Information theory and (F)RG

Francesco Caravelli,
Theory Division and CNLS,
Los Alamos National Laboratory

September 16, 2019

In collaboration with **Andrea Trombettoni** (CNR/SISSA) and **Nicolo Defenu** (Univ. Heidelberg).

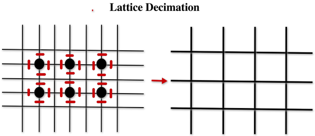
1. Exact Renormalization Group
   - Exact renormalization group

2. Information theory
   - General information theoretic framework

3. Conclusions and Outlook

The renormalization group

The renormalization group has been a central tool in condensed matter and high energy physics for over 50 years now.[1]



Basic idea is that an increasing correlation length near phase transition washes out degrees of freedom which are then incorporated in an effective Hamiltonian which keeps the free energy constant.
Also: **since the early days of Statistical Mechanics, Entropy and Information have been closely related concepts.**

[1] Kadanoff (60's)

From Kadanoff to Wilson

Consider a momentum dependent action. The coarse graining is introduced perturbatively by integrating out fast modes:

$$Z = \int D\phi e^{-S} \approx \int D\phi_< e^{-S_{\tilde{k}}[\phi_<]} \int D\phi_> e^{-S[\phi_>]-S'[\phi_<,\phi_>]} \qquad (1)$$

where $\phi_< \equiv \tilde{\phi}_{k\leq k^*}$ and $\phi_> \equiv \tilde{\phi}_{k>k^*}$ are the slow and fast modes respectively. Wilson introduced a momentum-shell coarse-graining procedure.

- Good: Phase diagrams, effective field theories and interesting phases *Outstanding success here, including BKT transition*
- OK: Critical Exponents to a very good level of approximation *Bootstrap is also extremely precise, based on CFT at the critical point* Addendum September 16th: Now also FRG seems to be doing very well, but need PMS
- "Distasteful": How do we choose coarse graining?

Intermezzo: Variational principles

Variational principles have a long history in the study of perturbation theory.

- Kadanoff (70s) suggested a variational principle in which the "optimal" coarse graining is one in which the free energy is minimized at the discrete level.
- Feynman and Kleinert (1986) introduced variational perturbation theory.

In the case of FK, write

$$S = \frac{1}{2}(m^2 - \Omega^2)\phi^2 + \underbrace{\frac{1}{2}\Omega^2\phi^2 + g\phi^4}_{perturbation} \tag{2}$$

and then finding Ω by minimizing the energy at the perturbative level.

Wetterich Equation

This whole workshop is based on the following equation:

$$\partial_k \Gamma_k[\{\phi\}] = \frac{1}{2}\mathrm{Tr}\left(\frac{\partial_k R_k}{\frac{\delta^2}{\delta^2\{\phi\}}\Gamma_k + R_k}\right) \tag{3}$$

which, rather than an equation, is a statement of intent!

- $\Gamma_k[\phi]$ is our guess for the the parameter space dependence
- R_k is a mass, e.g. a regulator: it is our recipe for how modes should be eliminated[2]

[2] Polchinsky (1988) , Wetterich (1993), Morris (1994)

Regulator dependence

One of the key problems of the Wetterich equation is the fact that the regulator R_k is actually arbitrary, e.g. you can introduce your favourite regulator. The flow equation is infinite dimensional, thus one has to work out various approximations to test R_k. In principle the flow should not depend on the regulator, but one uses oftentimes the optimized regulator[3]:

$$R_k(q) = (k^2 - q^2)\theta(k^2 - q^2) \equiv q^2 r(y) \tag{4}$$

as it simplifies the flow equations, with $y = q^2/k^2$. It is considered optimal because the equations converge fast to the UV theory at $k = 0$.

[3] Litim (2000); Litim (2001)

Compactly Supported Smooth (CSS) regulator

The Litim regulator can be generalized to a smooth one[4]:

$$R_k(q) = q^2 r_{css}(y) = \frac{e^{c\frac{y_0^b}{1-y_0^b}} - 1}{e^{c\frac{y^b}{1-y^b}} - 1}\theta(1-y) \tag{5}$$

where y_0, b and c are parameters. **But the key problem is: how to choose among the new parameter space?**

Variational principles (minimizing the free energy) do not seem to work, as one gets a monotonic free energy. Other options is the Principle of Minimum Sensitivity[5], e.g. use parameters such that physical parameter at the least dependent on.

[4] Nandori (2013)
[5] Stevenson (1981)

Information of the RG flow: fast and slow modes

Consider the probability function [6]

$$p(\phi, \psi) = \frac{1}{Z} e^{-S[\phi,\psi]} \quad (6)$$

for a certain physical system, where

$$Z = \int [D\phi][D\psi] e^{-S[\phi,\psi]}. \quad (7)$$

In particular, one can choose ϕ and ψ such that these **are the slow and fast modes in a renormalization group approach**, e.g. we study the effective action

$$e^{W_{eff}[\phi]]} = \int [D\psi] e^{-S[\phi,\psi]}. \quad (8)$$

where ψ are considered the fast mode that need to be integrated out.

[6] Apenko (2009)

Mutual information

Alternatively to the use of variational principles, one can think of information theoretic principles for the study of the renormalization group. For instance, one can think of using the mutual information along the renormalization group flow.Mutual information is defined as a relative entropy measure. If $p(\psi, \phi)$ is the distribution on two variables,

$$M = \sum_{\psi,\phi} p(\phi, \psi) \log \frac{p(\psi, \phi)}{p(\phi)p(\psi)}. \tag{9}$$

In the case of discrete this measure already saw some adopters, obtaining some optimal coarse graining procedures. [7]. Our purpose is to introduce the use of **Fisher information** for reasons that will be clear soon.

[7] Apenko (2009), Lenggenhager et al (2018), Koch-Janusz, Ringel (2018)

Central Limit Theorem and RG

Let us consider a probability distribution $p(x)$. We write these relations for simple 0-dimensional distributions, but these can be generalized to $d-$dimensional systems as well. The probability distribution can be written in terms of the characteristic function $\hat{f}(t)$ as

$$\begin{aligned} p(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\hat{f}(t)dt \\ \hat{f}(t) &= \int_{-\infty}^{\infty} e^{itx}p(x)dx \end{aligned} \tag{10}$$

Let us write the characteristic function in terms of the moments c_k cumulants γ_k, as

$$\hat{f}(t) = \sum_{r=0^\infty} c_r \frac{(it)^r}{r!} = e^{\sum_{r=1}^\infty \gamma_k \frac{(it)^r}{r!}} = e^{\hat{\psi}(t)}, \qquad (11)$$

from which we see that we can write

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx+\hat{\psi}(t)} dt \qquad (12)$$

Let X_i be 2^n random variables distributed according to $p(x_i) = \mathcal{N} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}$. The characteristic function of X_i is given by

$$\hat{f}_{X_i}(t) = e^{i\mu_i t - \frac{t^2\sigma_i^2}{2}}. \qquad (13)$$

Let us consider now the case $\mu_i = 0$, $\sigma_i = 1$ for a single variable.

Let $X_1 \cdots X_n$ be random variables sampled from $F(x)$. Let us assume that $\mathbb{E}X_i = 0, \mathbb{E}X_i^2 = 1$. We define $\gamma = \mathbb{E}X_i^3$, $\tau = \mathbb{E}X_i^4$. We want to study the distribution of the variable

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j. \tag{14}$$

Let us define the renormalization group transformation (grouping)

$$\phi_n(t) = \mathbb{E}e^{itZ_n} = \mathbb{E}[\hat{\phi}^n(\frac{t}{\sqrt{n}})] \equiv \mathbb{R}\phi(t), \tag{15}$$

which is a convolution, and which is known to have a finite variance if the sample variables comes from a distribution which has finite variance.This step can be interpreted as a renormalization group transformation.[8] **The scaling $\frac{t}{\sqrt{n}}$ is to fix the variance.**

[8] Jona-Lasinio (2001); see also notes by Codello

We now use the Taylor expansion for $\phi(\frac{t}{\sqrt{n}})$:

$$\psi(\frac{t}{\sqrt{n}}) = \mathbb{E}\left(1 + \frac{itX}{\sqrt{n}} + \frac{1}{2}\frac{(it)^2X^2}{n} + \frac{1}{6}\frac{(it)^3X^3}{n\sqrt{n}} + \frac{1}{24}\frac{(it)^4X^4}{n^2} + \cdots\right)$$
$$= 1 - \frac{1}{2}\frac{t}{n} - \frac{i}{6}\frac{t^3\gamma}{n\sqrt{n}} + \frac{1}{24}\frac{t^4\tau}{n^2} + \cdots, \quad (16)$$

from which, after elevating to the n-th power, we get

$$\psi_n(\frac{t}{\sqrt{n}}) = \Big((1 - \frac{t^2}{2n})^n + (1 - \frac{t^2}{2n})^{n-1}(-\frac{i}{6}\frac{t^3\gamma}{\sqrt{n}} + \frac{1}{24}\frac{t^4\tau}{n})$$
$$+ (1 - \frac{t^2}{2n})^{n-2}\frac{(n-1)\gamma^2(it)^6}{72n^2}\Big) + \mathcal{O}(\frac{1}{n}) \quad (17)$$

We have

$$\hat{\psi}_n(t) = e^{-\frac{t^2}{2}}\left(1+\frac{(it)^3\gamma}{6\sqrt{n}}+\frac{(it)^4(\tau-3)}{24n}+\frac{(it)^6\gamma^2}{72n}+\cdots\right) \quad (18)$$

We thus find that, using the Hermite polynomials, we have

$$\begin{aligned} P_n(x) &= \int_{-\infty}^{\infty} e^{-itx}\hat{\psi}_n(t)\frac{dt}{2\pi} \\ &= \left(1+\frac{H_3(x)\gamma}{6\sqrt{n}}+\frac{H_4(x)(\tau-3)}{24n}+\frac{H_6(x)\gamma^2}{72n}\right)\phi(x) \end{aligned} \quad (19)$$

which is called Edgeworth expansion. What signals the convergence of the CLT?

Fisher information

Let us now introduce the properties of the Fisher information. The Fisher information is a measure of the amount of information associated with a certain parameter b, specifically,
The quantity

$$I(x, b) = \frac{d}{db} \log p_b(x) \tag{20}$$

is called the score function and is the derivative of the log-likelihood of a certain model, and describes the sensitivity of the probability distribution on the parameter b. The Fisher information is thus just a measure of average score. Fisher information is defined as

$$\langle (I(x, b))^2 \rangle, \tag{21}$$

which is maximum when the Fisher information is high.

Fisher Information 2

There are two equivalent definitions of the Fisher information if the distribution is $\mathcal{C}^2$ in the parameter b:

$$I(\theta) = \int dx\ p_b(x) \left(\partial_b \log p_b(x)\right)^2 = \int dx\ \frac{(\partial_b p_b(x))^2}{p_b(x)}$$

$$\text{If } C^2 = -\int dx\ p_b(x) \partial_b^2 \log p_b(x) \leftarrow \text{our definition} \qquad (22)$$

Specifically, Fisher information is high if the parameter b is informative. Thus:

Fisher information is a measure of the dependence of a distribution on a parameter. The higher I, the more information b contains about the distribution.

Fisher Information 3

There are many definitions of the Fisher information, which we now describe. The Fisher information of a probability distribution $p(x, \vec{b})$ which depends on a set of parameters $\vec{b}$, is defined as

$$I(\vec{b}) = \int dx \; p(x, \vec{b}) \left(\vec{\nabla}_b \log \left(p(x, b) \right) \right)^2 , \tag{23}$$

Example For instance, suppose that for a normal distribution the value of μ is unkown, but σ is given. Then

$$\log p_\mu(x) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}, \tag{24}$$

and thus

$$I(\mu) = \frac{1}{\sigma^2} \leftarrow \textbf{Related to the Cramer-Rao bound} \tag{25}$$

Also you have to admit that Fishers usually get it right

 and 

And yes, it's not the same Fisher

Interestingly, it has been shown using the Edgeworth expansion that if we define the relative Fisher information for the CLT:

$$F(Z_n||Z) = F(Z) - F(Z_n) = \int \left(\frac{p'(x)}{p(x)} - \frac{p_n'(x)}{p_n(x)} \right)^2 p(x) dx \quad (26)$$

one has[9]

$$F(Z_n||Z) = \frac{c_1}{n} + \frac{c_2}{n^2} + \cdots , \quad (27)$$

e.g. Fisher information should be converging in the central limit theorem.

Hypothesis: can we use Fisher information along the flow? Can it tell something new?

[9] Bobkov, Chistyakov, Götze (2014)

Functional case

In the case of the Polchinski and Wetterich approaches for the functional renormalization group, one typically writes

$$Z_k = \int [D\phi] e^{-S[\phi]-S_k[\phi]}. \tag{28}$$

where $S_k[\phi]$ is a regulator which depends on a scale k. This scale is what defines fast and slow modes. Thus,

$$e^{W_k[\phi]]} = \int [D\phi] e^{-S[\phi]-S_k[\phi]}. \tag{29}$$

and where

$$S_k[\phi] = \int_{x,y} \sum_{\alpha,\beta} \frac{1}{2} \phi_\alpha (C_k)_{\alpha,\beta} \phi_\beta \tag{30}$$

and C_k is the regulator. The connection with the previous picture of fast and slow modes is the following.

First we consider a measure of the following type

$$d\mu_\Lambda(\phi) = D\phi e^{-\frac{1}{2}\int_{x,y}\phi(x)C_\Lambda^{-1}(x-y)\phi(y)} \quad (31)$$

with $C_\Lambda(p) = (1-\theta_\epsilon(P,\Gamma))\,C(p)$ where $\theta_\epsilon(P,\Gamma)$ simply represents a UV cutoff and ϵ is a parameter which controls its smoothness. The block spin idea is thus connected to the following decomposition. We assume that we can write in momentum space the following decomposition:

$$\phi_k = \phi_{\langle k} + \phi_{\rangle k} \quad (32)$$

where the scale k defines the modes that are fast and slow. Since now on we drop the symbol k in the fields.

We then divide the propagator in two parts for the two modes, for an arbitrary regulator as before, as

$$\phi_{\langle k} \to C_k(p), \qquad \phi_{\rangle k} \to C_\lambda - C_k(p)$$

where C_k defines our arbitrary scale. Then the partition function is written as

$$Z = \int d\mu_{C_\Lambda}(\phi) e^{-\int_p V(\phi)} = \int d\mu_{C_k}(\phi_\langle) d\mu_{(C_\Lambda - C_k)}(\phi_\rangle) e^{-\int V(\phi_\langle + \phi_\rangle)} \quad (33)$$

Thus, the probability distribution that we had introduced at the beginning of this section can be simply written as

$$p(\phi, \psi) \to p(\phi, k, \{b\}) \quad (34)$$

where $\{b\}$ are the parameters which enter in the family of regulators R_k, and

$$p(\phi, k, b) = \frac{e^{-S(\phi,k,b)}}{Z_k(b)} = \frac{e^{-S[\phi] - S_k(\phi,b)}}{Z_k(b)} \quad (35)$$

An open question is how to choose, given a family of regulators R_k, the right regulators in order to perform the renormalization group.

Imagine that Ötzi had to solve FRG

he would do it in zero dimensions first

Flow equation for 0d field theories, also called integrals

What is the FRG equivalent of the 0d probability "Jona-Lasinio" approach?[10] Consider an integral:

$$Z = \int_{-\infty}^{\infty} f(x)dx \tag{36}$$

with $f(x) \geq 0$, and can be interpreted as a . In this case, we can write [11]

$$f(x) = e^{-S(x)} \tag{37}$$

Consider now a "regulator":

$$f_k(x,j) = e^{-S(x)-\frac{1}{2}k^2x^2-jx} \to Z_k(j) = \int_{-\infty}^{\infty} f_k(x)dx \equiv e^{W_k(j)} \tag{38}$$

[10] Jona-Lasinio (2001)

[11] see for instance Flörchinger thesis(2014)

Following the same derivation for the Wetterich equation, we can write a flow equation as a function of the regulator momentum k

$$\partial_k \Gamma_k(\phi) = \frac{1}{2} \frac{\partial_k k^2}{\frac{\delta^2}{\delta^2 \phi} \Gamma_k(\phi) + k^2} \tag{39}$$

where $\Gamma_k(\phi) = j[\phi]\phi - W_k(j[\phi]) - \frac{1}{2}k^2 x^2$ and $j[\phi] = \sup_\phi \Big(j[\phi]\phi - W_k(j[\phi]) \Big)$. It can be shown that

$$Z_k = e^{-\Gamma_k[\phi=0]} \tag{40}$$

Now, we are interested in the regulator dependence. Consider

$$R_k = \frac{1}{2} k^{2\alpha} x^2 \tag{41}$$

which generalizes to a new parameter α the previous regulator. Can we see that $\alpha = 1$ should be "optimal"?

Danger! Danger! Danger! Simple case follows

Case under control (we know the solution):

$$S(x) = \frac{1}{2} m^2 x^2 \tag{42}$$

but we can solve any integrals. in this case, we must have

$$\lim_{k\to 0} Z_k(j=0) = \frac{\sqrt{2\pi}}{m}. \tag{43}$$

This means that for $m = 1$ this is a very complicated way of calculating $\sqrt{2\pi}$.

In order to solve this integral, we can choose an **initial condition for the (functional) differential equation** such that

$$\lim_{k\to\infty} \Gamma_k[\phi] = S(\phi) + \log\left(\frac{k^{2\alpha}}{\sqrt{2\pi}}\right) \tag{44}$$

this is a technical but important point.

Procedure: standard truncation scheme.

$$\Gamma_k[\phi] = \sum_{j=0}^{n} a_j[k]\phi^k \tag{45}$$

If $n = 3$, we have for instance

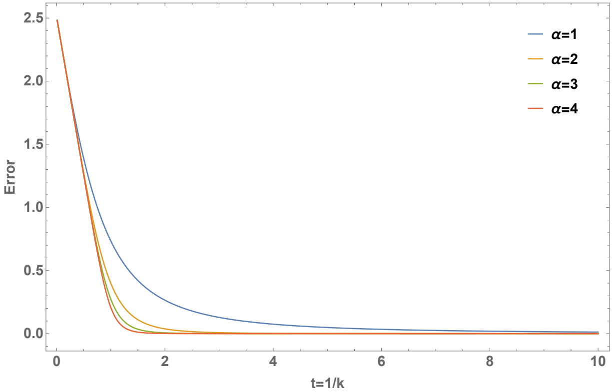
$$a_0'(k) - \frac{k^{2\alpha-1}\alpha}{2a_2 + k^{2\alpha}} = 0 \tag{46}$$

$$\frac{6a_3(k)\alpha 2k^{2\alpha-1}}{(2a_2 + k^{2\alpha})^2} + a_1' = 0 \tag{47}$$

$$a_2'(k) - \frac{36{a_3}^2 k^{2\alpha-1}\alpha}{(2a_2(k) + k^{2\alpha})^3} = 0 \tag{48}$$

$$\frac{216a_3(k)^3\alpha k^{2\alpha-1}}{(2a_2(k) + k^{2\alpha})^4} + a_3'(k) = 0 \tag{49}$$

$|Z_k - \sqrt{2\pi}/m| \to 0$ as $k \to 0$



Hypothesis: **Does the Fisher information say anything about the parameter α?**

Procedure for calculating Fisher

How to calculate $p_k(x,\alpha)$? Using the definition we had seen we have that

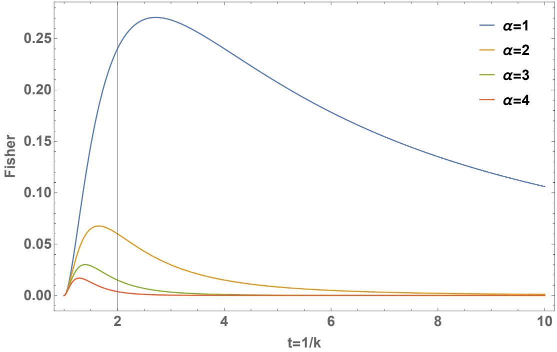
$$p(x,\alpha)=\frac{e^{-S(x)}}{Z}\to p_k(x,\alpha)=\frac{e^{-S(x)-\frac{1}{2}k^{2\alpha}x^2}}{Z_k}=e^{\Gamma_k(\alpha)-S(x)-\frac{1}{2}k^{2\alpha}x^2} \tag{50}$$

We can calculate Fisher information along the flow. As it turns out, we have actually in the Gaussian case

$$\begin{aligned} I(x,\alpha) &= -\int_{-\infty}^{\infty} p(x,k,\alpha)\partial_\alpha^2 \log p(x,k,\alpha) \\ &= \langle x^2\rangle_{p_k(x,\alpha)} f(k,\alpha)+G(k) \end{aligned} \tag{51}$$

with $G(k)\to 0$ as $k\to 0$, and one can use the fact that $\partial_\alpha\Gamma_k=\frac{k\log k}{\alpha}\partial_k\Gamma_k$.

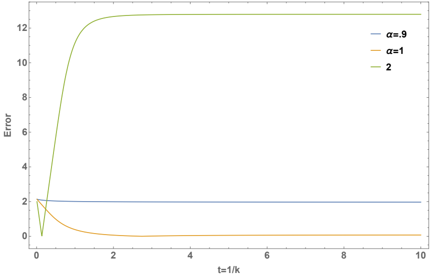
Fisher information seems to suggest a value of $\alpha = 1$



We see that the convergence of the Fisher has a hierarchy in the exponents, which is consistent with the convergence.

Fisher information seems to suggest a value of α

In case for $S[\phi] = \frac{1}{2}\phi^4$



The equations seem to converge only for $\alpha = 1$. *Again: this is a work in progress, take this with a grain of salt.*

Our preliminary results seem to suggest that Fisher information might be a good measure to estimate the role of the parameters: the one that contains the most amount of information about the system, at the expense of a slower convergence. Next steps:

- Test our hypothesis on Landau-Ginzburg $\rightarrow$ Check the critical exponents
- Future work: make $\alpha(k)$ dynamical. We have some ideas in this regard.

The end.

Many thanks to ACRI for support!

caravelli@lanl.gov