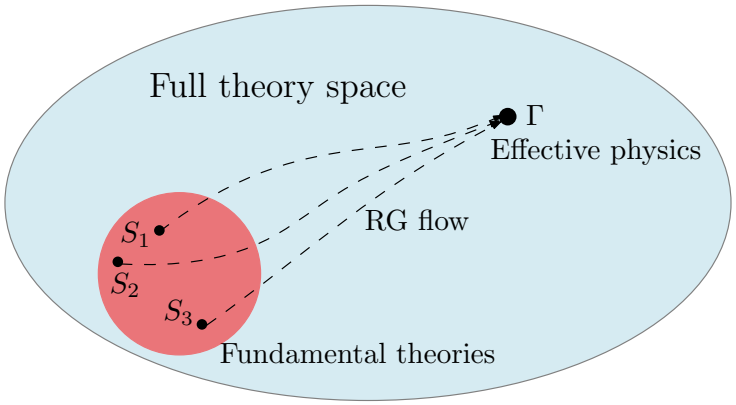

Figure 1. Qualitative picture of the RG flow through the theory space. The initial Hamiltonians $\mathcal{S}_1$, $\mathcal{S}_2$ ... in the red bubble correspond to different microscopic physics in the same basin of attraction toward IR scales described by effective Hamiltonian.

This point of view was the one developed in [14–17]. In these papers, the authors were able to characterize the presence of a signal, and estimate the breaking point between signal and noise, by the fact that the first non-Gaussian perturbation, which is relevant for a purely MP distribution, becomes irrelevant for a sufficiently strong signal. In this

paper, we focus precisely on the asymptotic aspects (IR) attached to the signal, and we show that a phase transition, corresponding to a breaking of reflection symmetry, can be associated with it. We justify the existence of an intrinsic detection threshold and show how this threshold could be considered for the construction of a functional detection algorithm. Finally, we mention some open questions.

The paper is organized as follows: In Section 2 we provide a short state of the art, allowing us to position our work in the existing literature, especially in regard to the continuous spectra issue. In Section 3 we present the theoretical framework and the functional renormalization group for this unconventional field theory. In Section 4 we investigate the behavior of the RG flow. We particularly focus on the shape of the effective potential, and show that the size of the so-called symmetric phase decreases as the signal strength disturbing a given background increases. This enforces a parallel between signal detection and symmetry breaking; an intrinsic notion of "detection threshold" emerging from the formalism. Section 5 is devoted to the conclusion and remarks.

2. Related Works

In this work, we continue our efforts to connect the important concept of renormalization group (RG) with machine learning and data analysis tools [27,31–40].

In this investigation, our work follows a fruitful path for the use of RG in a context connected with PCA [14,15,17]. Namely, the challenging context studied in this series of works tries to address, with the lens of RG, the problem of signal detection in a continuous spectrum where the distinction between signal and noise appears to be arbitrary with the standard PCA.

It has already been shown that this continuous spectrum problem arises quite naturally in different practical applications: in the study of neural activity data [41–43], in biology [44,45], in particular with the study of single-cell data [46,47], in genetic data [48], and in financial data [49,50]. Moreover, as this problem is related to the PCA, one might expect even more applications in the future.

Following the non-perturbative framework introduced in [14,15], we have shown in this study that this new way of detecting a signal in such continuous spectra is related to a symmetry-breaking phenomenon. Note that this phase transition finding is in the same vein as those well studied in the spiked model associated with the PCA [51–54].

Finally, we provide some arguments to emphasise the genericity of the proposed approach. First, recall that this framework is based on the following assumptions, which are quite general: (i) Sufficient proximity to the continuous approximation so that the sums can be suitably replaced by integrals on the eigenvalue spectrum distribution; (ii) The effective Gibbs distribution with which we are working can reproduce the exact correlations at two points (given by the correlation matrix) in the limit where the cumulants of order greater than two cancel each other out. Second, because of the generality of these assumptions, this result is of course not limited to the Marchenko-Pastur distribution that we explored in this paper as an example to illustrate the symmetry-breaking phenomenon. Indeed, there is some evidence for the universality of this result concerning general statistical noise models associated with continuous and positive spectra. In particular, it has been recently been shown in [16] that this result is also valid for tensor generalizations of the covariance matrix which are involved in the important tensor PCA problem [55–58].

3. Framework

3.1. The Model

We consider a set of data described by a big $N \times P$ matrix X_{ia} for $i = 1, 2, \cdots, N$ and $a = 1, 2, \cdots P$. We assume $N, P \gg 1$. The covariance matrix C is the $N \times N$ entries $C_{ij} = \sum_{a=1}^{P} X_{ia} X_{ja}$. Moreover, when the entries of X are *independent identically distributed* (i.i.d.) variables, the eigenvalues of the matrix C_{ij}/N converge in the weak topology in distribution toward the MP law [59]. Figure 2 below provides a typical spectrum for $P = 1500$ and $N = 2000$. We denote by $\mu_{\exp}(\lambda)$ the empirical eigenvalue distribution.

Let us recall the standard MP theorem [60]. We consider a $N \times P$ random matrix X, whose entries X_{ia} are (i.i.d.) with zero mean and finite variance σ^2 ($< \infty$). We moreover define the $N \times N$ Wishart matrix Y with entries $Y_{ij} := \frac{1}{P}\sum_{a=1}^{P} X_{ia}X_{ja}$. Taking limits $N, P \to +\infty$ such that $N/P = \alpha \in (0, +\infty)$, the eigenvalue distribution $\mu_{\exp}(\lambda)$ converge toward the MP law:

$$\mu_{MP}(\lambda > 0) = \frac{1}{2\pi\sigma^2}\frac{\sqrt{(\lambda - \lambda_-)(\lambda_+ - \lambda)}}{\lambda\alpha} 1_{[\lambda_-,\lambda_+]}(\lambda)\,, \tag{1}$$

where $\lambda_\pm = (1 \pm \sqrt{\alpha})^2$ and $1_{[a,b]}(x) = 1$ for $x \in [a, b]$ and zero otherwise. Note that for $P/N < 1$ a contribution $(1 - 1/\alpha)\delta(\lambda)$ have to be added. Restricting on the values $\lambda > 0$ avoids to consider this contribution which play no role in the following.

In [14,15] the authors introduced a field theoretical embedding aiming to reproduce data correlations. The framework describes a nearly continuous random field $\varphi(p) \in \mathbb{R}$, the variables p being defined such that p^2 is an eigenvalue of the inverse covariance matrix (Assuming zero modes of C are removed.) translated from its largest eigenvalue $1/\lambda_0$ (disregarding quantum corrections). The field is provided with a probability density $p[\varphi] := e^{-\mathcal{S}[\varphi]}$. The functional Hamiltonian $\mathcal{S}$ being defined as:

$$\mathcal{S}[\varphi] := \frac{1}{2}\sum_p \varphi(p)(p^2 + m^2)\varphi(-p) + gU[\varphi]\,. \tag{2}$$

For $g = 0$, the model is purely Gaussian, and the 2-point correlations functions $\langle\varphi(p)\varphi(p')\rangle = (p^2 + m^2)^{-1}\delta_{p,-p'}$, where δ is the Kronecker delta and the notation $\langle X[\varphi]\rangle$ denotes the mean value of the quantity X with respect to the probability measure $e^{-\mathcal{S}[\varphi]}\prod_p d\varphi(p)$. In that case, we reproduce exactly the experimental 2-point correlations given by the eigenvalues of the covariance matrix if, firstly $m^2 = 1/\lambda_0$, and secondly the momenta p are such that p^2 is distributed following the eigenvalue distribution of the covariance matrix. We denote as $\rho(p^2)$ this eigenvalue distribution inferred from the knowledge of $\mu_{\exp}(\lambda)$ (see [14,15]):

$$\rho(p^2) = \frac{\mu_{\exp}\left(\frac{1}{p^2 + 1/\lambda_0}\right)}{(p^2 + 1/\lambda_0)^2}\,, \tag{3}$$

the integration measure for the variable p reading as: $\rho(p^2)pdp$. For MP law, we get explicitly:

$$\rho_{MP}(p^2) = \frac{\sqrt{\lambda_+\lambda_-}}{2\pi\alpha\sigma^2}\frac{(p^2)^{1/2}}{(p^2 + 1/\lambda_+)^2}\left(\frac{\lambda_+ - \lambda_-}{\lambda_+\lambda_-} - p^2\right)^{1/2}, \tag{4}$$

and for p small enough, $\rho_{MP}(p^2) \sim (p^2)^{1/2}$.

The existence of n-points correlations functions which cannot be decomposed as a product of 2-point functions accordingly to the Wick theorem require to remove the condition $g = 0$. The functional $U[\varphi]$ is assumed to be a *conservative* and $\mathbb{Z}_2$-invariant polynomial in φ of the form:

$$U[\varphi] = \sum_{L=1}^{M} u_{2L}\,\delta_{0,\sum_{\alpha=1}^{2L} p_\alpha}\prod_{\alpha=1}^{2L}\varphi(p_\alpha)\,. \tag{5}$$

The model (2)–(5) is conservative in the usual sense in field theory, meaning that momenta are conserved at each vertex. The choices of these interactions and the reflection symmetry $\varphi \to -\varphi$, that we assume in this paper are extensively discussed in [14–16]. Indeed, we are aiming to only construct an approximation and extract some relevant features concerning the momenta distributions able to discriminate between data and noise.

Let us recall some of the main arguments underlying the choice of this model. In the cases of practical interest the data sets describe a large number of degrees of freedom that are not independent, but for which we have only partial knowledge of the probability distribution of the different configurations. Generally, what we know empirically are the

first and second moments, i.e., the mean and the variance. In such a situation, the *maximum entropy estimate* is the best compromise (i.e., the least structured distribution) corresponding to the empirical data [42,61,62]. One way to arrive at an action of the form of (2) is to consider a specific case, from which we can easily proceed to an analytical derivation. One can, for example, consider a system formed by N spins $s_i = \pm 1$ attached to the sites of a lattice having an arbitrarily complicated structure. Assuming that the distribution has 0 mean and covariance C, the maximum entropy estimate for probability distribution $p[\sigma]$ of spin configurations $\sigma = \{s_i\}$ match with the standard Gibbs-Boltzmann states,

$$p[\sigma] \propto \exp\left(\frac{1}{2}\sum_{i,j=1}^{N} s_i K_{ij} s_j\right), \tag{6}$$

for some matrix K_{ij} matching with the constraint $\sum_\sigma p[\sigma] s_i s_j = C_{ij}$. The derivation of the underlying field theory follows the standard strategy allowing for the construction of the Ginzberg-Landau model [63,64]. The corresponding distribution is now for a continuous field $\phi_i \in \mathbb{R}$, and we arrive to the formal expression [16] $p[\phi] = \exp(-H[\phi])$, with:

$$H[\phi] = \frac{1}{2}\sum_{i,j=1}^{N} \phi_i \left(C_{ij}^{-1} - \delta_{ij}\right)\phi_j + \frac{1}{12}\sum_{i=1}^{N} \phi_i^4 + \cdots, \tag{7}$$

whose model (2)–(5) can be seen as a continuous limit. Note that such a continuum limit is not too artificial. Indeed, nearly continuous spectra generally have a power-law behavior at the tail, and such an effective field as introduced above is expected to behave as do ordinary field theories. This, however, poses the problem of the dimension. For the MP law, for instance, the momentum distribution at the tail of the spectrum behaves like $\rho_{MP}(p^2) \propto (p^2)^{1/2}$, which corresponds to a field theory in dimension $d = 3$. Obviously, finite size effects, or the presence of a signal, will modify the value of this effective dimension, which in not an integer in general. To avoid this difficulty, we set the dimension of the space to 1 in the definitions (2)–(5), while keeping a non-trivial moment distribution $\rho(p^2)$; the last ensuring that the dominant effects of the RG are the same for the effective theory or for a more exact and particular model. Note that it has been shown for a 2D Ising model on a rectangular lattice [65,66], and making numerical simulations using the Metropolis algorithm as time-evolution that the spectral density changes shape in the vicinity of the critical temperature, behaving like a power law, whereas it agrees with the predictions of the MP law for high temperatures.

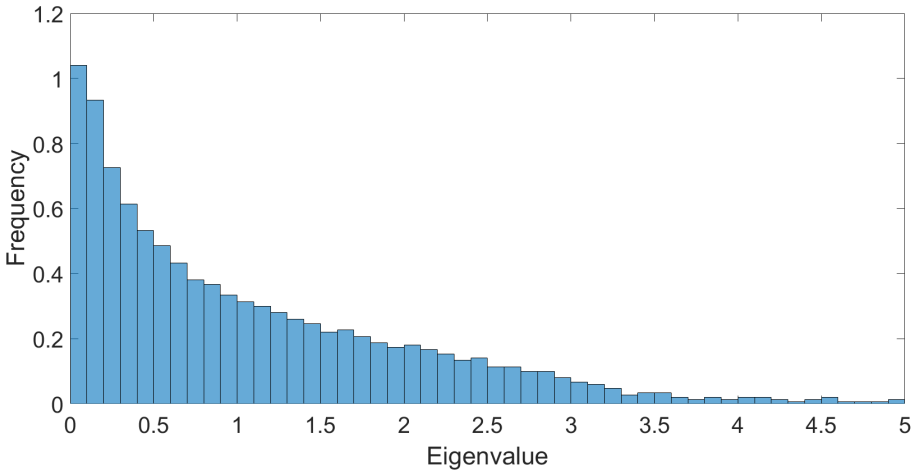

Figure 2. A nearly continuous spectra obtained as a disturbance (by adding a large number of spikes) to the spectrum corresponding to the covariance matrix of a random matrix with i.i.d. entries. For such a spectrum standard PCA fails to provide a clean separation between relevant and irrelevant degrees of freedom.

Everything that follows in this article could be concerned only with this particular model derived in the special context of discreet spins having complex correlations. But a moment's reflection shows that it must be more general. Indeed, one of our essential results is that the presence of a signal in a spectrum in the neighborhood of the universal MP law corresponds to a phase transition breaking the $\mathbb{Z}_2$ symmetry. But a spectrum like the one in Figure 2 is completely blind to the nature of the microscopic degrees of freedom. They could be spins, financial data, or anything else. The fact that one is in the neighborhood of a universality class (that of MP) implies that any particular effective model for the data must also have a universal aspect, independently of the particular nature of the microscopic degrees of freedom. This is exactly what the results of this paper and its predecessors [14–16] highlight. More details can be found in the references, as well as in the forthcoming paper [67].

3.2. Functional Renormalization Group Formalism

The RG flow can be constructed from the standard Wilson-Kadanoff procedure, partially integrating over modes having high momenta (ultraviolet (UV) modes). In such a field framework, it is suitable to use the functional renormalization group (FRG) to construct approximate solutions of the RG flow beyond perturbation theory (see [5–13]). The FRG is based on the effective Hamiltonian for integrated modes below some scale k rather than on the Hamiltonian for the remaining, not integrated modes above the scale k (infrared (IR) modes). The effective Hamiltonian for integrated degrees of freedom is denoted as $\Gamma_k[M]$ and obeys to the first order differential equation [10]:

$$k\frac{d}{dk}\Gamma_k = \frac{1}{2}\sum_p \dot{r}_k(p^2)\left(\Gamma_k^{(2)} + r_k\right)^{-1}_{p,-p}. \tag{8}$$

In this equation:

- $r_k(p^2)$, the regulator, plays the role of an effective mass, depending both on momenta and infrared cut-off k. It vanishes for high momenta with respect to k ($p^2/k^2 \gg 1$), whereas low momenta modes are frozen, and decouple from long distance physics. Moreover, $r_k(p^2)$ vanishes for $k = 0$, ensuring that all the modes are integrated out.
- The effective averaged Hamiltonian $\Gamma_k[M]$ is defined from a slight modified version of the Legendre transform for free energy $W_k[j]$:

$$\Gamma_k[M] + W_k[j] = \sum_p j(-p)M(p) + \Delta\mathcal{S}_k[M]\,, \tag{9}$$

  where $\Delta\mathcal{S}_k[\varphi] := \frac{1}{2}\sum_p \varphi(p)r_k(p^2)\varphi(-p)$. The free energy $W_k[j]$ being the generating functional of cumulants, $W_k[j] := \ln\left\langle \exp\left(\sum_p j(-p)\varphi(p) + \Delta\mathcal{S}_k[\varphi]\right)\right\rangle$. This definition ensures that Γ_k reduces to the microscopic Hamiltonian $\mathcal{S}$ in the deep UV ($k^2 \gg 1$), where $r_k(p^2)$ is expected to be of order k^2. Moreover, for $k = 0$, $r_k(p^2)$ vanishes, and Γ_k reduces formally to the full effective Hamiltonian Γ, with all modes integrated out.
- The notation $\Gamma^{(2)}$ means second derivative with respect to M, the classical field defined as:

$$\frac{\partial W_k[j]}{\partial j(-p)} = M(p)\,. \tag{10}$$

The exact flow equation (8) works in an infinite-dimensional space of functions, and cannot be solved exactly in general. A standard method to construct approximate solutions is to truncate into a finite-dimensional subspace, assumed to be relevant from physical conditions. In this paper, we focus on the local potential approximation (LPA), assuming that the non-quadratic part of Γ_k may be spanned by local interactions of the form (5). For the quadratic part, we use standard derivative expansion (DE), keeping only couplings of order p^2,

$$\Gamma_{k,\mathrm{kin}}[M] = \frac{1}{2}\sum_p M(-p)(p^2+u_2)M(p) + \mathcal{O}(p^2)\,, \tag{11}$$

assuming to work in the IR region, and following the standard LPA assumptions, we project the flow equation on a constant classical field, neglecting its momentum dependence: $M(p) = M\delta_{p0}$. It is suitable to include the term of order $(p^2)^0$ in the non-quadratic part. Denoting it as U_k, we assume the following expansion around non-vanishing vacuum κ for constant classical field:

$$U_k[\chi] = \frac{u_4}{2!}(\chi-\kappa)^2 + \frac{u_6}{3!}(\chi-\kappa)^3 + \cdots\,, \tag{12}$$

where $\chi = M^2/2$. It is well known that truncation procedure may introduce a spurious dependence on the regulator [68]. However, as our investigations focus on the actual potential, the potentially undesirable effects of the regulator are expected to be marginal. Moreover, we will focus on the Litim regulator, which is optimal for the kind of truncation we will consider [11,12] and allows us to compute analytically the loop integrals:

$$r_k(p^2) = (k^2-p^2)\theta(k^2-p^2)\,, \tag{13}$$

θ being the Heaviside step function.

The flow equation for the potential U_k can be deduced from the Equation (8), setting constant classical field:

$$k\frac{d}{dk}U_k[\chi] = \left(2\int_0^k \rho(p^2)pdp\right)\times\frac{k^2}{k^2+\partial_\chi U_k(\chi)+2\chi\partial_\chi^2 U_k(\chi)}\,.$$

It is suitable to introduce the flow parameter $\tau := \ln\int_0^k p\rho(p^2)dp$ rather than k. Moreover, from the interpretation of the parameter u_2 as the asymptotic effective mass, it is suitable to assume the scaling $u_2 \sim k^2$. In such a way, we can define a scaling (or canonical) dimension for all of the couplings. In standard field theory, this scaling dimension allows for the converting of the RG equations as an autonomous system. This is not true here, because the shape of the momentum distribution is not invariant from dilatation. As an analogy, it is as if the dimension of the effective space (fixing the shape of the distribution in moments) depends on the scale k. The dimension of the operators in this context must therefore also depend explicitly on the scale, and the equations never form an autonomous system. The best compromise one can hope for is to deport all the explicit dependence into the scaling dimensions. The reader may find some details in [14,15]. From this requirement, one expects to define dimensionless quantities denoted with a "bare" as:

$$\partial_\chi U_k(\chi)k^{-2} = \partial_{\bar\chi}\bar U_k(\bar\chi)\,,\quad \chi\partial_\chi^2 U_k(\chi)k^{-2} = \bar\chi\partial_{\bar\chi}^2\bar U_k(\bar\chi)\,, \tag{14}$$

leading to:

$$U_k'[\chi] = \left(\frac{dt}{d\tau}\right)^2\frac{k^2\rho(k^2)}{1+\partial_{\bar\chi}\bar U_k(\bar\chi)+2\bar\chi\partial_{\bar\chi}^2\bar U_k(\bar\chi)}\,; \tag{15}$$

with the notation $X' := dX/d\tau$. Hence from (15) and (14), it is suitable to define:

$$U_k[\chi] := \bar U_k[\bar\chi]k^2\rho(k^2)\left(\frac{dt}{d\tau}\right)^2,\quad \chi = \rho(k^2)\left(\frac{dt}{d\tau}\right)^2\bar\chi\,. \tag{16}$$

The flow equation for the "dimensionless" parameter follows:

$$\bar U_k'[\bar\chi] = -\dim_\tau(U_k)\bar U_k[\bar\chi] + \dim_\tau(\chi)\bar\chi\frac{\partial}{\partial\bar\chi}\bar U_k[\bar\chi] + \frac{1}{1+\partial_{\bar\chi}\bar U_k(\bar\chi)+2\bar\chi\partial_{\bar\chi}^2\bar U_k(\bar\chi)}\,,$$

where:

$$\dim_\tau(U_k) = t' \frac{d}{dt} \ln\left(k^2 \rho(k^2) \left(\frac{dt}{d\tau}\right)^2\right), \tag{17}$$

and:

$$\dim_\tau(\chi) = t' \frac{d}{dt} \ln\left(\rho(k^2) \left(\frac{dt}{d\tau}\right)^2\right). \tag{18}$$

The flow equations for couplings κ and u_{2n} maybe finally deduced from the condition:

$$\left.\frac{\partial U_k}{\partial \chi}\right|_{\chi=\kappa} = 0\,, \quad \left.\frac{\partial^n U_k}{\partial \chi^n}\right|_{\chi=\kappa} = u_{2n}\,. \tag{19}$$

We get for κ, u_4 and u_6:

$$\bar{\kappa}' = -\dim_\tau(\chi)\bar{\kappa} + 2\frac{3 + 2\bar{\kappa}\frac{\bar{u}_6}{\bar{u}_4}}{(1+2\bar{\kappa}\bar{u}_4)^2}\,, \tag{20}$$

$$\bar{u}_4' = -\dim_\tau(u_4)\bar{u}_4 + \dim_\tau(\chi)\bar{\kappa}\bar{u}_6 - \frac{10\bar{u}_6}{(1+2\bar{\kappa}\bar{u}_4)^2} + 4\frac{(3\bar{u}_4 + 2\bar{\kappa}\bar{u}_6)^2}{(1+2\bar{\kappa}\bar{u}_4)^3}\,. \tag{21}$$

and

$$\bar{u}_6' = -\dim(u_6)\bar{u}_6 - 12\frac{(3\bar{u}_4 + 2\bar{\kappa}\bar{u}_6)^3}{(1+2\bar{\kappa}\bar{u}_4)^4} + 40\bar{u}_6\frac{3\bar{u}_4 + 2\bar{\kappa}\bar{u}_6}{(1+2\bar{\kappa}\bar{u}_4)^3}\,. \tag{22}$$

The corresponding flow equation can be deduced following the same strategy. Figure 3 shows the canonical dimensions for the first local interactions with the pure MP law. This picture shows the existence of two regions. For the last third of the spectrum, only two couplings are relevant, the sixtic being asymptotically marginal in accordance to power-law counting (the MP law behaving as $\rho(p^2) \sim (p^2)^{1/2}$ for small p). In contrast, for the two first thirds of the distribution, the number of relevant interactions may be very large. As discussed in [15], standard methods in field theory do not work suitably in such a case. One should expect that the field theoretical approach is relevant only for the last level of the spectrum that we call the *learnable region*.

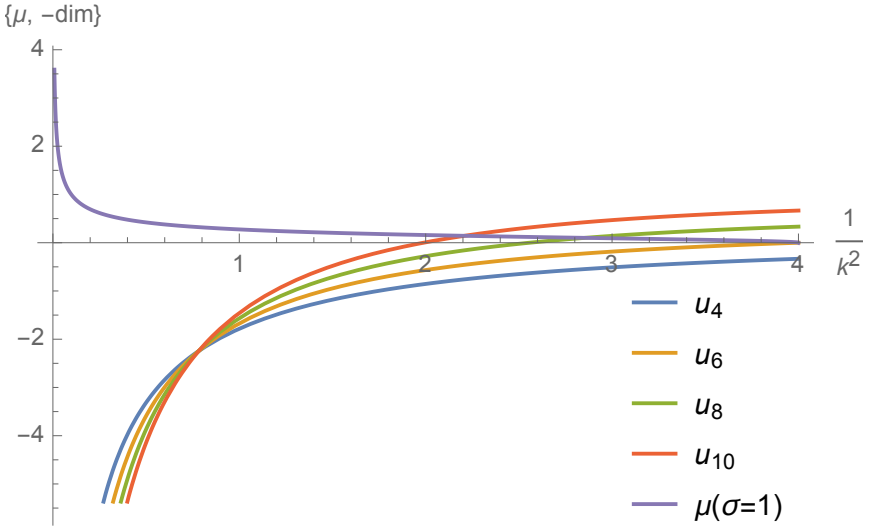

Figure 3. Canonical dimensions for the first even local interactions; for φ^4 (blue curve), φ^6 (orange curve), φ^8 (green curve) and φ^{10} (red curve) associated to the purely MP law (purple curve) with variance equals to 1.

4. $\mathbb{Z}_2$-Symmetry Breaking and Signal Detection

Besides these analytic considerations, we provide in this section the first look at a numerical investigation on a more realistic signal, as illustrated in Figure 2. In our experiments, we focus on the distribution of the eigenvalues for two types of covariance matrix in the regime of high dimensions (typically in our experiments we consider $P = 1500$ and $N = 2000$, which gives $K(= P/N) = 0.75$). First, we consider the covariance matrix associated with i.i.d. random entries. The distribution of the eigenvalues of such a matrix converges for large P and N but fixed ratio P/N toward the MP's law, which we interpret as a purely noisy data. Figure 2 corresponds to a random draw, for a matrix build as a perturbation of the pure noisy data with a deterministic matrix of rank $R = 65$ (defining the size of the signal). At this point, let us recall that we can characterize a signal by the localization property of the eigenvectors of the R rank matrix, contrasting with the delocalized nature of the degrees of freedom associated with noise (the bulk). In our experiments, we fix the variance to one and $K = 0.75$. For such a spectrum, the learnable region is expected between $\sim$2.5 and $\sim$3.4, where φ^4 and φ^6 are expected to be the only relevant interactions (this is information that one can get from the study of the canonical dimensions as illustrated in Figure 2).

To start, and following [14,15], we focus on the simpler version of the derivative expansion (DE), expanding the effective potential U_k as a power of $m := M/\sqrt{N}$:

$$U_k(m, \{u_{2n}\}) = \frac{1}{2} u_2 m^2 + \frac{u_4}{4!} m^4 + \frac{u_6}{6!} m^6 . \tag{23}$$

The derivation of the corresponding flow equations follows the same strategy as for (20), (21) and (22), see [14,15]. In Figure 4, we illustrated different viewpoints of the 3D compact region $\mathcal{R}_0$ in the vicinity of the Gaussian fixed point where the RG trajectories, obtained by the DE, ends in the symmetric phase, and thus are compatible with the symmetry restoration scenario for initial conditions corresponding to an explicit symmetry breaking. However, all these initial conditions are not expected to be physically relevant in the deep IR. Indeed, for scales $k^2 \sim 1/N$, one expects to obtain a good approximation for the exact covariance matrix. From construction, this imposes u_2 to reach a finite value of the order of the inverse of the larger eigenvalue of the spectrum. In turn, this imposes for the dimensionless parameter $\bar{u}_2$ to be of order N. The initial conditions compatibles with this requirement are pictured in blue on the figure.

In Figure 5 we show the same region $\mathcal{R}_0$ using LPA and Equations (20)–(22). We show that this region it is still compact, but its size reduces as the signal strength increases. This figure alone summarizes our findings. It shows that the appearance of a signal in the spectrum reduces the size of the symmetric phase, and that some trajectories that initially had their symmetry restored are now in the region of broken symmetry. The appearance of the signal thus leads to an effective breaking of the $\mathbb{Z}_2$ symmetry, which is illustrated in Figure 6 for a given trajectory in the purple region. On this figure we can follow the evolution of the effective potential in the deep IR (i.e., for $k^2 \sim 1/N$), for an increasing intensity of the signal which plays a role analogous to the temperature in the ordinary physics of phase transitions. The figure also allows us to understand the existence of a detection threshold. Indeed, if the presence of the signal reduces the size of the purple region, it still has to affect the physical states, i.e., the blue region. As long as we are in the intermediate region, the presence of the signal will have no appreciable effect on the physical states. Note that the very definition of these physical states depends on the precision of the measurements, and in particular on that of the largest eigenvalue.

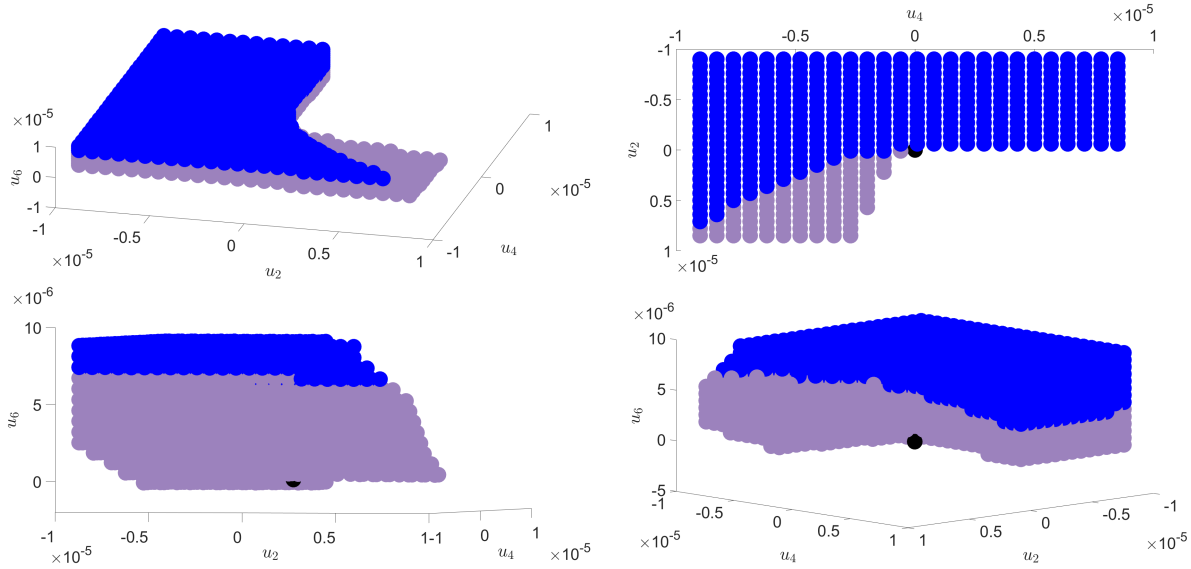

Figure 4. Different points of view of the compact region $\mathcal{R}_0$ (illustrated with purple dots) in the vicinity of the Gaussian fixed point (illustrated with a black dot) for the DE formalism. In this 3D region, corresponding to the case of pure noise, the RG trajectories ended in the symmetric phase, and thus are compatible with a symmetry restoration scenario for initial conditions corresponding to an explicit symmetry breaking. The blue dots correspond to RG trajectories associated to a physically relevant states in the deep infrared, i.e., the trajectories for which the values of $\bar{u}_2$ end with the same magnitude of $N = 2000$. Axes are relevant couplings u_2, u_4 and u_6.

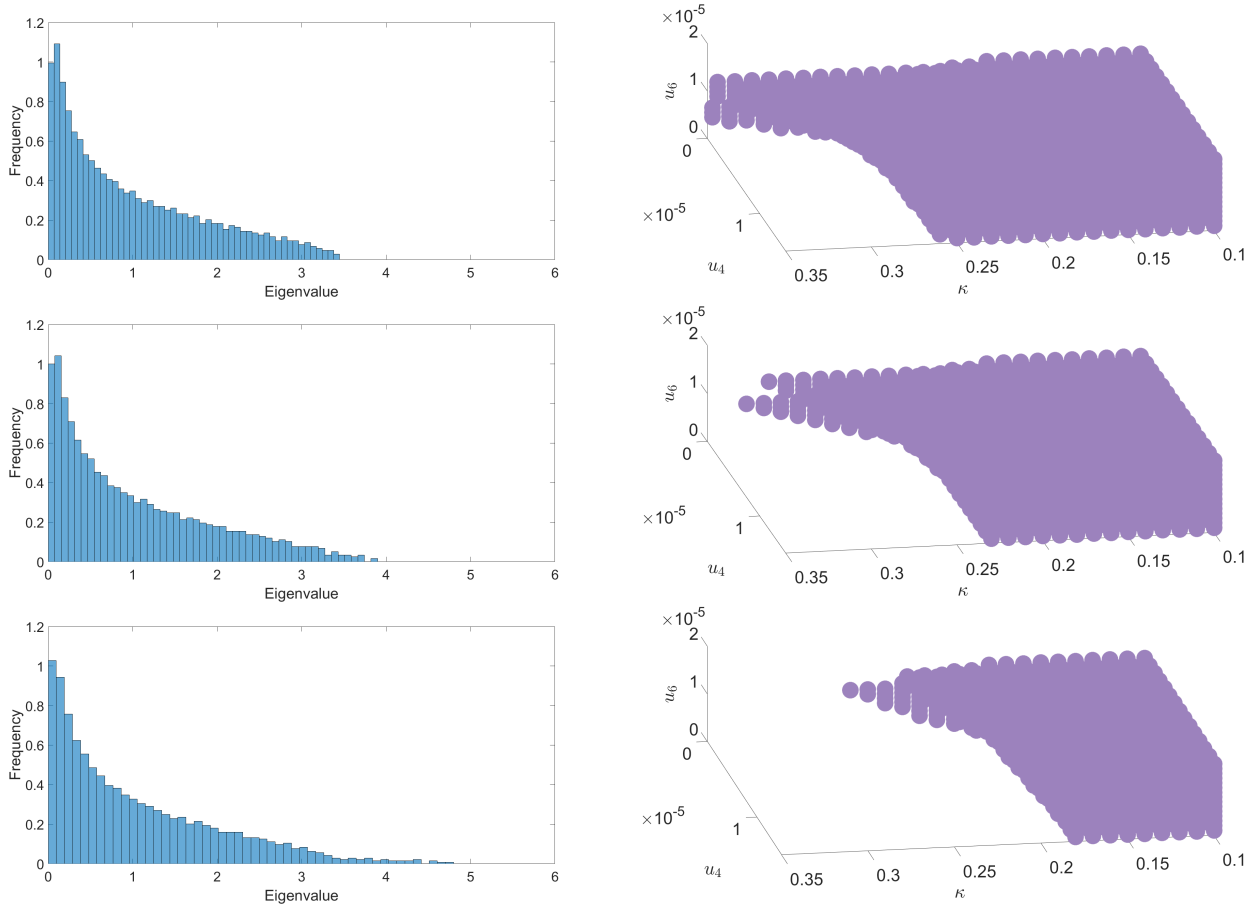

Figure 5. *Cont.*

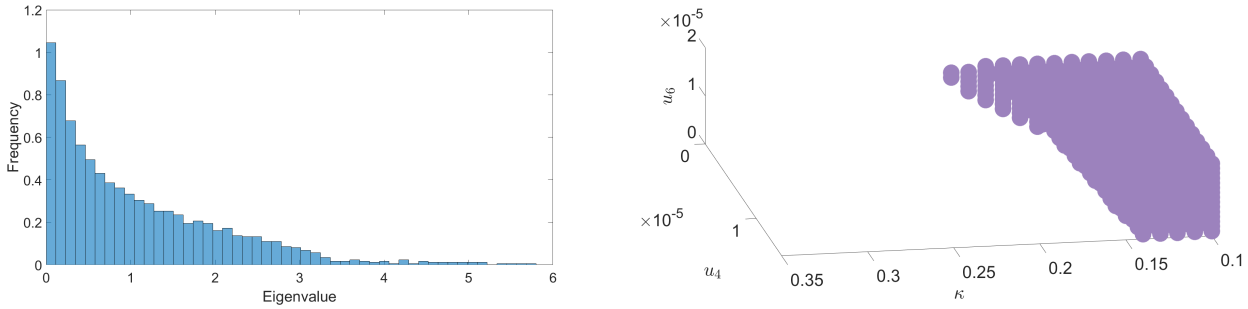

Figure 5. Left: The nearly continuous spectra obtained for different intensities of the signal. From top to bottom we apply respectively, 0, 40, 70 and 100 percent of the intensity of the original signal. Right: A view point of the respective 3D compact region $\mathcal{R}_0$ (illustrated with purple dots) for the LPA formalism. In these regions, the RG trajectories ended in the symmetric phase, and thus are compatible with a symmetry restoration scenario for initial conditions corresponding to an explicit symmetry breaking.

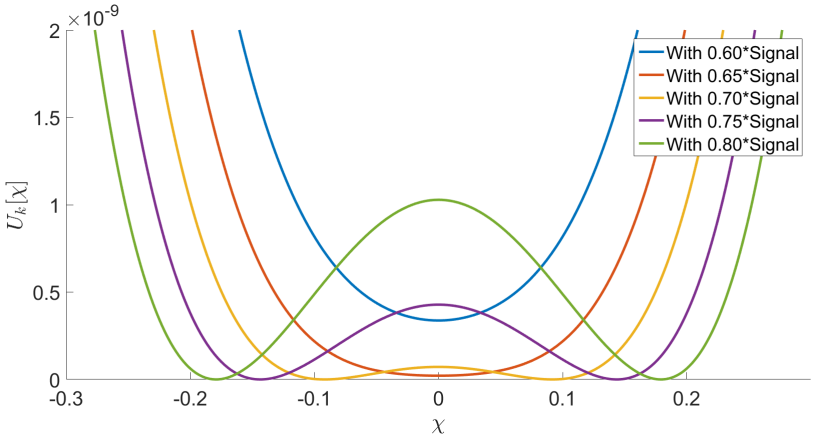

Figure 6. Illustration of the evolution of the potential with initial conditions in the purple region. The different curves correspond to different intensities of the signal.

5. Conclusions

In this paper, we investigated the RG of an effective field theory able to reproduce IR correlations at least partially in the learnable region, where both locals φ^4 and φ^6 are relevant. Focusing on local interactions, we constructed approximate solutions of the exact RG Equation (8) using standard DE and LPA. Some extended discussions can be found in [14–16], especially regarding the role of the anomalous dimension, which does not change our conclusions. Among the IR properties of the effective IR theories, we focused on the vacuum expectation value. We showed the existence of a nearly compact region $\mathcal{R}_0$ in the vicinity of the Gaussian fixed point where the $\mathbb{Z}_2$-symmetry is always restored in the deep IR for purely noisy signals that is described well by the MP law. Furthermore, we observed that the size of this region $\mathcal{R}_0$ is reduced when we consider a deviation by a signal to the asymptotic MP spectrum. Thus, this implies that some trajectories ending in the symmetric phase for pure noise end in a broken phase, with $\langle\varphi\rangle \neq 0$ when the signal is added. Moreover, among the initial conditions allowed by the region $\mathcal{R}_0$; only a subset of them are physically relevant, i.e., such that the inverse end mass u_2 is of the same magnitude as the expected largest eigenvalue of the (continuous part of the) spectrum. Thus, as soon as the deformation of the region $\mathcal{R}_0$ reaches one of these physical subregions, some physically relevant trajectories are affected and leave the symmetric phase in the deep IR. This observation exhibits the existence of an intrinsic sensitivity threshold for signal detection based on the asymptotic vacuum expectation value.

This observation allows for the consideration of a detection algorithm based on the existence of a phase transition in the deep IR region. This, however, remains an objective under investigation. In this study, we focus on synthetic data for which we have a good knowledge of the noise and signal notions, essentially to keep control on the perturbation. However, we plan to investigate this framework on real data. Other questions concern the phase transition, which seems to be able to be first or second order, depending on how the power count for the φ^6 coupling is affected. The nature of the transition could be linked to a finer detection criterion. Finally, other questions concern the approach. Investigations of regions of the more UV spectrum, for example, might require methods beyond standard DE. The validity of the field theory approximation could also be questioned in the UV. Moreover, the choice of the symmetry group can also be discussed. The $\mathbb{Z}_2$ symmetry seems to be a general and universal property at the neighborhood of the universality class of MP. However, other symmetries, more restrictive on the nature of the data and probably referring to finer levels of understanding, could be considered in the future. All of these questions are the subject of ongoing investigations [67].

Despite the fact our findings are based on a definition of noise based on the MP law, we planned to explore different mathematical incarnations of noisy signals, in a different context, to confirm the universal character of our conclusions. Thus, our investigations are also continuing for the Wigner distribution [69], as well as on more exotic distributions for data science based on tensors rather than on matrices, and for which standard tools are also more subject to technical limitations (see [70,71]).

To conclude, we would like to add that if we limited ourselves to artificial spectra to keep control of the signal strength and the different parameters in our experiments, the spectra themselves are not different from what we could obtain for real data in the neighborhood of a universality class like MP. Thus, our conclusions are indeed general properties of spectra, and could be used in practice through algorithms exploiting these properties, which is the subject of ongoing research.

Author Contributions: All the authors contributed with the same efficiencies in the realization of this manuscript. All authors have read and agreed to the published version of the manuscript..

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kadanoff, L.P.; Götze, W.; Hamblen, D.; Hecht, R.; Lewis, E.; Palciauskas, V.V.; Rayl, M.; Swift, J.; Aspnes, D.; Kane, J. Static phenomena near critical points: Theory and experiment. *Rev. Mod. Phys.* **1967**, *39*, 395. [CrossRef]
2. Wilson, K.G. The renormalization group: Critical phenomena and the Kondo problem. *Rev. Mod. Phys.* **1975**, *47*, 773. [CrossRef]
3. Bény, C. Coarse-grained distinguishability of field interactions. *Quantum* **2018**, *2*, 67. [CrossRef]
4. Bény, C.; Osborne, T.J. Information-geometric approach to the renormalization group. *Phys. Rev. A* **2015**, *92*, 022330. [CrossRef]
5. Delamotte, B. An introduction to the nonperturbative renormalization group. In *Renormalization Group and Effective Field Theory Approaches to Many-Body Systems*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 49–132.
6. Nagy, S. Lectures on renormalization and asymptotic safety. *Ann. Phys.* **2014**, *350*, 310–346. [CrossRef]
7. Blaizot, J.P.; Mendez-Galain, R.; Wschebor, N. Nonperturbative renormalization group and momentum dependence of n-point functions. I. *Phys. Rev. E* **2006**, *74*, 051116. [CrossRef] [PubMed]
8. Blaizot, J.P.; Mendez-Galain, R.; Wschebor, N. Nonperturbative renormalization group and momentum dependence of n-point functions. II. *Phys. Rev. E* **2006**, *74*, 051117. [CrossRef]
9. Berges, J.; Tetradis, N.; Wetterich, C. Non-perturbative renormalization flow in quantum field theory and statistical physics. *Phys. Rep.* **2002**, *363*, 223–386. [CrossRef]
10. Wetterich, C. Exact evolution equation for the effective potential. *Phys. Lett. B* **1993**, *301*, 90–94. [CrossRef]
11. Litim, D.F. Optimisation of the exact renormalisation group. *Phys. Lett. B* **2000**, *486*, 92–99. [CrossRef]
12. Litim, D.F. Derivative expansion and renormalisation group flows. *J. High Energy Phys.* **2001**, *2001*, 059. [CrossRef]
13. Manohar, A.V.; Nardoni, E. Renormalization Group Improvement of the Effective Potential: An EFT Approach. *arXiv* **2020**, arXiv:2010.15806 .
14. Lahoche, V.; Samary, D.O.; Tamaazousti, M. Generalized scale behavior and renormalization group for principal component analysis. *arXiv* **2020**, arXiv:2002.10574 .

15. Lahoche, V.; Ousmane Samary, D.; Tamaazousti, M. Field Theoretical Approach for Signal Detection in Nearly Continuous Positive Spectra I: Matricial Data. *Entropy* **2021**, *23*, 1132. [CrossRef]
16. Lahoche, V.; Ouerfelli, M.; Samary, D.O.; Tamaazousti, M. Field theoretical approach for signal detection in nearly continuous positive spectra II: Tensorial data. *Entropy* **2021**, *23*, 795. [CrossRef]
17. Bradde, S.; Bialek, W. Pca meets rg. *J. Stat. Phys.* **2017**, *167*, 462–475. [CrossRef]
18. Richard, E.; Montanari, A. A statistical model for tensor PCA. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2897–2905.
19. Woloshyn, R. Learning phase transitions: Comparing PCA and SVM. *arXiv* **2019**, arXiv:1905.08220 .
20. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417. [CrossRef]
21. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
22. Lu, H.; Plataniotis, K.N.; Venetsanopoulos, A.N. A survey of multilinear subspace learning for tensor data. *Pattern Recognit.* **2011**, *44*, 1540–1551. [CrossRef]
23. Guan, Y.; Dy, J. Sparse probabilistic principal component analysis. *Artif. Intell. Stat. J. Mach. Learn. Res.* **2009**, *5*, 185–192.
24. Seddik, M.E.A.; Tamaazousti, M.; Couillet, R. A kernel random matrix-based approach for sparse PCA. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
25. Foreman, S.; Giedt, J.; Meurice, Y.; Unmuth-Yockey, J. Machine learning inspired analysis of the Ising model transition. In Proceedings of the 36th Annual International Symposium on Lattice Field Theory (LATTICE 2018)—Theoretical Developments, East Lansing, MI, USA, 22–28 July 2018; Volume 334.
26. Bachtis, D.; Aarts, G.; Lucini, B. Adding machine learning within Hamiltonians: Renormalization group transformations, symmetry breaking and restoration. *Phys. Rev. Res.* **2021**, *3*, 013134. [CrossRef]
27. Bény, C. Inferring relevant features: From QFT to PCA. *Int. J. Quantum Inf.* **2018**, *16*, 1840012. [CrossRef]
28. Shlens, J. A tutorial on principal component analysis. *arXiv* **2014**, arXiv:1404.1100 .
29. Mahajan, M.; Nimbhorkar, P.; Varadarajan, K. The Planar k-Means Problem is NP-Hard. In Proceedings of the International Workshop on Algorithms and Computation, WALCOM, Kolkata, India, 18–20 February 2009; Volume 442, pp. 274–285. doi: [CrossRef]
30. Zinn-Justin, J. *From Random Walks to Random Matrices*; Oxford University Press: Oxford, UK, 2019.
31. Bény, C. Deep learning and the renormalization group. *arXiv* **2013**, arXiv:1301.3124 .
32. Mehta, P.; Schwab, D.J. An exact mapping between the variational renormalization group and deep learning. *arXiv* **2014**, arXiv:1410.3831 .
33. Shukla, M.; Thakur, A.D. An Enquiry on similarities between Renormalization Group and Auto-Encoders using Transfer Learning. *arXiv* **2021**, arXiv:2108.06157 .
34. Iso, S.; Shiba, S.; Yokoo, S. Scale-invariant feature extraction of neural network and renormalization group flow. *Phys. Rev. E* **2018**, *97*, 053304. [CrossRef]
35. Koch-Janusz, M.; Ringel, Z. Mutual information, neural networks and the renormalization group. *Nat. Phys.* **2018**, *14*, 578–582. [CrossRef]
36. Koch, E.D.M.; Koch, R.D.M.; Cheng, L. Is deep learning a renormalization group flow? *IEEE Access* **2020**, *8*, 106487–106505. [CrossRef]
37. Chung, J.H.; Kao, Y.J. Neural Monte Carlo Renormalization Group. *Phys. Rev. Res.* **2021**, *3*, 023230. [CrossRef]
38. Li, S.H.; Wang, L. Neural network renormalization group. *Phys. Rev. Lett.* **2018**, *121*, 260601. [CrossRef] [PubMed]
39. Halverson, J.; Maiti, A.; Stoner, K. Neural networks and quantum field theory. *Mach. Learn. Sci. Technol.* **2021**, *2*, 035002. [CrossRef]
40. Erbin, H.; Lahoche, V.; Samary, D.O. Nonperturbative renormalization for the neural network-QFT correspondence. *arXiv* **2021**, arXiv:2108.01403 .
41. Tkačik, G.; Marre, O.; Amodei, D.; Schneidman, E.; Bialek, W.; Berry, M.J. Searching for collective behavior in a large network of sensory neurons. *PLoS Comput. Biol.* **2014**, *10*, e1003408. [CrossRef]
42. Meshulam, L.; Gauthier, J.L.; Brody, C.D.; Tank, D.W.; Bialek, W. Collective behavior of place and non-place neurons in the hippocampal network. *Neuron* **2017**, *96*, 1178–1191. [CrossRef]
43. Meshulam, L.; Gauthier, J.L.; Brody, C.D.; Tank, D.W.; Bialek, W. Coarse–graining and hints of scaling in a population of 1000+ neurons. *arXiv* **2018**, arXiv:1812.11904 .
44. Agrawal, A.; Sarkar, C.; Dwivedi, S.K.; Dhasmana, N.; Jalan, S. Quantifying randomness in protein–protein interaction networks of different species: A random matrix approach. *Phys. A Stat. Mech. Its Appl.* **2014**, *404*, 359–367. [CrossRef]
45. Korošak, D.; Slak Rupnik, M. Random Matrix Analysis of Ca^{2+} Signals in β-Cell Collectives. *Front. Physiol.* **2019**, *10*, 1194. [CrossRef]
46. Aparicio, L.; Bordyuh, M.; Blumberg, A.J.; Rabadan, R. A random matrix theory approach to denoise single-cell data. *Patterns* **2020**, *1*, 100035. [CrossRef]
47. Johnson, E.; Kath, W.; Mani, M. EMBEDR: Distinguishing Signal from Noise in Single-Cell Omics Data. *Patterns* **2021**, *3*, 100443. [CrossRef]
48. Xu, Y.; Liu, Z.; Yao, J. ERStruct: An Eigenvalue Ratio Approach to Inferring Population Structure from Sequencing Data. *arXiv* **2021**, arXiv:2104.01944.

49. Laloux, L.; Cizeau, P.; Bouchaud, J.P.; Potters, M. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.* **1999**, *83*, 1467. [CrossRef]
50. Marsili, M. Dissecting financial markets: Sectors and states. *Quant. Financ.* **2002**, *2*, 297. [CrossRef]
51. Johnstone, I.M. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* **2001**, *29*, 295–327. [CrossRef]
52. Baik, J.; Arous, G.B.; Péché, S. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **2005**, *33*, 1643–1697. [CrossRef]
53. Paul, D. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Stat. Sin.* **2007**, *17*, 1617–1642.
54. Perry, A.; Wein, A.S.; Bandeira, A.S.; Moitra, A. Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Stat.* **2018**, *46*, 2416–2451. [CrossRef]
55. Montanari, A.; Richard, E. A statistical model for tensor PCA. *arXiv* **2014**, arXiv:1411.1076 .
56. Hopkins, S.B.; Shi, J.; Steurer, D. Tensor principal component analysis via sum-of-square proofs. In Proceedings of the Conference on Learning Theory, PMLR, COLT, Paris, France, 3–6 July 2015; pp. 956–1006.
57. Anandkumar, A.; Deng, Y.; Ge, R.; Mobahi, H. Homotopy analysis for tensor PCA. In Proceedings of the Conference on Learning Theory, PMLR, Amsterdam, The Netherlands, 7–10 July 2017; Volume 65, pp. 79–104.
58. Dudeja, R.; Hsu, D. Statistical query lower bounds for tensor pca. *J. Mach. Learn. Res.* **2021**, *22*, 1–51.
59. Marčenko, V.A.; Pastur, L.A. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik* **1967**, *1*, 457. [CrossRef]
60. Potters, M.; Bouchaud, J. In *A First Course in Random Matrix Theory (for Physicists, Engineers and Data Scientists)*; Cambridge University Press: Cambridge, UK, 2021. [CrossRef]
61. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. doi: 10.1103/PhysRev.106.620. [CrossRef]
62. Jaynes, E.T. Information Theory and Statistical Mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190. [CrossRef]
63. Itzykson, C.; Drouffe, J.M. *Statistical Field Theory: Volume 2, Strong Coupling, Monte Carlo Methods, Conformal Field Theory and Random Systems*; Cambridge University Press: Cambridge, UK, 1991; Volume 2.
64. Itzykson, C.; Drouffe, J.M. *Statistical Field Theory: Volume 1*; Cambridge University Press: Cambridge, UK, 1991; Volume 1.
65. Sps, V.; Seligman, T. Time series, correlation matrices and random matrix models. *AIP Conf. Proc.* **2014**, *1575*, 196. doi: [CrossRef]
66. Sps, V.; Prosen, T.; Buca, B.; Seligman, T. Correlation matrices at the phase transition of the Ising model. *arXiv* **2014**, arXiv:1403.7218.
67. Lahoche, V.; Samary, D.O.; Tamaazousti, M. Field theoretical approach for signal detection in nearly continuous positive spectra III: Universal features. *arXiv* **2022**, arXiv:2201.04250.
68. Pawlowski, J.M.; Scherer, M.M.; Schmidt, R.; Wetzel, S.J. Physics and the choice of regulators in functional renormalisation group flows. *Ann. Phys.* **2017**, *384*, 165–197. [CrossRef]
69. Wigner, E.P. On the distribution of the roots of certain symmetric matrices. *Ann. Math. Second Ser.* **1958**, *67*, 325–327. [CrossRef]
70. Lahoche, V.; Samary, D.O. Reliability of the local truncations for the random tensor models renormalization group flow. *Phys. Rev. D* **2020**, *102*, 056002. [CrossRef]
71. Lahoche, V.; Samary, D.O. Revisited functional renormalization group approach for random matrices in the large-N limit. *Phys. Rev. D* **2020**, *101*, 106015. [CrossRef]