Proceedings of the ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2021
August 17-August 20, 2021, Online, Virtual,

DETC2021-70505

A NEW TRIANGLE: FRACTIONAL CALCULUS, RENORMALIZATION GROUP, AND MACHINE LEARNING

Haoyu Niu
Department of Electrical Engineering and Computer Science
University of California
Merced, California 95340
Email: hniu2@ucmerced.edu

YangQuan Chen *
Department of Mechanical Engineering
University of California
Merced, California 95340
Email: ychen53@ucmerced.edu

Lihong Guo
Institute of Mathematics,
Jilin University,
Changchun, Jilin, 130012, China,
Email: guolh17@mails.jlu.edu.cn

Bruce J. West
retired
Email: brucejwest213@gmail.com

ABSTRACT

The emergence of the systematic study of complexity as a science has resulted from the growing recognition that the fundamental assumptions upon which Newtonian physics is based are not satisfied throughout most of science, e.g., time is not necessarily uniformly flowing in one direction, nor is space homogeneous. Herein we discuss how the fractional calculus (FC), renormalization group (RG) theory and machine learning (ML) have each developed independently in the study of distinct phenomena in which one or more of the underlying assumptions of Newtonian formalism is violated. FC has been shown to help us better understand complex systems, improve the processing of complex signals, enhance the control of complex networks, increase optimization performance, and even extend the enabling of the potential for creativity. RG allows one to investigate the changes of a dynamical system at different scales. For example, in quantum field theory, divergent parts of a calculation can lead to nonsensical infinite results. However, by applying RG, the divergent parts can be adsorbed into fewer measurable quantities, yielding finite results. To date, ML is a fashionable research topic and will probably remain so into the foreseeable future. How a model can learn efficiently (optimally) is always essential. The key to learnability is designing efficient optimization methods. Although extensive research has been carried out on the three topics separately, few studies have investigated the association triangle between the FC, RG, and ML. To initiate the study of their interdependence, herein the authors discuss the critical connections between them (Fig. 1). In the FC and RG, scaling laws reveal the complexity of the phenomena discussed. The authors emphasize that the FC's and RG's critical connection is the form of inverse power laws (IPL), and the IPL index provides a measure of the level of complexity. For FC and ML, the critical connections in big data, wherein variability, optimization, and non-local models are described. The authors introduce the derivative-free and gradient-based optimization methods and explain how the FC could contribute to these study areas. In the end, the association between the RG and ML is also explained. The mutual information, feature extraction, and locality are also discussed. Many of the cross-sectional studies suggest a con-

*Address all correspondence to this author.

nection between the RG and ML. The RG has a superficial similarity to deep neural networks (DNNs) structure in which one marginalizes over hidden degrees of freedom. The authors remark in the conclusions that the association triangle between FC, RG, and ML, form a stool on which the foundation to complexity science might comfortably sit for a wide range of future research topics.

1. Fractional Calculus

Fractional calculus (FC) is the quantitative analysis of functions using non-integer-order integration and differentiation. The order can be a real number, a complex number, or even the function of a variable. The first recorded query regarding the meaning of a non-integer order differentiation appeared in a letter written in 1695 by Guillaume de l'Hôpital to Gottfried Wilhelm Leibniz. Leibniz was a contemporary of Sir Isaac Newton, and independently of him, co-invented the infinitesimal calculus [1]. Since then, many contributors have provided definitions for fractional derivatives and integrals [2], and the theory along with the applications of FC expanded greatly over the centuries [3, 4, 5].

In more recent decades, the concept of fractional dynamics has merged and gained followers in the statistical and chemical physics communities [6, 7, 8]. For example, optimal image processing has improved through the use of fractional-order differentiation and fractional-order partial differential equations as summarized in Chen et al. [9, 10, 11]. Anomalous diffusion was described using fractional diffusion equations in [12, 13] and Metzler et al. used fractional Langevin equations to model viscoelastic materials [14].

Fractional-order thinking (FOT) is a way of thinking using FC. For example, there are non-integers between the integers. Between logic 0 and logic 1, there is the fuzzy logic [15]. Compared with integer-order splines, there are fractional-order splines [16]. Between the high-order integer moments, there are non-integer-order moments. FOT has been entailed by many research areas, for example, self-similar [17, 18], scale-free or scale-invariant, power law, long-range dependence (LRD) [19, 20], and $1/f^{\alpha}$ noise [21, 22].

2. Renormalization Group

The **Renormalization group (RG)** is a conceptual framework which contains multiple techniques, such as real-space RG [23], functional RG [24], density matrix renormalization group (DMRG) [25]. The RG was originally devised in particle physics. Stueckelberg et al. [26] anticipated the idea in quantum field theory and first proposed the field conceptually. They noted that RG exhibited a group of transformations that transferred quantities from the bare terms to the counter terms. To date, the RG has extended to many other areas, such as solid-state physics, fluid mechanics, and physical cosmology. In theoretical physics, the RG refers to a formal apparatus, allowing one to investigate a physical system's changes at different scales. A change in scale is usually defined as a scale transformation.

The RG is positively related to self-similarity. In quantum field theory, divergent parts of a calculation can cause nonphysical infinite results. Using the RG, the divergent parts can be redefined into a few measurable quantities, yielding finite results. For instance, the quantum field theory is adopted to calculate the effects of fundamental forces at the quantum level. In quantum electrodynamics, the electron can emit and reabsorb "virtual" photons continuously, which means that its total energy and mass can be infinite. However, the divergence problem can be resolved by using RG theory, which means defining the mass of the "bare" electron to include these virtual processes and setting it equal to the measured mass. Inspired by the success of quantum electrodynamics, some other quantum field theories are also thriving, such as the electroweak theory and quantum chromodynamics. However, a renormalizable theory covering all the fundamental forces, such as gravity, is still under investigation.

Two major areas of investigation not mentioned above in which the non-inevitability of RG is explicit are the chaotic behavior in dynamical systems and the universality theory of critical points in statistical mechanics. Of more recent origin is the RG application to describe the phase transition of complex dynamic networks, where at the critical point a physical observable becomes discontinuous. Both social and biological complex dynamic networks have been shown to be members of the Ising universality class even though they are finite dimensional and therefore are not in the thermodynamic limit imposed for its use in the historical applications, see, e.g., West et al. [27] for a more extended discussion. A two-level model of cognition was developed using these ideas and provided insight into how linear logical thinking is disrupted by paradox only to be resolved using FOT [28].

3. Machine Learning

Machine Learning (ML) is the science (and art) of programming computers so they can learn from data [29]. A more engineering-oriented definition was given by Tom Mitchell in 1997 [30], "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E". In 2006, Hinton et al. [31] trained a DNN to recognize handwritten digits with an accuracy of more than 98%. Since then, researchers are more and more interested in Deep Learning (DL), and this enthusiasm extends to many areas of ML, such as image processing [32, 33], natural language processing [34], and even precision agriculture [35, 36, 37].

Why do we need ML? In summary, ML algorithms can usually simplify a solution and perform better than traditional methods, which may require much more hand-tuning rules. Further-

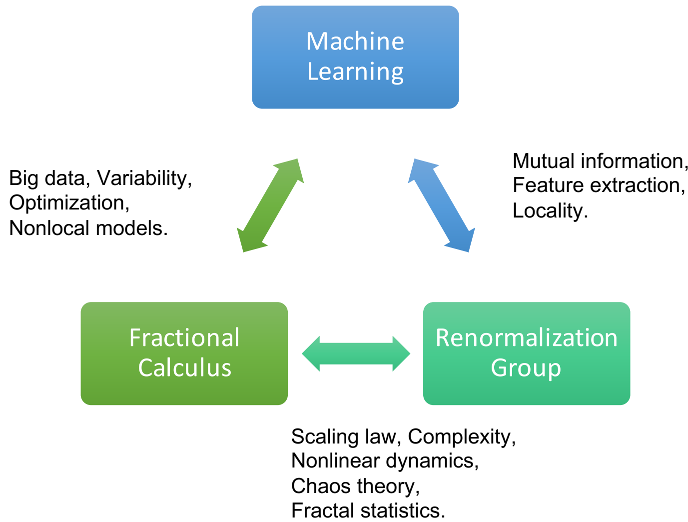

FIGURE 1: The new triangle between the FC, RG, and ML. For FC and RG, scaling laws and complexity are discussed in the text. The authors point out that FC's and RG's essential connection is the IPL. For FC and ML, the critical connections in big data, variability, optimization, and nonlocal models are described. The authors introduce the derivative-free and gradient-based optimization methods and explain how the FC contributes to these study areas. In the end, the association between the RG and ML is explained. The mutual information, feature extraction, and locality are subsequently discussed. Many of the cross-sectional studies suggest a connection between the RG and ML. The RG has a superficial similarity to DNNs structure in which one marginalizes over hidden degrees of freedom.

more, there may not exist a right solution for the complex phenomena by traditional methods. The ML techniques can help explain that kind of complexity and can adapt to new data better. The ML algorithms can obtain the variability about the complex problems and big data [38]. There are many different types and ways for ML algorithms classification (Fig. 2). ML can be classified as supervised, unsupervised, semi-supervised, and Reinforcement Learning (RL) based on whether human supervision is included. According to whether or not the ML algorithms can learn incrementally on the fly, they can be classified into online and batch learning. Based on whether or not the ML algorithms detect the training data patterns and create a predictive model, the ML can be classified into instance-based and model-based learning [29].

Most ML algorithms perform training by solving optimization problems that rely on the first-order derivatives (Jacobians), which decide whether to increase or decrease weights. For a huge speed boost, faster optimizers are being used instead of traditional Gradient Descent (GD) optimizers. For example, the most popular boosters are Momentum optimization [39], Nesterov Accelerated Gradient [40], AdaGrad [41], RMSProp [42], and Adam optimization [43]. The second-order (Hessian) optimization methods usually find the solutions with faster convergence rates but higher computational costs. Therefore, the answer to the following question is important: What is a more optimal ML algorithm? What if the derivative is fractional-order instead of integer order?

4. Fractional Calculus and Renormalization Group

In the study of phase transitions and critical phenomena, the RG can successfully characterize the self–similarity of systems near critical points [44]. Since FC is another tool to explain the self-similarity or complexity, suggest that there might be strong connections between the FC and RG. Guo et al. [45] discussed RG with scaling laws and complex phenomena to better understand their relationship, which covers nonlinear dynamics, chaos theory, fractal statistics, and FC from different disciplines.

The power law is usually described as:

$$f(x) = ax^k, \tag{1}$$

when k is a negative constant, $f(x)$ is an IPL. One important characteristic of this power law is scale invariance [46] determined by:

$$f(cx) = a(cx)^k = c^k f(x) \propto f(x). \tag{2}$$

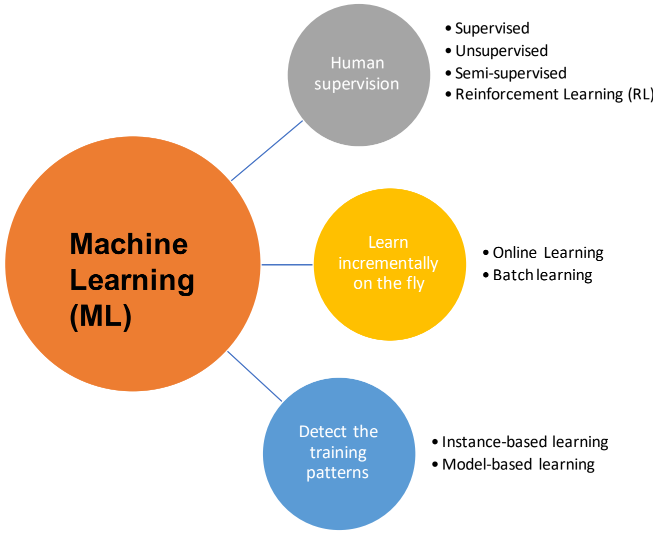

FIGURE 2: The ML can be classified as supervised, unsupervised, semi-supervised, and Reinforcement Learning (RL) based on whether or not human supervision is included. According to whether or not the ML algorithms can learn incrementally on the fly, they can be classified into online and batch learning. Based on whether or not the ML algorithms detect the training data patterns and create a predictive model, the ML can be classified into instance-based and model-based learning [29].

Note that when x is the time, the scaling depicts a property of the system dynamics. However, when the system is stochastic, the scaling is a property of the PDF (or correlation structure) and is a constraint on the collective properties of the system.

When complexity is under scrutiny, it is fair that we ask what it means? At what point do investigators begin identifying a system, network, or phenomenon as complex [47,48]? It seems that a clear and unified definition of complexity is still unknown to us. The joint use of FC and RG can potentially answer the following two questions [45]:

1. How can we characterize complexity?
2. What method should be used for the analysis of complexity in order to better understand real–world complex phenomena?

There is agreement among a significant fraction of the scientific community that when the distribution of the data associated with the process of interest is IPL, the phenomenon is complex. In the book by West and Grigolini [49], there is a table listing a sample of the empirical power laws and IPLs uncovered in the past two centuries. For example, in scale-free networks, the degree distributions follow an IPL in connectivity [50,51], in the processing of signals containing pink noise the power spectrum is IPL [20]. For other examples, such as the PDF, the auto-correlation function (ACF) [52], allometry ($Y = aX^b$) [53], anomalous relaxation (evolving over time) [54], anomalous diffusion (mean squared dissipation versus time) [13], self-similar, they can all be described by an IPL.

Douglas [55] investigated the surface–interacting polymers model and achieved the exact solution of the partition function using the FC. Furthermore, Douglas first explained the fundamental correlation between the RG scaling functions and the exact scaling functions. Qian [56] explained the correlation between the fractional Brownian motion (FBM) and the RG in statistical physics, and analyzed the statistical, geometric, and fractal properties of the complex phenomena. To prove that the connection between the FC and RG is IPL, Guo et al. [45] gave an example of the Weierstrass Random Walk (WRW) [57]. Then, the RG method was applied to the lattice structure function and determined the scaling properties of the WRW. Zaslavsky [58] used non-integrable Hamiltonians to study the chaotic particle kinetics. The fractional Fokker-Planck-Kolmogorov (FFPK) equation was derived to explain the ensemble behavior of the random walk in the fractal space–time system. The FFPK equation was solved by using RG methods.

5. Fractional Calculus and Machine Learning

As mentioned previously, ML is a much discussed research topic and will probably remain so for some time. How a machine can learn efficiently (optimally) is always important. The key for the learning process is the optimization method. Thus, designing an efficient optimization method is necessary to answer the following three questions,

1. What is the optimal way to optimize?
2. What is the **more optimal** way to optimize? Multiple optima?
3. Can we demand **"More Optimal Machine Learning"**, for example, DL with minimum/smallest labelled data?

During the learning process of ML, we care about both of the speed and the accuracy of the process. The algorithm is important, otherwise, data labelling and other labor can be costly. Therefore, the key to ML is the optimization methods being applied. The convergence rate and global searching are two important parts of the optimization method. Theoretically, there are two broad optimization categories, the derivative-free and gradient-based. For the derivative-free methods, there are the direct-search techniques, consisting of particle swarm optimization (PSO) [59, 60], etc. For the gradient-based methods, there are GD and its variants. Both of these categories have shown better performance when using the FC as demonstrated below.

For derivative-free methods, there are single agent search and swarm-based search methods. Exploration is often achieved by randomness or random numbers in terms of some predefined PDFs. Exploitation uses local information such as gradients to search local regions more intensively, and such intensification can enhance the rate of convergence. Then, the question was posed: What is the optimal randomness? Wei et al. [61] investigated the optimal randomness in a swarm-based search. Four asymptotic heavy-tailed PDFs, the Mittag-Leffler, Pareto, Weibull and Cauchy distributions, have been used for sample paths analysis. Based on the experimental results, the randomness-enhanced cuckoo search (CS) algorithms [62, 63, 64] can identify the unknown specific parameters of the fractional-order system with better effectiveness and robustness. The randomness-enhanced CS algorithms can be considered as a promising tool for solving the real-world complex optimization problems. The reason is that optimal randomness is being applied with fractional-order noise during the exploration, which is more optimal than, the "optimized PSO", CS. The fractional-order noise refers to the stable PDFs [65]. In other words, when we are discussing optimal randomness, we are discussing the FC [66]!

The GD is a very common optimization algorithm, which can find the optimal solutions by iteratively tweaking parameters to minimize the cost function. The stochastic gradient descent (SGD) random selects times during the training process. Therefore, the cost function will bounce up and down, decreasing on average, which is good for escape from local optima. Sometimes noise is added into the GD method and usually such noise follows Gaussian PDF in the literature. We ask "Why not heavy-tailed PDFs"? The answer to this question can lead to interesting future research.

The key to developing an efficient learning process is the method of optimization. Thus, it is important to design an efficient optimization method. The derivative-free methods, as well as the gradient-based methods, such as the Nesterov accelerated gradient descent (NAGD) can contribute to this study area. For NAGD, a main idea of Jordan's work [67] is to analyze the iteration algorithm in the continuous-time domain. For differential equations, one can use the Laypunov or variational method to analyze its properties, for example the convergence rate is $O(\frac{1}{t^2})$. One can also use the variational method to derive the master differential equation for an optimization method, such as the Least Action Principle [68], Hamilton's Variational Principle [69], and the Quantum-Mechanical Path Integral Approach [70].

Jordan's work revealed that one can transform an iterative (optimization) algorithm to its continuous-time limit case, which can simplify the analysis (Lyapunov methods). One can directly design a differential equation of motion (EOM) and then discretize it to derive an iterative algorithm (variational method). The key is to find a suitable Lyapunov functional to analyze the stability and convergent rate. The new exciting fact due to Jordan is that optimization algorithms can be systematically synthesized using Lagrangian mechanics (Euler-Lagrange), EOM.

Inspired by M. Jordan's idea in frequency domain, a continuous time fractional-order system was designed in [38]. Furthermore, it is shown to be possible, following the internal model principle (IMP), to design and analyze the ML algorithms in S or Z transform domain [38]. In general, M. Jordan asked the question: "Is there an optimal way to optimize"? Our answer is yes, by limiting dynamics analysis, discretization and SGD with other randomness, such as Langevin motion. Herein, the question posed was: "Is there a more optimal way to optimize"? Again the answer is yes, but it requires the FC to be used to optimize randomness in SGD, random search, and IMP. There are more potentials for further investigations along this line of ideas.

Nonlocal models have been commonly used to describe physical systems and/or processes which can not be accurately described by classical approaches [71]. For example, fractional nonlocal Maxwell's equations and the corresponding fractional wave equations have been applied in [72] for fractional vector calculus [73]. The nonlocal differential operators [74], including nonlocal analogs of the gradient/Hessian, are the key of these nonlocal models, which can be very interesting research with FC in the near future.

6. Renormalization Group and Machine Learning

The objective of ML research is to learn and extract significant features from the training data. As a sub-discipline of ML, DL uses multiple layers of representation to learn features directly from training data, which has been successfully utilized in many research topics, such as precision agriculture [75, 76, 77], and object detection [78, 79, 80]. Despite the success of DL algorithms, there remains a paucity of evidence on why ML techniques perform well on feature learning.

Recently, a number of cross-sectional studies suggest an association between the RG and ML [81, 82, 83]. Jefferson pointed out that the connection between RG and ML was originally made in the context of certain lattice models [84]. The decimation RG has a superficial similarity to the structure of DNNs in which one marginalizes over hidden degrees of freedom. For example, Mehta et al. proposed that the ML algorithms may be employing a generalized RG-like scheme to learn features from data [82]. In DNN, the low-level features were fed into higher layers with more abstract higher-level features. During the feature extraction process, the DNNs learn to ignore the irrelevant features while keeping the relevant ones. This continuous coarse-graining procedure is similar to the RG, which extracts relevant features of a physical system to describe phenomena at large length scales by integrating short-distance degrees of freedom. To illustrate their ideas, they constructed a mapping from the variational RG and DL architectures based on the Restricted Boltzmann Machines.

Lin et al. [83] explored how physics properties, such as symmetry, locality, compositionality, and polynomial log-probability, can translate into neural networks. They argued that when the statistical process belongs to some hierarchical form common in physics and ML, a DNN can perform better than a shallow one. For instance, one of the principles of physics is locality, which means that things only directly affect their immediate vicinity. When the locality is considered in ML applications, an arbitrary transformation of a collection of local random variables will result in a non-local collection [83]. However, the locality may be preserved. For example, spins are grouped into blocks in the simple block-spin RG, which are then treated as random variables. According to a high degree of accuracy, these blocks are only coupled to their nearest neighbors. This kind of locality is commonly exploited by biological and artificial visual systems. However, the lower layers perform fairly local operations.

In physical systems, the universal properties can determine the physical characteristics at most time, which can be revealed by the RG procedure. However, the critical degrees of freedom may not be easy to figure out. Therefore, Koch-Janusz et al. proposed an artificial neural network based on a model-independent, information-theoretic characterization of a real-space RG procedure [85]. Based on the ML algorithms, the model can identify the physically relevant degrees of freedom in a spatial region and performing an RG coarse-graining step iteratively. The input data follows a Boltzmann distribution, and no further knowledge about the microscopic details of the system is provided. The parameters of the NNs are optimized by a training algorithm based on evaluating real-space mutual information (RSMI) between spatially separated regions. The Ising and dimer models of classical statistical physics in two dimensions were used for the method validation.

7. The New FC-RG-ML Triangle and Emerging Opportunities

In general, the authors discussed the connections between the FC, RG, and ML in this article. It is shown that there exists a new triangle relationship between FC, RG, and ML. In this section, views on RG and physics informed ML with FC for future research opportunities are presented.

1. Generalization
   The generalization can be enhanced using the FC, RG, and ML. The ML models have great potential to be enhanced, referring to the generalization part being connected to FC and RG methods. For example, Mehta et al. proposed that the ML algorithms may be employing a generalized RG-like scheme to learn features from data [82]. The whole complex system usually has many small components that can interact with their nearby components and the environment, making their behavior challenging to predict. As a result, the ensemble PDF dynamics cannot be described by the standard partial differential equations in phase space. The PDF equations have been determined to be fractional generalizations of the traditional phase-space equations [45].

2. Training dataset variability and diversity
   The variability is the most critical characteristic being discussed during the training of ML algorithms. Variability can refer to several properties of the training dataset. First, the number of inconsistencies in the data needs to be understood using anomaly and outlier detection methods. Second, variability can also refer to diversity [86, 87], resulting from disparate data types and sources, for example, healthy or unhealthy [88, 89]. As mentioned for RG methods, in quantum field theory, divergent parts of a calculation can lead to nonsensical infinite research results. By applying RG, the divergent parts can be adsorbed into fewer measurable quantities, yielding finite results. RG and FC techniques can enhance the training data diversity for ML algorithms. In turn, we could expect "smaller data" rather than "big data" for ML under the same performance requirement.

3. Accelerated learning
   In DNNs, the low-level features were fed into higher layers with more abstract higher-level features. During the feature extraction process, the DNNs can learn to ignore the irrelevant features while keeping the relevant ones, which will make the learning process faster and efficient. As mentioned previously, this procedure is similar to the RG, which extracts relevant features of a physical system to describe phenomena at large length scales by integrating short-distance degrees of freedom. To illustrate this idea, Mehta et al. [82] constructed a mapping from the varia-

tional RG, and DL architectures based on the Restricted Boltzmann Machines. The ML algorithms can be shown to benefit from the RG and FC.

4. Optimal randomness
Randomness plays a significant role in both exploration and exploitation. A good NNs architecture with randomly assigned weights can easily outperform a more deficient architecture with finely tuned weights [90]. Therefore, it is critical to discuss the optimal randomness for ML algorithms by using RG and FC methods. For FC, there are rich forms in stochasticity [91], for example, heavytailedness, which corresponds to fractional-order master equations [92]. In probability theory, heavy-tailed distributions are PDFs whose tails do not decay exponentially [93]. Consequently, they have more weight in their tails than does an exponential distribution. Heavy-tailed distributions are widely used for modeling in different disciplines, such as finance [94], insurance [95], and medicine [96].
In [83], Lin et al. explored how physics properties can translate into neural networks (NNs) with RG methods. In [85], Koch-Janusz et al. proposed an artificial neural network based on a model-independent, information-theoretic characterization of a real-space RG procedure. Based on the ML algorithms, the model could identify the physically relevant degrees of freedom in a spatial region and performing an RG coarse-graining step iteratively [85]. These examples, we believe, can open a new horizon to the optimal randomness study using the FC, RG, and ML together.

8. Concluding Remarks

FC, RG, and ML are three of the most productive research areas and will probably remain so into the near future. Extensive research has been carried out on the three topics separately. However, few studies have investigated the association triangle between the FC, RG, and ML. Thus, the authors have explained the critical connections among the three points of the triangle in this article. When dealing with complex systems, the connection between FC and RG is revealed and explained. For FC and RG, scaling laws and complexity are discussed and here we note the scaled form of the PDF that solves the fractional diffusion equation [43]. The authors point out that the key connection between FC and RG is the IPL, which is essential to studying complex systems.

For FC and ML, the critical connections in big data, variability, optimization, and nonlocal models are described. The authors introduced the derivative-free and gradient-based optimization methods and explained how the FC could contribute to these study areas. Fractional dynamics responds to the more advanced characterization of our more complex world to capture structures at a too small or too large scale that had previously been smoothed over. Suppose one wishes to obtain better results than the best possible using integer order calculus-based methods, or to be "more optimal", we advocate to apply FOT (fractional order thinking). In the era of big data, decision and control need FC, such as fractional-order signals, systems, and controls.

In the end, the association between the RG and ML was explained. The mutual information, feature extraction, and locality were discussed. Many of the cross-sectional studies suggested a connection between the RG and ML. The RG has a similarity to DNNs structure in which one marginalizes over hidden degrees of freedom. The future of ML should be physics-informed, scientific (cause-effect embedded or cause-effect discovery), and involving the use of FC and RG, where the modeling is closer to the true nature.

REFERENCES

[1] Vinagre, B. M., and Chen, Y., 2002. "Lecture notes on fractional calculus applications in automatic control and robotics". In Proceedings of the 41st IEEE CDC Tutorial Workshop, Vol. 2, pp. 1–310.

[2] Valério, D., Machado, J., and Kiryakova, V., 2014. "Some pioneers of the applications of fractional calculus". *Fractional Calculus and Applied Analysis,* ***17***(2), pp. 552–578.

[3] Abel, N., 1823. "Solution of a couple of problems by means of definite integrals". *Magazin for Naturvidenskaberne,* ***2***(55), p. 2.

[4] Podlubny, I., Magin, R. L., and Trymorush, I., 2017. "Niels Henrik Abel and the birth of fractional calculus". *Fractional Calculus and Applied Analysis,* ***20***(5), pp. 1068–1075.

[5] Ross, B., 1977. "The development of fractional calculus 1695–1900". *Historia Mathematica,* ***4***(1), pp. 75–89.

[6] Tarasov, V. E., 2011. *Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Fields and Media*. Springer Science & Business Media.

[7] Klafter, J., Lim, S., and Metzler, R., 2012. *Fractional Dynamics: Recent Advances*. World Scientific.

[8] Pramukkul, P., Svenkeson, A., Grigolini, P., Bologna, M., and West, B., 2013. "Complexity and the fractional calculus". *Advances in Mathematical Physics,* ***2013***, pp. 1–7.

[9] Chen, D., Xue, D., and Chen, Y., 2013. "More optimal image processing by fractional order differentiation and fractional order partial differential equations". In Proceedings of the International Symposium on Fractional PDEs.

[10] Chen, D., Sun, S., Zhang, C., Chen, Y., and Xue, D., 2013. "Fractional-order TV-L 2 model for image denoising". *Central European Journal of Physics,* ***11***(10), pp. 1414–1422.

[11] Yang, Q., Chen, D., Zhao, T., and Chen, Y., 2016. "Fractional calculus in image processing: A review". *Fractional Calculus and Applied Analysis,* ***19***(5), pp. 1222–1249.

[12] Seshadri, V., and West, B. J., 1982. "Fractal dimensionality of Lévy processes". *In Proceedings of the National Academy of Sciences of the United States of America,* ***79***(14), p. 4501.

[13] Metzler, R., and Klafter, J., 2000. "The random walk's guide to anomalous diffusion: A fractional dynamics approach". *Physics Reports,* ***339***(1), pp. 1–77.

[14] Metzler, R., Glöckle, W. G., and Nonnenmacher, T. F., 1994. "Fractional model equation for anomalous diffusion". *Physica A: Statistical Mechanics and its Applications,* ***211***(1), pp. 13–24.

[15] Zadeh, L. A., 1988. "Fuzzy logic". *Computer,* ***21***(4), pp. 83–93.

[16] Unser, M., and Blu, T., 2000. "Fractional splines and wavelets". *SIAM Review,* ***42***(1), pp. 43–67.

[17] Samoradnitsky, G., 2017. *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Routledge.

[18] Crovella, M. E., and Bestavros, A., 1997. "Self-similarity in World Wide Web traffic: Evidence and possible causes". *IEEE/ACM Transactions on Networking,* ***5***(6), pp. 835–846.

[19] Burnecki, K., and Weron, A., 2004. "Lévy stable processes. From stationary to self-similar dynamics and back. An application to finance". *Acta Physica Polonica Series B,* ***35***(4), pp. 1343–1358.

[20] Pesquet-Popescu, B., and Pesquet, J. C., 2002. "Synthesis of bidimensional α-stable models with long-range dependence". *Signal Processing,* ***82***(12), pp. 1927–1940.

[21] Hartley, T. T., and Lorenzo, C. F., 2003. "Fractional-order system identification based on continuous order-distributions". *Signal Processing,* ***83***(11), pp. 2287–2300.

[22] Wolpert, R. L., and Taqqu, M. S., 2005. "Fractional Ornstein–Uhlenbeck Lévy processes and the telecom process: Upstairs and downstairs". *Signal Processing,* ***85***(8), pp. 1523–1545.

[23] Kadanoff, L. P., 1966. "Scaling laws for Ising models near Tc". *Physics Physique Fizika,* ***2***(6), pp. 263–272.

[24] Wetterich, C., 1993. "Exact evolution equation for the effective potential". *Physics Letters B,* ***301***(1), pp. 90–94.

[25] White, S. R., 1992. "Density matrix formulation for quantum renormalization groups". *Physical Review Letters,* ***69***(19), pp. 2863–2866.

[26] Stuckelberg, E., 1953. "La normalisation des constantes dans la theorie des quanta". *Helvetica Physica Acta,* ***26***, pp. 499–520.

[27] West, B. J., Turalska, M., and Grigolini, P., 2014. "Complex networks: From social crises to neuronal avalanches". *Hoboken, NJ: Wiley-VCH Verlag GmbH*, pp. 509–524.

[28] West, B. J., 2019. "Paradox entails new kinds of medical knowledge". *Medical Research Archives,* ***7***(8), pp. 1–16.

[29] Géron, A., 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.

[30] Mitchell, T. M., 1997. *Machine Learning*. McGraw-hill New York.

[31] Hinton, G. E., Osindero, S., and Teh, Y. W., 2006. "A fast learning algorithm for deep belief nets". *Neural Computation,* ***18***(7), pp. 1527–1554.

[32] Zhao, T., Yang, Y., Niu, H., Wang, D., and Chen, Y., 2018. "Comparing U-net convolutional network with mask R-CNN in the performances of pomegranate tree canopy segmentation". In Proceedings of the Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII, International Society for Optics and Photonics.

[33] Zhao, T., Koumis, A., Niu, H., Wang, D., and Chen, Y., 2018. "Onion irrigation treatment inference using a low-cost hyperspectral scanner". In Proceedings of the Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII, International Society for Optics and Photonics.

[34] Hirschberg, J., and Manning, C. D., 2015. "Advances in natural language processing". *Science,* ***349***(6245), pp. 261–266.

[35] Zhao, T., Niu, H., de la Rosa, E., Doll, D., Wang, D., and Chen, Y., 2018. "Tree canopy differentiation using instance-aware semantic segmentation". In Proceedings of the 2018 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers.

[36] Niu, H., Zhao, T., Wang, D., and Chen, Y., 2019. "A UAV resolution and waveband aware path planning for onion irrigation treatments inference". In Proceedings of the 2019 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE, pp. 808–812.

[37] Niu, H., Wang, D., and Chen, Y., 2020. "Estimating crop coefficients using linear and deep stochastic configuration networks models and UAV-based normalized difference vegetation index (NDVI)". In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), IEEE.

[38] Niu, H., Chen, Y., and West, B. J., 2021. "Why do big data and machine learning entail the fractional dynamics?". *Entropy,* ***23***(3), p. 297.

[39] Polyak, B. T., 1964. "Some methods of speeding up the convergence of iteration methods". *USSR Computational Mathematics and Mathematical Physics,* ***4***(5), pp. 1–17.

[40] Nesterov, Y., 1983. "A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$". *Doklady Akademii Nauk, Russian Academy of Sciences,* ***269***, pp. 543–547.

[41] Duchi, J., Hazan, E., and Singer, Y., 2011. "Adaptive subgradient methods for online learning and stochastic optimization.". *Journal of Machine Learning Research,* ***12***(7),

pp. 2121–2159.
[42] Hinton, G., and Tieleman, T. Slide 29 in lecture 6, 2012, rmsprop: "Divide the gradient by a running average of its recent magnitude".
[43] Kingma, D. P., and Ba, J., 2014. "Adam: A method for stochastic optimization". *arXiv preprint arXiv:1412.6980*.
[44] Wilson, K. G., 1983. "The renormalization group and critical phenomena". *Reviews of Modern Physics,* ***55***(3), pp. 583–600.
[45] Guo, L., Chen, Y., Shi, S., and West, B. J., 2021. "Renormalization group and fractional calculus methods in a complex world: A review". *Fractional Calculus and Applied Analysis,* ***24***(1), pp. 5–53.
[46] Kello, C. T., Brown, G. D., Ferrer Cancho, R., Holden, J. G., Linkenkaer Hansen, K., Rhodes, T., and Van Orden, G. C., 2010. "Scaling laws in cognitive sciences". *Trends in Cognitive Sciences,* ***14***(5), pp. 223–232.
[47] West, B. J., Geneston, E. L., and Grigolini, P., 2008. "Maximizing information exchange between complex networks". *Physics Reports,* ***468***(1-3), pp. 1–99.
[48] West, B. J., 2020. "Sir Isaac Newton stranger in a strange land". *Entropy,* ***22***(11), p. 1204.
[49] West, B. J., and Grigolini, P., 2010. *Complex Webs: Anticipating the Improbable*. Cambridge University Press.
[50] Barabási, A. L., and Albert, R., 1999. "Emergence of scaling in random networks". *Science,* ***286***(5439), pp. 509–512.
[51] Sun, W., Li, Y., Li, C., and Chen, Y., 2011. "Convergence speed of a fractional order consensus algorithm over undirected scale-free networks". *Asian Journal of Control,* ***13***(6), pp. 936–946.
[52] Li, M., 2007. "Modeling autocorrelation functions of long-range dependent teletraffic series based on optimal approximation in Hilbert space—A further study". *Applied Mathematical Modelling,* ***31***(3), pp. 625–631.
[53] Zhao, Z., Guo, Q., and Li, C., 2008. "A fractional model for the allometric scaling laws". *The Open Applied Mathematics Journal,* ***2***(1), pp. 26–30.
[54] Sun, H., Chen, Y., and Chen, W., 2011. "Random-order fractional differential equation models". *Signal Processing,* ***91***(3), pp. 525–530.
[55] Douglas, J. F., 1989. "Surface-interacting polymers: An integral-equation and fractional-calculus approach". *Macromolecules,* ***22***(4), pp. 1786–1797.
[56] Qian, H., 2003. "Fractional Brownian motion and fractional Gaussian noise". In *Processes with Long-Range Correlations*. Springer, pp. 22–33.
[57] Hughes, B., Montroll, E., and Shlesinger, M., 1982. "Fractal random walks". *Journal of Statistical Physics,* ***28***(1), pp. 111–126.
[58] Zaslavsky, G., 1994. "Renormalization group theory of anomalous transport in systems with Hamiltonian chaos". *Chaos: An Interdisciplinary Journal of Nonlinear Science,* ***4***(1), pp. 25–33.
[59] Wei, J., Yu, Y., and Wang, S., 2015. "Parameter estimation for noisy chaotic systems based on an improved particle swarm optimization algorithm". *The Journal of Applied Analysis and Computation,* ***5***(2), pp. 232–242.
[60] Wei, J., Yu, Y., and Cai, D., 2018. "Identification of uncertain incommensurate fractional-order chaotic systems using an improved quantum-behaved particle swarm optimization algorithm". *Journal of Computational and Nonlinear Dynamics,* ***13***(5).
[61] Wei, J., Chen, Y., Yu, Y., and Chen, Y., 2019. "Optimal randomness in swarm-based search". *Mathematics,* ***7***(9), p. 828.
[62] Wei, J., and Yu, Y., 2019. "A novel cuckoo search algorithm under adaptive parameter control for global numerical optimization". *Soft Computing*, pp. 1–24.
[63] Wei, J., Chen, Y., Yu, Y., and Chen, Y., 2019. "Improving cuckoo search algorithm with Mittag-Leffler distribution". In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 1–9.
[64] Wei, J., and Yu, Y., 2018. "An adaptive cuckoo search algorithm with optional external archive for global numerical optimization". In Proceedings of the International Conference on Fractional Differentiation and its Applications (ICFDA).
[65] Gorenflo, R., and Mainardi, F., 1998. "Fractional calculus and stable probability distributions". *Archives of Mechanics,* ***50***(3), pp. 377–388.
[66] Niu, H., Wei, J., and Chen, Y., 2021. "Optimal randomness for stochastic configuration network (SCN) with heavy-tailed distributions". *Entropy,* ***23***(1), p. 56.
[67] Wilson, A. C., Recht, B., and Jordan, M. I., 2016. "A Lyapunov analysis of momentum methods in optimization". *arXiv preprint arXiv:1611.02635*.
[68] Feynman, R. P., 2005. "The principle of least action in quantum mechanics". In *Feynman's Thesis—A New Approach to Quantum Theory*. World Scientific, pp. 1–69.
[69] Hamilton, S. W. R., 1834. *On a General Method in Dynamics*. Richard Taylor.
[70] Hawking, S. W., 1979. "The path-integral approach to quantum gravity". In *General Relativity: An Einstein centenary survey*. pp. 746–789.
[71] Nagaraj, S., 2020. "Optimization and learning with nonlocal calculus". *arXiv preprint arXiv:2012.07013*.
[72] Tarasov, V. E., 2008. "Fractional vector calculus and fractional Maxwell's equations". *Annals of Physics,* ***323***(11), pp. 2756–2778.
[73] Ortigueira, M., and Machado, J., 2018. "On fractional vectorial calculus". *Bulletin of the Polish Academy of Sciences:*

Technical Sciences, ***66***(4), pp. 389–402.

[74] Feliu Faba, J., Fan, Y., and Ying, L., 2020. "Meta-learning pseudo-differential operators with deep neural networks". *Journal of Computational Physics,* ***408***, p. 109309.

[75] Niu, H., Zhao, T., Westphal, A., and Chen, Y., 2020. "A low-cost proximate sensing method for early detection of nematodes in walnut using Walabot and scikit-learn classification algorithms". In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping V, International Society for Optics and Photonics.

[76] Niu, H., Zhao, T., Zuber, C., Vasquez-Mendoza, J., Doll, D., Arnold, K., and Chen, Y., 2020. "A low-cost stem water potential monitoring method using proximate sensor and scikit-learn classification algorithms". In Proceedings of the ASABE Annual International Virtual Meeting, American Society of Agricultural and Biological Engineers, pp. 1–6.

[77] Niu, H., Hollenbeck, D., Zhao, T., Wang, D., and Chen, Y., 2020. "Evapotranspiration estimation with small UAVs in precision agriculture". *Sensors,* ***20***(22), p. 6427.

[78] Niu, H., Zhao, T., and Chen, Y., 2019. "Intelligent bugs mapping and wiping (iBMW): An affordable robot-driven robot for farmers". In Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, pp. 397–402.

[79] Wang, Y., Niu, H., Zhao, T., Liao, X., Dong, L., and Chen, Y., 2019. "Contactless Li-Ion battery voltage detection by using Walabot and machine learning". In Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

[80] Wang, Y., Chen, Y., Liao, X., and Dong, L., 2019. "Lithium-ion battery face imaging with contactless Walabot and machine learning". In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, pp. 1067–1072.

[81] Bény, C., 2013. "Deep learning and the renormalization group". *arXiv preprint arXiv:1301.3124*.

[82] Mehta, P., and Schwab, D. J., 2014. "An exact mapping between the variational renormalization group and deep learning". *arXiv preprint arXiv:1410.3831*.

[83] Lin, H. W., Tegmark, M., and Rolnick, D., 2017. "Why does deep and cheap learning work so well?". *Journal of Statistical Physics,* ***168***(6), pp. 1223–1247.

[84] Jefferson, R., 2019. Deep learning and the renormalization group. Available online at the website of: https://rojefferson.blog/2019/08/04/deep-learning-and-the-renormalization-group/. (Accessed: 4th August 2019).

[85] Koch Janusz, M., and Ringel, Z., 2018. "Mutual information, neural networks and the renormalization group". *Nature Physics,* ***14***(6), pp. 578–582.

[86] Nakahira, Y., Liu, Q., Sejnowski, T. J., and Doyle, J. C., 2019. "Diversity-enabled sweet spots in layered architectures and speed-accuracy trade-offs in sensorimotor control". *arXiv preprint arXiv:1909.08601*.

[87] Arabas, J., and Opara, K., 2019. "Population diversity of non-elitist evolutionary algorithms in the exploration phase". *IEEE Transactions on Evolutionary Computation,* ***24***(6), pp. 1050–1062.

[88] Ko, M., Stark, B., Barbadillo, M., and Chen, Y., 2015. "An evaluation of three approaches using Hurst estimation to differentiate between normal and abnormal HRV". In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers.

[89] Li, N., Cruz, J., Chien, C. S., Sojoudi, S., Recht, B., Stone, D., Csete, M., Bahmiller, D., and Doyle, J. C., 2014. "Robust efficiency and actuator saturation explain healthy heart rate control and variability". *Proceedings of the National Academy of Sciences,* ***111***(33), pp. E3476–E3485.

[90] Scardapane, S., and Wang, D., 2017. "Randomness in neural networks: An overview". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* ***Volume 7***(2).

[91] Montroll, E. W., and West, B. J., 1979. "On an enriched collection of stochastic processes". *Fluctuation Phenomena,* ***66***, p. 61.

[92] Li, Z., Liu, L., Dehghan, S., Chen, Y., and Xue, D., 2017. "A review and evaluation of numerical tools for fractional calculus and fractional order controls". *International Journal of Control,* ***90***(6), pp. 1165–1181.

[93] Asmussen, S., 2003. "Steady-state properties of GI/G/1". *Applied Probability and Queues*, pp. 266–301.

[94] Bernardi, M., and Petrella, L., 2015. "Interconnected risk contributions: A heavy-tail approach to analyze US financial sectors". *Journal of Risk and Financial Management,* ***8***(2), pp. 198–226.

[95] Ahn, S., Kim, J. H., and Ramaswami, V., 2012. "A new class of models for heavy tailed distributions in finance and insurance risk". *Insurance: Mathematics and Economics,* ***51***(1), pp. 43–52.

[96] Resnick, S. I., 2007. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer Science & Business Media.