Renormalizing complex models: It is hard without Landau!

Parameter Space Compression Underlies Emergent Theories and Predictive Models. Benjamin B. Machta, Ricky Chachra, Mark K. Transtrum, and James P. Sethna. *Science* **342**, 604-607 (2013).

PCA meets RG. Serena Bradde and William Bialek. *arXiv*:1610.09733 (2016).

Recommended with a commentary by Ilya Nemenman, Emory University.

Renormalization group (RG) is one of the defining frameworks of theoretical physics of the second half of the 20th century. The goal of RG in statistical physics is to extract features of a theory that are *relevant* for describing phenomena at large length scales, in addition to later computing measurable quantities, such as critical exponents. For this, the method starts with the joint probability distribution describing microscopic degrees of freedom conditional on various parameter values (in traditional physical problems, this is typically the Boltzmann distribution with the Hamiltonian defined by the microscopic variables, and parameters of the Hamiltonian represent various coupling constants). Then some microscopic degrees of freedom are integrated out (marginalized). This changes parameters of the distribution, now defined over a reduced, coarse-grained set of degrees of freedom: some parameters get changed (*renormalized*), new couplings may emerge, etc. One repeats the procedure iteratively, progressing to larger and larger scales. The goal is then to understand which parameters or other features of the original joint probability distribution remain relevant for describing the coarse-grained theory, and which have the ever-diminishing effects on the observable properties of the system at large scales.

Crucially, succesfull application of RG is predicated on using various known symmetries of the problem to limit the set of possible interactions in the microscopic Hamiltonian. In turn, this also limits interactions emerging during the coarse-graining step. Here I use the word "symmetries" very broadly to denote all constraints on the involved coupling constants, including various homogenities (coupling constants at different points in time and space being equal to each other), localities (only self-couplings or near-neighbor couplings being nonzero), or absence of simultaneous many-body contacts (only low-order interactions being allowed), etc. Colloquially speaking, RG starts with Landau-style enumeration of symmetries, which reduce the number of distinct possible interactions among N degrees of freedom from $O(2^N)$ to $O(1)$ of those that satisfy the symmetries. One then uses Wilson's (or Kadanoff's, or any other) RG methodology to identify which even smaller subset of the interaction parameters or their combinations is relevant at large scales.

RG has had a dramatic success when applied to traditional physics problems in condensed matter and high energy physics. Not surprisingly, there have been recent elegant attempts to harness the power of RG for less well understood systems, such as those coming from the biological or the social sciences domain. Machta and colleagues in Ref. (1) implemented a method that is reminiscent of *real-space*, decimation approach to RG on the one hand, but also, surprisingly, connects to the field of statistical inference and machine learning, where *relevance* has a slightly different, but a related, meaning (see also (2) for a discussion of an exact correspondence between relevance in the machine learning and the variational RG context). In Machta et al. approach, one starts with many samples of configurations $\mathbf{x}$ of microscopic variables of the system. A model of the system is given by a probability distribution $P(\mathbf{x}|\vec{\theta})$, where $\vec{\theta}$ are the parameters. One can disregard small-scale features and focus on $P(\hat{\mathbf{x}}|\vec{\theta})$, where $\hat{\mathbf{x}}$ are coarse-grained subsamples of the original data. Parameters that are relevant for the large-scale description are the ones whose change significantly changes the distribution $P(\hat{\mathbf{x}}|\vec{\theta})$. Importantly, due to the Bayes theorem, $P(\hat{\mathbf{x}}|\vec{\theta}) \propto P(\vec{\theta}|\hat{\mathbf{x}})$, so that

large changes in the distribution of data given the parameters are equivalent to large changes in the distribution of the parameters given the observed data. In other words, relevant parameters are the ones that are *inferable* from the large scale, coarse-grained data. One can coarse-grain the system iteratively then and observe that some parameters remain inferable, while others become less so on large scales. And, indeed, when used to derive diffusion as a long-time limit of random walks, or to understand the relevant features of the Ising model, the inferable parameters were found in (1) to correspond to what we know to be the relevant interactions in the RG treatment.

The advantage of this approach is that it does not require a carefully constructed coarse-graining procedure and enumeration of new interactions that might be emerging in the coarse-grained model: the inference procedure replaces both. And yet it is unclear how the method could be used for systems where "symmetry" properties are not known *a priori*. For example, one needs to know the relative spatial arrangement (or, at least, some measure of closeness) of the variables x_i and x_j. This is not obvious when, for example, these microscopic variables represent neurophysiological multielectrode recordings from the brain or different mRNA expression levels in cells. Further, the approach assumes implicitly that dimensionality of $\vec{\theta}$ is not too large (inference of a large number of θ's is not practical) and that the small number of kept parameters in $P(\mathbf{x}|\vec{\theta})$ includes the ones that really matter (i. e., the model is correct and fits the data), and not much beyond that. All of this requires knowing the correct "symmetries" of the problem. In other words, the Machta et al. approach allows to replace the Wilson / Kadanoff RG with the inference step, and to go from a handful of parameters to the few that matter. However, it cannot do this without the (possibly much more important and harder) Landau step beforehand, which decreases the number of parameters from astronomically many to just the handful.

In a parallel approach, Bradde and Bialek (3) have developed a *momentum-space* version of renormalization of complex data. They avoid requiring the knowledge of the spatial relation among the variables, and instead work in the perturbative (small coupling) limit, where $P(\mathbf{x}|\vec{\theta})$ is nearly Gaussian. The usual kinetic energy and mass terms in the Hamiltonian are replaced with $1/2\mathbf{x}^T C^{-1}\mathbf{x}$, where C is the covariance matrix of the data. In the traditional momentum-space RG, one can estimate which interactions will be relevant on large scales based on the dimensional analysis, and the Bradde-Bialek formulation replaces this with the analysis of the spectrum of C instead. One then performs the coarse-graining step by integrating out eigenvectors (Principal Components) of $\mathbf{x}$ that correspond to small eigenvalues, and seeing how this affects the rest of the covariance matrix as well as the higher-order, perturbatively small interaction terms. As always, the iterative application of the coarse-graining step then allows to understand which combinations of the coupling constants become relevant at large scales and to what extent.

The approach was used to analyse various neural and financial datasets with some encouraging success, specifically because no assumptions about the spatial structure of the data were made: unlike in (1) and traditional applications of RG, here the scale is defined not by the spatial arrangement, but by the magnitude of a corresponding eigenvalue. This potentially requires a lot more data than (1), and also requires assuming that deviations from the Gaussianity are perturbative in nature. More importantly, the method still does not solve the "symmetry" problem. Indeed, the Bradde and Bialek model allows terms like $\lambda_i x_i^4$ in the Hamiltonian (and even then, to complete the renormalization step, requires an assumption that λ_is are weakly dependent on i). However, there is no reason why terms such as $\lambda_{ijkl}x_i x_j x_k x_l$ or even $\lambda_{ijk\ldots p}x_i x_j x_k \ldots x_p$ are impossible if locality is forgone. The number of such possible terms is astronomical, which makes it practically impossible to complete the renormalization step and calculate the renormalized values of all coupling constants. Even worse, with so many possible interaction terms, they can conspire to affect even the dimensional analysis calculations: a lot of small interactions may combine into large effects, so that the relevance of an interaction is not necessarily determined by the spectrum of the covariance

matrix alone. In other words, the Bradde and Bialek approach can calculate relevant parameters in the perturbative regime starting with some specific sets of $O(N)$ interactions (which is a huge step forward compared to methods that start with $O(1)$), but it still requires a Landau step to go from 2^N interactions to the *correct* set of $O(N)$.

In summary, we now have a variety of promising methods that have a chance to develop into replacements of some aspects of the Wilson / Kadanoff steps of the RG analysis in previously inaccessible contexts. This are important advances. Nonetheless, the harder problem of the Landau step remains elusive in complex data, such as that coming from biology, where symmetries of the data are unknown. Can we learn the appropriate exact or approximate, averaged symmetries of data from the data itself, and thus reduce the dimensionality of the problem and make various renormalization approaches practically applicable? Or can we build on strong-disorder RG methods (4) to embrace the heterogeneity of such complex data? This remains to be seen.

References

[1] B Machta, R Chachra, M Transtrum, and J Sethna. Parameter space compression underlies emergent theories and predictive models. *Science*, 342:604, 2013.

[2] P Mehta and D Schwab. An exact mapping between the variational renormalization group and deep learning. 2014. arXiv:1410.3831.

[3] S Bradde and W Bialek. PCA meets RG. 2016. arXiv:1610.09733.

[4] D Fisher. Critical behavior of random transverse-field Ising spin chains. *Phys Rev B*, 51:6411, 1995.