1 Basic Idea

1.1 From Self-Similarity to Pattern Recognition

There are many kinds of self-similarity in Nature. Turbulence, for instance, has self-similarity at the critical point of parameters. This self-similarity indicates that, when you zoom-in a picture of turbulence, you should find that the original consists of many smaller turbulence each of which looks quite like the original. By saying "looks like", we mean they share the same "pattern". That is, *they are not exactly the same, but same in pattern, which is recognized by our brain.*

Pattern recognization can also be made by Boltzmann machine (BM) [1], which is a simplified but still efficient model of human brain. In Boltzmann machine, two pictures are recognized as the same pattern if they both locate within the same area of attractor of the corresponding Langevin dynamics.

So, we should connect the operation on the picture of turbulence with the Boltzmann machine that recognizes the patterns. That is, *the pictures before and after the operation should obey the same Boltzmann machine.*

In the next several sections, we expand the theme carefully, declaring what the configuration space and operation should be described in mathematics. And how Boltzmann machine is changed by the operation. This gives birth to renormalization group (RG). After all has been clarified, we can see what self-similarity really means.

2 Renormalization Group

2.1 Configuration Space and Operations

First of all, we declare what the configuration space should be. A picture is numerically described by a 2D array of float type, the size of which determines the precision of the picture. Generally, we should consider the continuous version, while the discrete or array version can be deduced from it, no matter what the precision is. So, a configuration should be described by a real scalar field, say $\varphi(x)$, where x in the region A and $\varphi(x) \in \mathbb{R}$ for each $x \in A$.

Then, the operation of zooming in is nothing but marginalizing some component $\varphi(x)$ in the probability density functional (PDF) of φ, $p[\varphi]$, which gives the probability density on a configuration φ.

Apart from the operation of zooming in picture, there are many kinds of operation that may be interested in. This hint us to generalize the discussion to the most generic case. The mathematical tool for this purpose is representation theory [2]. Let $|\varphi\rangle$ the state of a configuration, and $\{|x\rangle | x \in \mathcal{X}\}$ a complete orthogonal base, which may not be spatial coordinate. The configuration is described by the mode like $\varphi(x) := \langle x|\varphi\rangle$. With this, the general operation should be nothing but marginalizing some mode in the probability density functional $p[\varphi]$.

2.2 Boltzmann Machine

Boltzmann machine describes the probability density functional of configuration φ by a functional called **action** in physics, or **energy** in machine learning, $S[\varphi]$, as

$$\frac{\mathrm{e}^{-S[\varphi]}}{\prod_{x \in \mathcal{X}} \int_{\mathbb{R}} \mathrm{d}[\psi(x)]\, \mathrm{e}^{-S[\psi]}}.$$

1. An example of deep Boltzmann machine used for pattern abstraction on the MNIST dataset can be found here.

2. For representation theory, see Dirac's *The Principles of Quantum Mechanics*.

If the action functional depends only a subset of all modes, say $\{\varphi(x)|x\in V\}$ with $V\subset\mathcal{X}$, then we should add a subscript V to action functional, and the probability density functional becomes

$$\frac{\mathrm{e}^{-S_V[\varphi]}}{\prod_{x\in V}\int_{\mathbb{R}}\mathrm{d}[\psi(x)]\,\mathrm{e}^{-S_V[\psi]}}.$$

2.3 Continuous Symmetries and Gauge Fixing

We may have translational symmetry. Let relaxation $\varphi\to\varphi_*$, which φ_* denotes the attractor on the area of which φ sits. Let $\psi(x):=\varphi(x+z)$ for constant z, and relaxation $\psi\to\psi_*$. If translational symmetry holds, we shoud expect that $\psi_*(x)=\varphi_*(x+z)$ and that $S[\varphi_*]=S[\psi_*]$. This implies a gauge problem: the extremum of S is not a single value, but a sub-manifold along the symmetry.

The same holds for any other continous symmetry, such as rotational symmetry.

To deal with this gauge problem, consider a Boltzmann machine that is to learn a rectangle. It will relax a perturbed rectangle to the "standard" one, the learned pattern. This learning task encounters the translational symmetry: a rectangle is still the same rectangle after being translationally moved. The method to solve this problem is gauge fixing. For instance, the dataset is a colleciton of hand-drawn rectangle images, and φ represents the gray level in range $[0,1]$. We are to move all images in dataset to be centered at the original. This can be done by shifting $x\to x-m$ where $m:=\mathrm{mean}(\{x|\varphi(x)>0\})$. Because of central limit theorem, this m is stable for random perturbation. After this shifting, all images are properly centered, and the gauge is fixed.

This can be seen as a re-definition of coordinates. Indeed, in the case of rotational symmetry, we re-define the Cartesian coordinates to polar coordinates. As in the case of translational symmetry, this re-definition of coordinates fixes the gauge caused by rotational symmetry.

2.4 Renormalization Group

Next, we perform the operation that marginalizes some modes. Let $V'\subset V$. Marginalizing the modes in $V\setminus V'$ results in

$$\prod_{x\in V\setminus V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]\frac{\mathrm{e}^{-S_V[\varphi]}}{\prod_{x\in V}\int_{\mathbb{R}}\mathrm{d}[\psi(x)]\,\mathrm{e}^{-S_V[\psi]}}.$$

On the other hand, this probability density functional of configuration should also be described by a Boltzmann machine, which has action functional $S_{V'}$.

$$\prod_{x\in V\setminus V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]\frac{\mathrm{e}^{-S_V[\varphi]}}{\prod_{x\in V}\int_{\mathbb{R}}\mathrm{d}[\psi(x)]\,\mathrm{e}^{-S_V[\psi]}}=\frac{\mathrm{e}^{-S_{V'}[\varphi]}}{\prod_{x\in V'}\int_{\mathbb{R}}\mathrm{d}[\psi(x)]\,\mathrm{e}^{-S_{V'}[\psi]}}.$$

This equation has the solution

$$\mathrm{e}^{-S_{V'}[\varphi]}=C\prod_{x\in V\setminus V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]\,\mathrm{e}^{-S_V[\varphi]},$$

where C is independent of φ. This is the **renormalization group**.

Indeed, by applying $\prod_{x\in V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]$ on both sides, we find up to a constant,

$$\prod_{x\in V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]\mathrm{e}^{-S_{V'}[\varphi]}=\prod_{x\in V}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]\,\mathrm{e}^{-S_V[\varphi]},$$

which is the starting point of deriving non-perturbative renormalization group equation given by Aoki, equation (77). If $V'\approx V$, the integration in the solution can be simplified by linear approximation, which turns to be the renormalizaiton group equation.

It should be noted that the expression of renormalization group is independent of the choice of complete orthogonal base. Indeed, changing bases to another complete orthogonal base $\{|\,y\rangle|y\in\mathcal{Y}\}$ involves a unitary transformation $\hat{U}$, which is a collection of $\langle x|\,y\rangle$. Since the product in $\prod_{x\in V\setminus V'}\int_{\mathbb{R}}\mathrm{d}[\varphi(x)]$ is in fact a wedged product. So, formally, this unitary transformation results in an extra term $\det(\hat{U})$, which is independent of φ, thus can be absorbed into the factor C. This will leave the expression formally invariant.

2.5 Self-Similarity in Renormalization Group

By the previous discussion, the same in pattern means the same in Boltzmann machine. This implies the equality of actional functional, before and after the operation. That is, $S_V = S_{V'}$.

3 Renormalization Group Equation

3.1 Renormalization Group Equation

Consider a continuous family of V, $\{V(t)|t\in[0,1]\}$, such that $V(0)=V$ and $V(1)=V'$, and that $V(t)\subset V(t')$ as long as $t>t'$. This family describes a "continuous compression" from V to V', which in turn gives birth to a functional autonomous differential equation of $S_{V(t)}$, called renormalization group equation (RGE).

Now, we are to derive the explicit form of this equation. Given t, we start at separating $\varphi(x)$ as $\{\varphi(x)|x\in V(t)\}$ and $\{\varphi(x)|x\in \mathrm{d}V(t)\}$, where $\mathrm{d}V(t):=V(t)\setminus V(t+\mathrm{d}t)$ To make it apparent, we use ϕ for the later. So, the action functional $S_{V(t)}[\varphi]$ is turned to be $S_{V(t)}[\varphi,\phi]$, wherein the φ with $x\in V(t)$ may be coupled with the ϕ. When $\phi=0$, φ is decoupled with ϕ in $S_{V(t)}$. Our aim is to derive the difference between $S_{V(t)}[\varphi,0]$ and $S_{V(t+\mathrm{d}t)}[\varphi]$, where the φ in both action functional run over the same $V(t+\mathrm{d}t)$. With this declaration, the renormalization group becomes

$$\exp(-S_{V(t+\mathrm{d}t)}[\varphi]) = C\prod_{x\in\mathrm{d}V(t)}\int_{\mathbb{R}}\mathrm{d}[\phi(x)]\exp(-S_{V(t)}[\varphi,\phi]).$$

By multiplying $\exp(S_{V(t)}[\varphi,0])$ on both sides, we get

$$\exp(-S_{V(t+\mathrm{d}t)}[\varphi]+S_{V(t)}[\varphi,0]) = C\prod_{x\in\mathrm{d}V(t)}\int_{\mathbb{R}}\mathrm{d}[\phi(x)]\exp(-S_{V(t)}[\varphi,\phi]+S_{V(t)}[\varphi,0]).$$

We expand the term in the integrand as

$$\begin{aligned}
&-S_{V(t)}[\varphi,\phi]+S_{V(t)}[\varphi,0]\\
=&-\int_{\mathrm{d}V(t)}\mathrm{d}x\,\frac{\delta S_{V(t)}}{\delta\phi(x)}[\varphi,0]\phi(x)\\
&-\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\int_{\mathrm{d}V(t)}\mathrm{d}x'\,\frac{\delta^2 S_{V(t)}}{\delta\phi(x)\delta\phi(x')}[\varphi,0]\phi(x)\phi(x')\\
&-\frac{1}{6}\int_{\mathrm{d}V(t)}\mathrm{d}x\int_{\mathrm{d}V(t)}\mathrm{d}x'\int_{\mathrm{d}V(t)}\mathrm{d}x''\,\frac{\delta^3 S_{V(t)}}{\delta\phi(x)\delta\phi(x')\delta\phi(x'')}[\varphi,0]\phi(x)\phi(x')\phi(x'')\\
&-\cdots
\end{aligned}$$

Plugging this expansion back, we find that the first two terms furnishes a functional Gaussian integral

$$C\prod_{x\in\mathrm{d}V(t)}\int_{\mathbb{R}}\mathrm{d}[\phi(x)]\exp\left(-\int_{\mathrm{d}V(t)}\mathrm{d}x\,\frac{\delta S_{V(t)}}{\delta\phi(x)}[\varphi,0]\phi(x)-\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\int_{\mathrm{d}V(t)}\mathrm{d}x'\,\frac{\delta^2 S_{V(t)}}{\delta\phi(x)\delta\phi(x')}[\varphi,0]\phi(x)\phi(x')\right).$$

Since $(\delta^2 S_{V(t)}/\delta\phi(x)\delta\phi(x'))[\varphi,0]$ is real and symmetric on x and x', and $\{|x\rangle | x\in \mathrm{d}V(t)\}$ is complete on the sub-Hilbert-space where the ϕ sits, this integral can be integrated out, as [3]

$$C\exp\left\{\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\int_{\mathrm{d}V(t)}\mathrm{d}x'\frac{\delta S_{V(t)}}{\delta\phi(x)}[\varphi,0]\left[\frac{\delta^2 S_{V(t)}}{\delta\phi\delta\phi}[\varphi,0]\right]^{-1}(x,x')\frac{\delta S_{V(t)}}{\delta\phi(x')}[\varphi,0]-\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\ln\left(\frac{\delta^2 S_{V(t)}}{\delta\phi(x)\delta\phi(x)}[\varphi,0]\right)\right\}$$

The other terms can be seen as a perturbative expansion based on this Gaussian term, and thus all are proportional to higher order of $|\mathrm{d}V(t)|$, thus omittable. So, letting

$$\mathrm{d}S_{V(t)}[\varphi]:=S_{V(t+\mathrm{d}t)}[\varphi]-S_{V(t)}[\varphi,0],$$

we arrive at a differential equation, up to a φ-independent term,

$$\begin{aligned}\mathrm{d}S_{V(t)}[\varphi]=&\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\int_{\mathrm{d}V(t)}\mathrm{d}x'\frac{\delta S_{V(t)}}{\delta\phi(x)}[\varphi,0]\left(\frac{\delta^2 S_{V(t)}}{\delta\phi\delta\phi}[\varphi,0]\right)^{-1}(x,x')\frac{\delta S_{V(t)}}{\delta\phi(x')}[\varphi,0]\\&-\frac{1}{2}\int_{\mathrm{d}V(t)}\mathrm{d}x\ln\left(\frac{\delta^2 S_{V(t)}}{\delta\phi(x)\delta\phi(x)}[\varphi,0]\right),\end{aligned}$$

which is called (non-perturbative) **renormalization group equation**.

3.2 Self-Similarity May Relate to Limit Circle

How is self-similarity characterized by this renormalization group equation? An educated guess is that self-similarity is a limit circle of this autonomous differential equation. It starts at a point and travels along a circle. Finally it goes back to the starting point: the self-similarity. And then, it starts the same trip again.

4 Construction of Action Functional

4.1 Vanilla Boltzmann Machine with Locality

We are to consider the explicit form of the action functional for the pictures of, for instance, turbulence. For vanilla Boltzmann machine, the action functional would be

$$S[\varphi]=\int\mathrm{d}x\,J(x)\,\varphi(x)+\int\mathrm{d}x\,\mathrm{d}y\,\varphi(x)\,W(x,y)\,\varphi(y)$$

for some bias J and kernel W. In the case of field theory, the kernel would be $W(x,y)=-\delta(x,y)\times[(1/2)\,(\partial^2/\partial y^2)+V(y)]$ for some "mass function" V. This form of kernel is local which means $W(x,y)\propto\delta(x,y)$. Under the locality assumption, we have the most general form of kernel: $W(x,y)=\delta(x,y)\,w(y)$, where

$$w(x)=a_0(x)+a_1(x)\,\partial^2+\cdots+a_n(x)\partial^{2n}+\cdots.$$

In this case, action functional is reduced to

$$S[\varphi]=\int\mathrm{d}x\,[J(x)\varphi(x)+a_0(x)\,\varphi^2(x)+a_1(x)\varphi(x)\,\partial^2\varphi(x)+\cdots]$$

The higher derivatives are involved, the larger range of "connections" between the "neurons". [4] Indeed, a function can be recovered in a larger range if we have higher derivatives on the origin.

3. For functional Gaussian integral, see appendix A.

4. Here the words "connection" and "neuron" come from the analogy of Boltzmann machine with human brain. The $W(x,y)$ is analogy to the weight between the neurons at x and y.

A Functional Gaussian Integral

Theorem 1. *[Functional Gaussian Integral] Let $A(x,x')$ and $b(x)$ real functions with $A(x,x')=A(x',x)$ for each x and x'. Suppose that the orthogonal base $|x\rangle$ is complete, that is $\int_V \mathrm{d}x|x\rangle\langle x|=1$ for some set V. Then we have the functional integral*

$$\prod_{x\in V}\int_{\mathbb{R}}\mathrm{d}[\phi(x)]\exp\left(-\frac{1}{2}\int_V\mathrm{d}x\int_V\mathrm{d}x'\,A(x,x')\,\phi(x)\phi(x')-\int_V\mathrm{d}x\,b(x)\,\phi(x)\right)$$
$$=C\exp\left\{\frac{1}{2}\int_V\mathrm{d}x\int_V\mathrm{d}x'\,b(x)(A^{-1})(x,x')b(x')-\frac{1}{2}\int_V\mathrm{d}x\,\ln A(x,x)\right\},$$

where C is independent of ϕ, and $\int_V\mathrm{d}x'\,A(x,x')\,(A^{-1})(x',x'')=\delta(x-x'')$.

The proof of this theorem needs several tools.

Lemma 2. *[Multi-Dimensional Gaussian Integral] We have*

$$\prod_{\alpha=1}^{n}\int_{\mathbb{R}}\mathrm{d}x^{\alpha}\exp\left(-\frac{1}{2}A_{\alpha\alpha'}\,x^{\alpha}\,x^{\alpha'}+b_{\alpha}\,x^{\alpha}\right)=\sqrt{\frac{(2\pi)^n}{\det(A)}}\exp\left(\frac{1}{2}(A^{-1})^{\alpha\alpha'}\,b_{\alpha}\,b_{\alpha'}\right),$$

where A is real symmetric matrix and b is real vector.

Proof. Left to reader. ☐

Lemma 3. *For a real matrix A, we have*

$$\det(A)=\mathrm{tr}\ln(A).$$

Proof. Left to reader ☐

Lemma 4. *Given an operator A and a complete orthogonal base $\{|x\rangle|x\in V\}$ with continuous spectrum, we have*

$$\ln A=\int\mathrm{d}x\,\mathrm{d}x'\ln\langle x|A|x'\rangle|x\rangle\langle x'|.$$

Proof. Left to reader. ☐

Lemma 5. *Given an operator A and a complete orthogonal base $\{|x\rangle|x\in V\}$ with continuous spectrum, we have*

$$\mathrm{tr}\ln A=\int_V\mathrm{d}x\ln\langle x|A|x\rangle.$$

Proof. To get this result, we start at the conclusion

$$\ln A=\int\mathrm{d}x\,\mathrm{d}x'\ln\langle x|A|x'\rangle|x\rangle\langle x'|.$$

Recall that, $\operatorname{tr}\ln A$ is well-defined in discrete spectrum. For instance, given a complete orthogonal base $|\alpha\rangle$ with discrete spectrum, $\operatorname{tr}\ln A := \sum_\alpha \langle\alpha|\ln A|\alpha\rangle$. Then we have

$$
\begin{aligned}
\operatorname{tr}\ln A &:= \sum_\alpha \langle\alpha|\ln A|\alpha\rangle \\
\{\ln A = \cdots\} &= \int \mathrm{d}x\,\mathrm{d}x' \sum_\alpha \langle x'|\alpha\rangle\langle\alpha|x\rangle \ln\langle x|A|x'\rangle \\
\left\{\sum_\alpha |\alpha\rangle\langle\alpha| = 1\right\} &= \int \mathrm{d}x \ln\langle x|A|x\rangle,
\end{aligned}
$$

which is the $\operatorname{tr}\ln A$ expressed in continuous spectrum. □

Now, we come to prove the main theorem 1.

Proof. The completeness of $|x\rangle$ helps convert to a discrete spectrum $\{|\alpha\rangle | \alpha \in I\}$, which results in $\phi(x) := \langle x|\phi\rangle = \sum_\alpha \langle x|\alpha\rangle\langle\alpha|\phi\rangle =: \sum_\alpha \langle x|\alpha\rangle \phi^\alpha$ where we suppose that ϕ^α is real for each α,

$$
\begin{aligned}
& - \int_V \mathrm{d}x\, b(x)\phi(x) \\
\{\phi \text{ is real}\} &= - \int_V \mathrm{d}x\, b(x)\phi^*(x) \\
\left\{\int_V \mathrm{d}x\, |x\rangle\langle x| = 1\right\} &= -\langle\phi|b\rangle \\
\left\{\sum_\alpha |\alpha\rangle\langle\alpha| = 1\right\} &= -b_\alpha \phi^{*\alpha} \\
\{\phi^\alpha \text{ is real}\} &= -b_\alpha \phi^\alpha,
\end{aligned}
$$

and

$$
\begin{aligned}
& -\frac{1}{2}\int_V \mathrm{d}x \int_V \mathrm{d}x'\, A(x,x)\phi(x)\phi(x') \\
\left\{\int_V \mathrm{d}x\, |x\rangle\langle x| = 1\right\} &= -\frac{1}{2}\langle\phi|A|\phi\rangle \\
\left\{\sum_\alpha |\alpha\rangle\langle\alpha| = 1\right\} &= -\frac{1}{2}A_{\alpha\alpha'}\phi^{*\alpha}\phi^{\alpha'} \\
\{\phi^\alpha \text{ is real}\} & -\frac{1}{2}A_{\alpha\alpha'}\phi^\alpha\phi^{\alpha'}.
\end{aligned}
$$

With these, the integral turns to be

$$
C \prod_{\alpha=1}^{n} \int_{\mathbb{R}} \mathrm{d}\phi^\alpha \exp\left(-\frac{1}{2}A_{\alpha\alpha'}\phi^\alpha\phi^{\alpha'} - b_\alpha\phi^\alpha\right),
$$

where $C := \det(U)$ and unitary matrix $U_{x,\alpha} := \langle x|\alpha\rangle$. It gives

$$
C' \exp\left(\frac{1}{2}\, b_\alpha (A^{-1})^{\alpha\alpha'} b_{\alpha'} - \frac{1}{2}\operatorname{tr}\ln A\right),
$$

where $C' := \sqrt{(2\pi)^n}\, C$. The final step is converting from $|\alpha\rangle$ back to $|x\rangle$. The first term naturally goes to

$$
\frac{1}{2}\int_V \mathrm{d}x \int_V \mathrm{d}x'\, b(x)(A^{-1})(x,x')b(x').
$$

And with the aid of formula $\operatorname{tr}\ln A = \int_V \mathrm{d}x \ln \langle x|A|x\rangle$, the second term gives

$$-\frac{1}{2}\int_V \mathrm{d}x \ln A(x,x).$$

Altogether, we get the final expression

$$C' \exp\left\{\frac{1}{2}\int_V \mathrm{d}x \int_V \mathrm{d}x'\, b(x)(A^{-1})(x,x')b(x') - \frac{1}{2}\int_V \mathrm{d}x \ln A(x,x)\right\}. \qquad \square$$