

Research Article

Renormalization Group in Physics and Beyond

Antal Jakovac^{1,2*}

¹Department of Computational Sciences, Institute for Particle and Nuclear Physics, HUN-REN Wigner Research Centre for Physics, 29-33 Konkoly-Thege Mikló's Street, H-1121 Budapest, Hungary

²Department of Statistics, Institute of Data Analytics and Information Systems, Corvinus University of Budapest, 8 Fővám Square, H-1093 Budapest, Hungary

Abstract

We explain qualitatively in this short paper the essence of the renormalization group, showing that these ideas apply not only in physics but also beyond, forming the foundation of artificial intelligence.

1. Introduction

The world is full of facts: even in simplified systems that can be represented in computers, the configurations are usually represented by a lot of numbers. A hydrodynamic state, a configuration in a Monte Carlo simulation contains a lot of information.

On the other hand, observations are less numerous. In reality, we are only capable of making observations discretely, with finite temporal and spatial resolution, or, if the observation depends on parameters, with a discrete set of parameters. In computers we do not consider all possible measurable quantities "observation". Usually we consider the microscopic structure as unobservable, irrelevant details, and the long time and long distance "infrared" quantities matter. In thermodynamics we speak about "microstates" and "macrostates" exactly to represent this phenomenon [1,2].

To model this situation we may tell that the possible states of the world are $W \subset X$ where $X \sim \mathbb{R}^N$, while the observations form a functional space $O \subset \{W \rightarrow Y\}$ where $Y \sim \mathbb{R}^M$. We can also say that we coordinatize the world with N real micro degrees of freedom, and observe it through M macro degrees of freedom. According to the above argumentation $N \gg M$, the number of microscopic degrees of freedom is much larger than the number of macroscopic degrees of freedom [3,4].

But this also means that there are coordinate transformations that do not influence any of the observables. Let us denote

More Information

***Address for correspondence:**

Antal Jakovac, Department of Computational Sciences, Institute for Particle and Nuclear Physics, HUN-REN Wigner Research Centre for Physics, 29-33 Konkoly-Thege Mikló's Street, H-1121 Budapest, Hungary, Email: jakovac.antal@wigner.hun-ren.hu

Submitted: August 12, 2025

Approved: August 21, 2025

Published: August 22, 2025

How to cite this article: Jakovac A. Renormalization Group in Physics and Beyond. Int J Phys Res Appl. 2025; 8(8): 259-262. Available from: <https://dx.doi.org/10.29328/journal.ijpra.1001132>

Copyright license: © 2025 Jakovac A. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Functional renormalization group; Relevant-irrelevant coordinates; Modeling in physics; Artificial intelligence



$$W_y = o^{-1}(y), \tag{1}$$

the inverse image of $y \in Y$. This leads to a foliation of W , a coordinate system that is (locally) $R \times I$, where the coordinates in R do change some of the observables, while the ones in I do not [5]. We will refer to the coordinates in R as relevant degrees of freedom, and those in I as irrelevant, in accordance with the renormalization group (RG) nomenclature [6-9].

To describe in words what these symbols mean let us take the following example. We have two real facts, i.e. two real numbers, say (x_1, x_2) . The world W is then identifiable with \mathbb{R}^2 . We choose the set of observations $O = \{o\}$ with a $o(x, y) = \sqrt{x_1^2 + x_2^2}$ single element. Here there is a transformation of this observable, namely a rotation that leaves this observable intact. Then we have a subspace in \mathbb{R}^2

$$W_r = o^{-1}(r) = \{(r \cos \phi, r \sin \phi) \mid \phi \in [0, 2\pi)\} \tag{2}$$

belonging to the same value of the observable: this is a circle in this particular case. Therefore instead of the original coordinates it is worth to change the relevant-irrelevant (R-I) basis:

$$r = \sqrt{x_1^2 + x_2^2}, \phi = \arctan\left(\frac{x_2}{x_1}\right). \tag{3}$$

Here the irrelevant coordinate does not influence the value of the observable, while the relevant coordinate does.

Actually, the problem of all modeling tasks is to find this coordinate system. Indeed, if we originally have observations depending on two coordinates, say x and y , i.e. $o(x, y)$, but it turns out that all components of o are insensitive to the value of $x - y$, then it is worth – in the spirit of Occam’s razor – to introduce new coordinates $u = x+y$ and $v = x-y$, and state that $o(u)$ is, in fact, a single-variable function.

Thus, we obtain a minimal set of relevant quantities, independent in the sense that every nontrivial variation of them modifies at least one observable. This minimal number is not necessarily equal to the number of observables we consider, just in opposite, usually the number of relevant coordinates is much smaller. According to the above argumentation, all the observables are functions of the relevant coordinates

$$o_\alpha = o_\alpha(x_r)_{r=1, \dots, |R|}. \quad (4)$$

We can also associate an entropy to the learning process, which is largest when the number of relevant coordinates is the smallest [10].

Note that only the number of the relevant quantities are fixed, and we can use any other form that contains the x_r coordinates in a bijective way. In particular we can use a designated set of observables to express the other

$$o_\alpha = o_\alpha(x_r)_{r=1, \dots, |R|} \Rightarrow o_{\alpha_j} = f_{\alpha_j}(o_{j \leq |R|}). \quad (5)$$

2. Modeling in physics

In physics modeling consists of two parts. First we use a basis $\{\phi_\alpha\}$ that can produce all observables of the given environment – these correspond to the microstates. In mechanics these are the point mass coordinates and velocities, in a field theory these are the elementary fields: electric fields, wave functions or quantum fields, respectively, depending on the actual system.

In statistical systems we build up the ensemble by weighting the basis element according to some probability distribution (or, in quantum systems, with complex weights). Deterministic systems can also be described in this way, but the distributions are sharp (Dirac deltas). So the model is equivalent to giving the statistical weight factors associated to the basis elements $P(\phi)$. Usually, we do not speak about the statistical weights themselves, but about its logarithm, which is the Hamiltonian [11,12].

$$H(\phi) = -\frac{1}{\beta} \ln P(\phi), \quad (6)$$

where β is a formal temperature, determining how precisely should the system remain in a given energy shell.

If we single out a set of observables $\{o_\alpha\}$ in this system, then, as it was argued above, there are equivalent states that do not influence the result. Therefore, there are numerous equivalence relations between the microstates

$$\phi' = X_i(\phi), \quad o_\alpha(\phi') = o_\alpha(\phi). \quad (7)$$

This means that there is an equivalence relation between Hamiltonians

$$H' = H \circ X_i \sim H \quad (8)$$

for all X_i transformations.

As was argued above, there are only $|R|$ coordinates that relevant from the point of view of defining the Hamiltonian. So we can choose $|R|$ variables $h_r(\phi)$ (actually, they do not need to be observables themselves), and treat the Hamiltonian as the sum of these variables

$$H = \sum_{r \in R} c_r h_r \quad (9)$$

$$r \in R$$

where the c_r coefficients are also called coupling constants.

In physics it is usual that the number of relevant terms, $|R|$ is small, if the observables are the long range (infrared, IR) correlation functions. In Ising model it is just three (temperature, magnetic field and the coupling constant), but even in the Standard Model it is just 21, which is a much smaller number than the number of observables. Actually, we are tempted to say that a “good” theory consists of a few relevant quantities.

3. Modeling and AI

The same reasoning applies outside of physics as well. In all systems that are defined by “microscopic” variables that are not important from a certain point of view we can define the R-I system.

What is considered “important” is, of course, not given a priori. It is defined by the environment and, in practice, usually determined by specifying samples that should be regarded as equivalent. For example, if we want to tell apart dog and cat images, all dog images are considered equivalent, and all cat images are considered equivalent. If we want to separate different dog breeds, then dog images must not be equivalent, but all images showing for example, German shepherds, are equivalent.

For any of the tasks we may define the R-I foliation and find the corresponding coordinate system. Moving along the relevant coordinates we change the observables, moving along the irrelevant, we remain in the same equivalence class.

In fact, all the AI methods aim to find this foliation [13-16], usually not a complete coordination, but finding the

most relevant directions. In Support Vector Machine (SVM) algorithms we seek that direction, where the projection on this direction separates the equivalence classes the most. The Principal Component Analysis tries to find those coordinates along which a dataset varies the most. If we think the dataset as examples of equivalent samples, then the highest eigenvalues in PCA are associated with directions of maximum variance; clarification is needed regarding whether these are considered relevant or irrelevant, while the lowest ones are the relevant coordinates.

Since this can be somewhat counterintuitive, let us have a closer look what PCA coordinates mean. PCA assumes that the data lie on a linear subspace (eventually having a finite width), and tries to find those directions that are aligned with this subspace. If we consider the data equivalent, then that directions that do not lead out from this subspace do not change the equivalence class. According to the above nomenclature these are called irrelevant coordinates. The directions that are perpendicular to this subspace change the equivalence class, and so they are relevant.

Another view is that the relevant directions v are the ones that give $v \cdot x = 0$ for all $x \in \text{data}$. Therefore, a relevant direction has zero eigenvalue. Otherwise, these are the laws, the rules that are obeyed by all data elements.

To give a concrete example about the meaning of relevant and irrelevant directions, consider human faces: faces of the same person are treated as equivalent. In this case, it is useful to adopt a coordinate system where the relevant directions take to another person, the Irrelevants, just change the facial expression of the same person (like the “smile vector” in VAE).

The deep neural network models designed for classification determine those coordinates, using a nonlinear transformation of the original degrees of freedom, where the classes are different the most. We can also think that different layers represent different relevant coordinates, which loosely correspond to a physical scale, in particular if pooling layers are put in between.

In autoencoders, we try to build in the hardware the relevant coordinates, and in the internal layer we reveal those coordinates that characterize the internal part of the data belonging to the given equivalence class.

We may note here that the number of relevant coordinates has no reason to be small, contrary to our assumption in physics systems. The idea of Boltzmann machine, that assumes it, fails in most cases. For example, in face recognition we need thousands of base points to identify the different facial expressions and tell apart different people. Actually, this is the biggest difference between the areas where physics (in general science) can be used, and where we should use intelligent systems: in science the number of relevant quantities is small, in intelligence-approachable systems it is large.

4. Renormalization

The data model, which, according to the above discussion, is equivalent to the determination of the R-I coordinate system, depends on the designation of the observables. We cannot expect to use the same relevant quantities for different observables. Sometimes yes, but usually not.

In physical renormalization group studies, we usually change only the scale where the observables are defined. If we use infinitely large lattice, and change the lattice spacing, we consider formally the same observables (n -point functions), but the meaning of the observables changes. Therefore, we can expect that for changing the lattice spacing a bit, we see a small change in the Hamiltonian. In this case we can speak of a “running coupling”, whose value depends on the actual scale k

$$h_r \rightarrow h_r(k). \quad (10)$$

It can happen, however, that coefficients that were small for a given scale may increase in value. This process can lead to a violent change in the Hamiltonian (c.f. IR Landau poles [17]). After this change we cannot use the same Hamiltonian anymore, and we arrive at a completely different system using completely different relevant quantities. This crossover can be observed in the QCD \rightarrow nuclear physics change.

In AI systems small changes are rare. Even if we have a system that is capable to separate N classes, the inclusion of the $(N+1)^{\text{th}}$ class brings in a huge change in the R-I coordination. This is the problem of catastrophic forgetting [18,19], meaning that we must re-train the network after the changing of the observation objectives.

Using the flexibility in defining the relevant quantities, we can find that R-I coordination that works for a lot of situations, allowing that in some cases certain coordinates are relevant, in other cases the same are irrelevant. Actually language works in this way, where the relevant coordinates are the different words. A word is not necessarily relevant in a given situation, for example a stone does not have taste, but all (important) observations in the world can be expressed with the help of the words. This process that includes several environments and works out their common R-I system, goes beyond the usual renormalization approach, and it leads to generalization and extension [4].

Conclusion

We have argued in this short paper that the most important task of all modelling, whether it is a scientific physics model or an artificial intelligence model, is to find the R-I (relevant-irrelevant) foliation for a given set of observables o_r , minimizing the number of relevant coordinates. All observables remain unchanged when we change any of the irrelevant coordinates, and there is no such a combination of the relevant coordinates that is irrelevant.

This coordinate system can be manifested in different ways. In physics, or in general in statistical systems we try to find the logarithm of the statistical weight, called Hamiltonian. The presence of irrelevant directions means that there is an equivalence relation between the Hamiltonians, leading to the same expected values of the observables. This makes it possible to represent the Hamiltonian as a weighted sum of certain terms.

In AI the R-I system usually shows up as the architecture of the applied network. In PCA the eigenvectors corresponding to the large eigenvalues are irrelevant in the above nomenclature, while the small eigenvalues remain the same in a given data set, so they play the role of relevant coordinates. In Deep Neural Networks the architecture provides the relevant coordinates, and in the last layer we build up those observables that can tell apart the different classes the most effectively.

Since the coordinate system depends on the singled-out observables, the data that require modelling, it changes whenever we change the task. In physics we consider the change in the scale, and follow the evolution of the Hamiltonian under this process. But this is only a small part of the possible changes, in fact all AI tasks represent a different observable set. In all cases, we expect that small change led to a small change in the R-I basis, in particular a small change in the Hamiltonian. Then we can speak about renormalization of the different terms, since the same terms remain, only their coefficients change. But even in physics it can happen that coordinates that are irrelevant in a given scale, become relevant in other cases. In AI problems it is a very common observation, leading to the catastrophic forgetting problem, and it requires an extra effort to use generalized relevant coordinates that can function across different situations.

Acknowledgment

This research was supported by the Ministry of Innovation and Technology, HUNREN Office, within the framework of the MI-LAB Artificial Intelligence National Laboratory Program.

References

1. <https://doi.org/10.1103/PhysicsPhysiqueFizika.2.263>
2. Shalizi CR, Moore C. What is a Macrostate? Subjective Observations and Objective Dynamics. arXiv preprint. 2003. Available from: <https://doi.org/10.1007/s10701-024-00814-1>
3. Kurbucz MT, Pósfay P, Jakovác A. Facilitating time series classification by linear law-based feature space transformation. *Sci Rep.* 2022;12(1):18026. Available from: <https://doi.org/10.1038/s41598-022-22829-2>
4. Jakovác A, Telcs A. Representation and abstraction. *Mathematics.* 2025;13(10):1666. Available from: <https://doi.org/10.3390/math13101666>
5. Jakovác A, Telcs A. A note on representational understanding. *Entropy.* 2022;24(9):1313. Available from: <https://doi.org/10.3390/e24091313>
6. Wilson KG. The renormalization group: Critical phenomena and the Kondo problem. *Rev Mod Phys.* 1975;47:773–840. Available from: <https://doi.org/10.1103/RevModPhys.47.773>
7. Wilson KG, Kogut J. The renormalization group and the ϵ expansion. *Phys Rep.* 1974;12(2):75–199. Available from: [https://doi.org/10.1016/0370-1573\(74\)90023-4](https://doi.org/10.1016/0370-1573(74)90023-4)
8. Cardy J. *Scaling and Renormalization in Statistical Physics.* Cambridge: Cambridge University Press; 1996. Available from: <https://doi.org/10.1017/CBO9781316036440>
9. Goldenfeld N. *Lectures on Phase Transitions and the Renormalization Group.* Boca Raton: CRC Press; 1992. Available from: <https://doi.org/10.1201/9780429493492>
10. Biró TS, Jakovác A. Entropy of Artificial Intelligence. *Universe.* 2022;8(1):53. Available from: <https://doi.org/10.3390/universe8010053>
11. Polchinski J. Renormalization and effective lagrangians. *Nucl Phys B.* 1984;231:269–95. Available from: [https://doi.org/10.1016/0550-3213\(84\)90287-6](https://doi.org/10.1016/0550-3213(84)90287-6)
12. Rosten OJ. Fundamentals of the exact renormalization group. *Phys Rep.* 2012;511(4):177–272. Available from: <https://doi.org/10.1016/j.physrep.2011.12.003>
13. Mehta P, Schwab DJ. An exact mapping between the variational renormalization group and deep learning. arXiv preprint. 2014. Available from: <https://doi.org/10.48550/arXiv.1410.3831>
14. Bény C. Deep learning and the renormalization group. arXiv preprint. 2013. Available from: <https://doi.org/10.48550/arXiv.1301.3124>
15. Lin HW, Tegmark M, Rolnick D. Why does deep and cheap learning work so well? *J Stat Phys.* 2017;168:1223–47. Available from: <https://doi.org/10.1007/s10955-017-1836-5>
16. Cammarota C. A renormalization group approach to data analysis. *Nat Commun.* 2020;11:1573. Available from: <https://doi.org/10.1038/s41467-020-15353-8>
17. Itzykson C, Zuber J-B. *Quantum Field Theory.* New York: McGraw-Hill; 1980. Available from: <https://archive.org/details/quantumfieldtheo0000itzy>
18. French RM. Catastrophic forgetting in connectionist networks. *Trends Cogn Sci.* 1999;3(4):128–35. Available from: [https://doi.org/10.1016/s1364-6613\(99\)01294-2](https://doi.org/10.1016/s1364-6613(99)01294-2)
19. Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, et al. Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci U S A.* 2017;114(13):3521–6. Available from: <https://doi.org/10.1073/pnas.1611835114>