Extension of information geometry to non-statistical systems: some examples

Jan Naudts and Ben Anthonis

Universiteit Antwerpen

Saclay, October 2015

J. Naudts and B. Anthonis, *Extension of Information Geometry to Non-statistical Systems: Some Examples,* in: Geometric Science of Information, GSI 2015 LNCS proceedings, F. Nielsen and F. Barbaresco eds., (Springer, 2015), ISBN 978-3-319-25039-7, p. 427–434

Outline

I An abstract setting

II The free Bose gas in the grand canonical ensemble

III Quantum measurements

IV Weak measurements

I An abstract setting

The geometry of a statistical manifold $\mathbb{M}$ can be studied by means of the Kullback-Leibler divergence

$$D_{\mathrm{KL}}(p,q) = \sum_i p(i) \ln \frac{p(i)}{q(i)}.$$

(p and q are probability distributions)

It is called a relative entropy in Statistical Physics.

The Fisher information at a point p_θ of $\mathbb{M}$ is given by

$$I_{k,l}(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta_k} \ln p_\theta \right) \left(\frac{\partial}{\partial \theta_l} \ln p_\theta \right).$$

The Fisher information can be obtained from the KL divergence by

$$I_{k,l}(\theta) = \frac{\partial^2}{\partial\theta_k \partial\theta_l} D(p||p_\theta)\Big|_{p=p_\theta}.$$

In this expression it is not needed that p belongs to $\mathbb{M}$.

A straightforward calculation shows this.

Definition The *extended* Fisher information of a pdf p (p not necessarily in $\mathbb{M}$) is

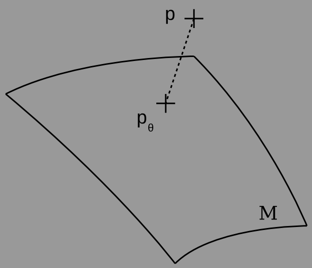
$$I_{k,l}(p) = \frac{\partial^2}{\partial\theta_k \partial\theta_l} D(p||p_\theta)\Big|_{\theta:\, p\in F_\theta}.$$

F_θ denotes the '**fiber**' of all p best fitted by p_θ.

Interestingly, the meaning of $I_{k,l}(\theta)$ becomes clearer when considering p not in $\mathbb{M}$.

p is the empirical measure, the result of an experiment.

Given p, look for p_θ which minimizes $D(p||p_\theta)$.



The divergence $D(p||p_\theta)$ is a measure for the distance between an arbitrary p and a point p_θ of the statistical manifold $\mathbb{M}$.

The inverse of the Fisher information matrix now indicates how well the model point p_θ is determined by the experimental data p.

Proposition $I_{k,l}(p)$ is covariant.

Proposition If p_θ belongs to the exponential family then $I_{k,l}(p)$ is constant on the fibre F_θ.

Proof follows from a Pythagorean relation.

- J. Naudts and B. Anthonis, The exponential family in abstract information theory, in: Geometric Science of Information, GSI 2013 LNCS proceedings, F. Nielsen and F. Barbaresco eds., (Springer, 2013), p. 265–272.
- B. Anthonis, Extension of information geometry for modelling non-statistical systems, PhD Thesis December 2014, arXiv:1501.00853.

These results hold in an abstract setting.

Abstract setting It is not needed that
p and p_θ are probability distributions.

Space $\mathbb{X}$ of possible experimental outcomes,
Space $\mathbb{M}$ manifold of model points,

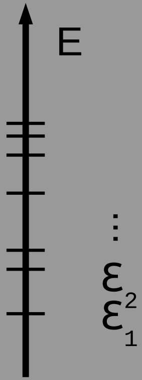
$D(x||m) : \mathbb{X} \times \mathbb{M} \to [0, +\infty]$ divergence function.

2 examples:

- Bose gas in the grand canonical ensemble
- Quantum measurements

II The free Bose gas in the grand canonical ensemble

For instance, a gas of photons.



- The energy levels are denoted $\varepsilon_1, \varepsilon_2, \cdots$.
- Each level ε_j can contain an arbitrary number n_j of photons.
- The total number $N = \sum_j n_j$ of photons is usually not known in advance.
- A measurement yields the sequence of numbers $n_1, n_j, \cdots$.
- The standard trick is to define probabilities $p_j = n_j/N$.
- Working with the empirical measure on the space of all finite sequences of integers is better, but is technically involved.
- Why not take directly $\mathbb{X} = \{n_1, n_2, \cdots, 0, 0, 0, \cdots : \ n_i \in \mathbb{N}\}$?

The manifold $\mathbb{M}$ is a two-parameter family of probability distributions

$$p_{\beta,\mu}(n) \quad = \quad \frac{1}{Z(\beta,\mu)} \exp(-\beta \sum_j \epsilon_j n_j + \beta\mu \sum_j n_j),$$

$\beta > 0$ and $\mu < \epsilon_j$ for all j, normalization given by

$$Z(\beta,\mu) = \prod_j \frac{1}{1 - \exp(-\beta(\epsilon_j - \mu))}.$$

Choose the divergence function

$$D(n||\beta,\mu) \quad = \quad \ln Z(\beta,\mu) - \sum_j n_j(-\beta\epsilon_j + \beta\mu).$$

Minimization of $\beta, \mu \to D(n||\beta, \mu)$ reproduces standard textbook results for β and μ.

The metric tensor $g(\beta, \mu)$ can be calculated.
The extended Fisher information matrix is given by

$$I_{k,l}(n) = g(\beta, \mu) \quad \text{for all } n \in F_{\beta,\mu}.$$

It indicates how well β and μ are determined by the data set n.

In the PhD thesis covariant derivatives ∇_β and ∇_μ are introduced.
A method is described to calculate connection coefficients ω^c_{ab} for which $\nabla_a \partial_b = \omega^c_{ab} \partial_c$.

III Quantum measurements

The state of a quantum system is a wave function ψ or
a density matrix ρ

- ψ is a normalized element of a Hilbert space
- ρ is a non-negative trace-class operator with trace 1
- in the finite case ρ is an Hermitean matrix with eigenvalues $p_i \geq 0$, $\sum_i p_i = 1$.
- ψ determines the density matrix $|\psi\rangle\langle\psi|$ which is an orthogonal projection onto $\mathbb{C}\psi$.

A quantum experiment measuring ψ
reveals the diagonal part of the matrix $|\psi\rangle\langle\psi|$
in a basis dictated by the experimental setup.

Example

Let $\psi = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{C}^2$.

Then $|\psi\rangle \langle\psi| = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

The result of a measurement in the basis $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

is $\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Axiom The result of a quantum measurement is always a conditional expectation.

The condition is that the resulting density matrix must be diagonal in the basis of the measurement.

A definition of a **quantum conditional expectation** is found in

D. Petz, Quantum Information Theory and Quantum Statistics (Springer, 2008)

The theory of quantum conditional expectations

- originated with the work of Accardi and Cecchini;
- relies on Tomita-Takesaki theory.

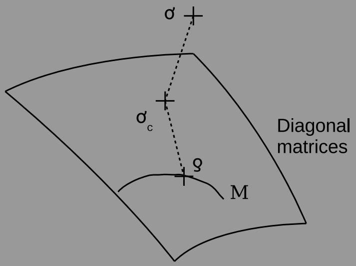
The map $|\psi\rangle\langle\psi| \to \mathrm{diag}\,|\psi\rangle\langle\psi|$ is a quantum conditional expectation in the sense of Petz.

The quantum analogue of the Kullback-Leibler divergence is given by

$$D(\sigma||\rho) \quad = \quad \mathrm{Tr}\,\sigma \ln \sigma - \mathrm{Tr}\,\sigma \ln \rho.$$

It satisfies $D(\sigma||\rho) \geq 0$, with equality if and only if $\sigma = \rho$.

Choose a model manifold $\mathbb{M}$ consisting of diagonal matrices.
Let $\sigma_c = \mathrm{diag}\,\sigma$.



Then the Pythagorean relation holds
(special case of Theorem 9.3 of Petz)

$D(\sigma||\rho) = D(\sigma||\sigma_c) + D(\sigma_c||\rho)$
for all σ, for all $\rho \in \mathbb{M}$.

Hence, minimizing $\rho \in \mathbb{M} \to D(\sigma||\rho)$ may be replaced by minimizing $\rho \in \mathbb{M} \to D(\sigma_c||\rho)$.

IV Weak measurements

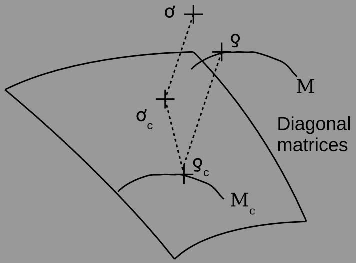
- Quantum measurements concern tiny effects
- The measurements usually destroy the state of the measured system
- This is known as **the collapse of the wave function**
- Recent experiments overcome this difficulty
  $\Rightarrow$ **weak measurements**

Theoretical support comes from

Y. Aharonov, D.Z. Albert, L. Vaidman,
Phys. Rev. Lett. 60, 1351–1354 (1988)

Our interpretation

What happens if the model manifold $\mathbb{M}$ does not consist of matrices diagonal in the basis of the experiment?



- Assume $\rho > 0$ for all $\rho \in \mathbb{M}$.
- Project $\mathbb{M}$ onto $\mathbb{M}_c = \{\rho_c : \rho \in \mathbb{M}\}$.
- Minimize $\rho \to D(\sigma_c||\rho_c)$.
- Then also $\rho \to D(\sigma||\rho_c)$ is minimized.

If now $\mathbb{M}_c$ lies in the border of $\{\rho : \rho > 0\}$ then the value of $D(\sigma||\rho_c)$ can become large and very sensitive to the experimental outcome σ.

We conjecture that this amplification effect is exploited in weak measurements.