Information geometry and the renormalization group

Stefan Floerchinger (Friedrich-Schiller-Universität Jena)

12th International Conference on the Exact Renormalization Group,
September 2024
Les Diablerets, Suisse

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

The ideas of information geometry

[Ronald A. Fisher, Calyampudi R. Rao, Shun'ich Amari, Nikolai N. Chentsov, ...]

- studies spaces of probability distributions $p(x, \xi)$ with parameters ξ^α
- Fisher information metric (symmetric, positive semi-definite)

$$G_{\alpha\beta}(\xi) = \int dx\, p(x,\xi) \left(\frac{\partial}{\partial \xi^\alpha} \ln p(x,\xi) \right) \left(\frac{\partial}{\partial \xi^\beta} \ln p(x,\xi) \right)$$

- unique Riemannian metric that is invariant under sufficient statistics [Chentsov 1972]
- higher geometric structure: pair of dual connections, non-metricity etc. [Amari, Chentsov, ...]
- extension to quantum states $\rho(\xi)$
- geometric structure follows from a *divergence* or *relative entropy*

$$D(p\|q) = \int dx\; p(x) \ln(p(x)/q(x))$$

Sufficient statistics and Chentsov's theorem

- start from random variable x with probability distribution $p(x, \xi)$ where ξ^α are parameters
- consider map to new random variable $x \to y = f(x)$ with probability distribution $q(y, \xi)$
- information about ξ^α could get lost in the map
- new random variable y is called *sufficient statistic* for ξ when no information about ξ is lost:

$$p(x, \xi) = p(x|y, \xi) q(y, \xi) = r(x) q(y, \xi) \qquad \text{factorizes}$$

or

$$p(x|y, \xi) = \frac{p(x, \xi)}{q(y, \xi)} = r(x) \qquad \text{independent of} \quad \xi^\alpha$$

- Chentsov's theorem: unique invariant metric for sufficient statistic is

$$\begin{aligned} G_{\alpha\beta}(\xi) &= \int dx\, p(x, \xi) \left(\frac{\partial}{\partial \xi^\alpha} \ln p(x, \xi) \right) \left(\frac{\partial}{\partial \xi^\beta} \ln p(x, \xi) \right) \\ &= \int dy\, q(y, \xi) \left(\frac{\partial}{\partial \xi^\alpha} \ln q(y, \xi) \right) \left(\frac{\partial}{\partial \xi^\beta} \ln q(y, \xi) \right) \end{aligned}$$

Relative entropy

- classical relative entropy or Kullback-Leibler divergence

$$D(p\|q) = \sum_j p_j \ln(p_j/q_j)$$

- not symmetric distance measure, but a *divergence*

$$D(p\|q) \geq 0 \qquad \text{and} \qquad D(p\|q) = 0 \quad \Leftrightarrow \quad p = q$$

- quantum relative entropy of two density matrices (also a *divergence*)

$$D(\rho\|\sigma) = \mathsf{Tr}\,\{\rho\,(\ln\rho - \ln\sigma)\}$$

- signals how well state ρ can be distinguished from a model σ
- Gibbs inequality: $D(\rho\|\sigma) \geq 0$
- $D(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$

Significance of Kullback-Leibler divergence

Uncertainty deficit

- true distribution p_j and model distribution q_j
- *uncertainty deficit* is expected surprise $\langle -\ln q_j \rangle = -\sum_j p_j \ln q_j$ minus real information content $-\sum_j p_j \ln p_j$

$$D(p\|q) = -\sum_j p_j \ln q_j - \left(-\sum_j p_j \ln p_j \right)$$

Asymptotic frequencies

- true distribution q_j and frequency after N drawings $p_j = \frac{N(x_j)}{N}$
- probability to find frequencies p_j for large N (similar: Sanov theorem)

$$\sim e^{-ND(p\|q)}$$

- probability for fluctuation around expectation value $\langle p_j \rangle = q_j$ tends to zero for large N and when divergence $D(p\|q)$ is large

Advantages of relative entropy: continuum limit

- consider transition from discrete to continuous distributions

$$p_j \to f(x)\,dx \qquad\qquad q_j \to g(x)\,dx$$

- not well defined for entropy

$$S = -\sum p_j \ln p_j \overset{\lightning}{\to} -\int dx\, f(x)\,[\ln f(x) + \ln dx]$$

- relative entropy remains well defined

$$D(p\|q) \to D(f\|g) = \int dx\, f(x) \ln(f(x)/g(x))$$

Information geometry for Euclidean quantum fields

[S. Floerchinger, 2303.04081 and Phys. Lett B 846, 138244 (2023)]

- classical statistical field theories
- bosonic quantum fields with real action in Euclidean space
- extension of information geometry
- new flow equation for divergence functional

Probabilities for Euclidean fields: exponential family

- probability density for Euclidean field theory with respect to measure $D\chi$

$$p[\chi, J] = \exp\left(-I[\chi] + J^{\alpha}\phi_{\alpha}[\chi] - W[J]\right)$$

- uses abstract index notation

$$J^{\alpha}\phi_{\alpha} = \int_x \sum_n J_n(x)\phi_n(x)$$

- partition function

$$e^{W[J]} = \int D\chi \exp\left(-I[\chi] + J^{\alpha}\phi_{\alpha}[\chi]\right)$$

- sources J^{α} could also compromise coupling constants
- will be considered as coordinates on space of probability distributions
- known as exponential family in information geometry

Affine geometry for sources

- exponential family is closed with respect to affine transformations

$$J^\alpha \to J'^\alpha = M^\alpha_{\ \beta} J^\beta + c^\alpha$$

- affine transformations respect convexity of $W[J]$
- so-called e-geodesics

$$J^\alpha(t) = (1-t)J'^\alpha + tJ''^\alpha$$

characterized by differential equation

$$\frac{d^2}{dt^2} J^\alpha(t) + (\Gamma_\mathsf{E})_\beta{}^\alpha{}_\gamma[J] \left(\frac{d}{dt} J^\beta(t)\right)\left(\frac{d}{dt} J^\gamma(t)\right) = 0$$

where the connection vanishes in terms of source coordinates

$$(\Gamma_\mathsf{E})_\beta{}^\alpha{}_\gamma[J] = 0$$

Fisher information metric

- Fisher information metric

$$G_{\alpha\beta}[J] = \int D\chi\, p[\chi, J]\, \frac{\delta}{\delta J^\alpha} \ln p[\chi, J]\, \frac{\delta}{\delta J^\beta} \ln p[\chi, J]$$
$$= -\int D\chi\, p[\chi, J]\, \frac{\delta^2}{\delta J^\alpha \delta J^\beta} \ln p[\chi, J]$$

- Fisher-Rao distance between nearby probability distributions

$$ds^2 = G_{\alpha\beta}[J]\, dJ^\alpha\, dJ^\beta$$

- for the exponential family

$$G_{\alpha\beta}[J] = \frac{\delta^2}{\delta J^\alpha \delta J^\beta} W[J] = \langle \phi_\alpha[\chi]\phi_\beta[\chi]\rangle - \langle \phi_\alpha[\chi]\rangle\langle \phi_\beta[\chi]\rangle$$

- equal to connected two-point correlation function !
- generalization of Zamolodchikov metric for conformal field theories

Expectation value coordinates

- can also use field expectation values as coordinates for $p[\chi, \Phi]$

$$\Phi_\alpha = \langle \phi_\alpha[\chi] \rangle = \frac{\delta}{\delta J^\alpha} W[J] = \int D\chi \, p[\chi, J] \, \phi_\alpha[\chi]$$

- best described in terms of quantum effective action

$$\Gamma[\Phi] = \sup_J \left(J^\alpha \Phi_\alpha - W[J] \right) = - \inf_J \left(- \int D\chi \, p[\chi, J] \, \ln p[\chi, J] \right)$$

- Fisher-Rao distance

$$ds^2 = G_{\alpha\beta}[J] \, \delta J^\alpha \delta J^\beta = G^{\alpha\beta}[\Phi] \, \delta\Phi_\alpha \delta\Phi_\beta = \delta J^\alpha \delta \Phi_\beta$$

- Fisher metric in expectation value coordinates

$$G^{\alpha\beta}[\Phi] = - \int D\chi \, p[\chi, \Phi] \, \frac{\delta^2}{\delta\Phi_\alpha \delta\Phi_\beta} \ln p[\chi, \Phi] = \frac{\delta^2 \Gamma[\Phi]}{\delta\Phi_\alpha \delta\Phi_\beta}$$

- another affine structure, dual to the one for sources

$$\Phi_\alpha \to \Phi'_\alpha = N_\alpha{}^\beta \Phi_\beta + d_\alpha$$

- defines so-called m-connection

Divergence functional in source coordinates

- functional generalization of Kullback-Leibler divergence

$$D[J\|J'] = \int D\chi \, p[\chi, J] \ln \left(p[\chi, J]/p[\chi, J'] \right)$$

- compares two probability distributions in asymmetric way
- non-negative

$$D[J\|J'] \geq 0$$

- equals Fisher-Rao distance for close-by distributions

$$D[J\|J'] = \frac{1}{2} G_{\alpha\beta}[J] \delta J^\alpha \delta J^\beta + \ldots$$

- characterizes probabilities for large deviations (Sanovs theorem)
- can be written as Bregman divergence

$$D[J\|J'] = (J^\alpha - J'^\alpha) \frac{\delta W[J]}{\delta J^\alpha} - W[J] + W[J']$$

- functional derivatives w.r.t. second argument yield connected correlation functions !

Divergence functional in expectation value coordinates

- Divergence functional in terms of expectation values

$$D[\Phi\|\Phi'] = \int D\chi\, p[\chi,\Phi] \ln\left(p[\chi,\Phi]/p[\chi,\Phi']\right)$$
$$= \Gamma[\Phi] - \Gamma[\Phi'] - \frac{\delta\Gamma[\Phi']}{\delta\Phi'_\lambda}(\Phi_\lambda - \Phi'_\lambda)$$

- functional derivatives w.r.t. first argument yield one-particle irreducible correlation functions (for $n \geq 2$)

$$D^{(n,0)}[\Phi\|\Phi'] = \Gamma^{(n)}[\Phi],$$

- mixed representation generates connected and 1-P.I. correlation functions

$$D[\Phi\|J'] = \Gamma[\Phi] + W[J'] - J'^{\alpha}\Phi_\alpha$$

Functional integral representations

- divergence functional in source coordinates

$$e^{-D[J\|J']} = \frac{e^{W[J]-J^\alpha\Phi_\alpha}}{e^{W[J']-J'^\alpha\Phi_\alpha}} = \frac{\int D\chi\ \exp\left(-I[\chi] + J^\alpha(\phi_\alpha[\chi] - \Phi_\alpha)\right)}{\int D\tilde{\chi}\ \exp\left(-I[\tilde{\chi}] + J'^\alpha(\phi_\alpha[\tilde{\chi}] - \Phi_\alpha)\right)}$$

- well defined as ratio of functional integrals
- similar in expectation value coordinates

Geometry from divergence

- Fisher metric from functional derivative of divergence

$$G_{\alpha\beta}[J] = -\frac{\delta^2}{\delta J^\alpha \delta J'^\beta} D[J\|J']\big|_{J=J'}$$

- transforms automatically as a metric under coordinate changes $J \to K[J]$
- m-connection symbols

$$(\Gamma_{\mathsf{M}})_{\alpha\beta\gamma}[J] = -\frac{\delta^2}{\delta J^\alpha \delta J^\gamma} \frac{\delta}{\delta J'^\beta} D[J\|J']\big|_{J=J'}.$$

- e-connection symbols

$$(\Gamma_{\mathsf{E}})_{\alpha\beta\gamma}[J] = -\frac{\delta}{\delta J^\beta} \frac{\delta^2}{\delta J'^\alpha \delta J'^\beta} D[J\|J']\big|_{J=J'}$$

- automatically transform like connections under $J \to K[J]$
- information geometry nicely encoded in divergence functional !
- expectation values are another useful coordinate choice

Regularized probability distribution

- introduce now quadratic regulator in probability density

$$p_k[\phi, J] = \exp\left(-S[\phi] - \frac{1}{2}R_k^{\alpha\beta}\phi_\alpha\phi_\beta + J^\alpha\phi_\alpha - W_k[J]\right),$$

- with modified partition function

$$e^{W_k[J]} = \int D\phi \exp\left(-S[\phi] - \frac{1}{2}R_k^{\alpha\beta}\phi_\alpha\phi_\beta + J^\alpha\phi_\alpha\right).$$

- regulator can be chosen to suppress fluctuations, e. g.

$$R_k^{\alpha\beta} = k^2\delta^{\alpha\beta}$$

Divergence functionals with regulator

- divergence functional with regulator

$$\tilde{D}_k[J\|J'] = \int D\phi\, p_k[\phi, J] \ln(p_k[\phi, J]/p_k[\phi, J'])$$
$$= (J^\alpha - J'^\alpha)\frac{\delta W_k[J]}{\delta J^\alpha} - W_k[J] + W_k[J'],$$

- flowing divergence in expectation value coordinates with regulator terms subtracted

$$D_k[\Phi\|\Phi'] = \tilde{D}_k[\Phi\|\Phi'] - \frac{1}{2} R_k^{\alpha\beta}(\Phi_\alpha - \Phi'_\alpha)(\Phi_\beta - \Phi'_\beta)$$
$$= \Gamma_k[\Phi] - \Gamma_k[\Phi'] - \frac{\delta \Gamma_k[\Phi']}{\delta \Phi'_\lambda}(\Phi_\lambda - \Phi'_\lambda).$$

Limit of large and small regulator

- for large k saddle point approximation becomes valid

$$\lim_{k\to\infty} D_k[\Phi\|\Phi'] = S[\Phi] - S[\Phi'] - \frac{\delta}{\delta\Phi'_\alpha} S[\Phi'](\Phi_\alpha - \Phi'_\alpha)$$

- for small k the full Kullback-Leibler divergence functional is recovered

$$\lim_{k\to 0} D_k[\Phi\|\Phi'] = D[\Phi\|\Phi']$$

Flow equation for the divergence functional

[S. Floerchinger, Phys. Lett B 846, 138244 (2023)]

- exact flow equation

$$\frac{\partial}{\partial k} D_k[\Phi\|\Phi'] = \frac{1}{2}\left(\frac{\partial}{\partial k} R_k^{\alpha\beta}\right)\Big[(G_k[\Phi])_{\alpha\beta} - (G_k[\Phi'])_{\alpha\lambda}\,(\tilde{D}_k^{(0,2)}[\Phi\|\Phi'])^{\lambda\kappa}\,(G_k[\Phi'])_{\kappa\beta}\Big]$$

- close relative of Polchinskis and Wetterichs equations
- starting point for approximate solutions
- can be used to flow from large to small regulators
- flow vanishes when $\Phi = \Phi'$
- general coordinates changes possible

Conclusions

- information geometry concepts can be applied to quantum and statistical field theories
- divergence functional encodes the information about geometry: metric, e-connection, m-connection etc.
- divergence functional is generating functional for connected and one-particle irreducible correlation functions
- new exact flow equation for divergence functional

Backup

Advantages / disadvantages of divergence functional

- information theoretic meaning
- positivity $D[\Phi\|\Phi'] \geq 0$ instead of convexity for $\Gamma[\Phi]$
- geometric realization
  - connected correlation functions: e-connection
  - one-particle irreducible: m-connection
- general coordinate changes $\Phi \to \Psi[\Phi]$

$$D[\Psi\|\Psi'] = D[\Phi[\Psi]\|\Phi'[\Psi']]$$

  preserve geometric structure
- equilibrium expectation value Φ_{eq} corresponding to $J = 0$ must be known in addition

Square roots of probabilities

- Fisher information metric

$$G_{\alpha\beta}(\xi) = \int dx\, p(x,\xi) \left(\frac{\partial}{\partial \xi^\alpha} \ln p(x,\xi) \right) \left(\frac{\partial}{\partial \xi^\beta} \ln p(x,\xi) \right)$$
$$= 4 \int dx \left(\frac{\partial}{\partial \xi^\alpha} \sqrt{p(x,\xi)} \right) \left(\frac{\partial}{\partial \xi^\beta} \sqrt{p(x,\xi)} \right)$$

- for discrete random variable, take coordinates

$$p_j = \xi_j^2, \qquad j = 1, \ldots, N.$$

- normalization implies

$$\xi_1^2 + \ldots + \xi_N^2 = 1$$

- Fisher information metric is just induced Euclidean metric on the sphere!