数理解析研究所講究録 1350 巻 2004 年 39-52

Riemannian geometry of sets of probability distributions in physics *

R. Mrugała
Institute of Physics, N. Copernicus University,
Grudziądzka 5, 87-100 Toruń, Poland
(e-mail: mrugala@phys.uni.torun.pl)

Abstract

Two topics are discussed in the paper. The first one concerns information thermodynamics, in particular the maximum entropy principle. This principle provides a constructive criterion allowing one to find multiparameter statistical models based on incomplete experimental data (on partial knowledge). As the next topic there are investigated differential-geometric structures for a few statistical models. As the source of these structures serves the Kullback-Leibler information and resulting from it a Riemannian metric, equivalent to the Fisher information matrix. Scalar curvature analyzed for a few statistical models corresponding to some known physical systems leads to the hypothesis that its inverse may serve as a plausible measure of stability of these systems.

1 Introduction

Differential geometry, in particular Riemannian geometry and the theory of spaces with affine connections, found many important applications in mathematical statistics and in information theory. In mathematical statistics one is interested in finding a so-called statistical model consisting of a family of probability distribution (density) functions which depend not only on random variables but usually depend also on a number of parameters. Then to every set of numerical values of these parameters corresponds one probability density called statistical hypothesis. Riemannian distance in the parameter space is used to make comparisons between different statistical hypotheses. These geometrical methods soon found successful applications in statistical physics, and hence in thermodynamics.

Thermodynamics is a branch of physics which deals with thermal phenomena in macroscopic bodies. It uses two methods: phenomenological and statistical. The phenomenological approach is based on direct macroscopic observations and assumes continuous model of matter disregarding its microscopic structure. The statistical approach (statistical physics, statistical mechanics) is based on microscopic (discrete) model of matter and on classical or quantum mechanics, as well as on probability and information theory. The basic problem of statistical physics consists in finding probability distribution (a function in the classical case and an operator in the quantum case) which describes the collective statistical behaviour of the system and allows one to calculate mean values and fluctuations of all plausible orders of various physical quantities. Such a probability distribution may be postulated or guessed, 'derived' from the (postulated) microcanonical distribution or derived from a variational principle. By a variational principle

*Research supported in part by European Commission under the Inco-Copernicus Programme, grant ICOP-DISS-2168-96.

we mean the *maximum uncertainty principle*, or the *maximum information* (*entropy*) *principle* [1, 2]. Due to this we use the term *information thermodynamics*.

2 Information thermodynamics of classical systems

Let us consider a classical N-particle mechanical system with a configuration space Q and the corresponding phase space Γ which usually is the cotangent bundle of Q, i.e. $\Gamma = T^*Q$. Points of Γ, parameterized locally by positions q and momenta p,

$$(q,p) = (q^1, \dots, q^{3N}; p_1, \dots, p_{3N}) \in \Gamma \,, \tag{2.1}$$

are called *microscopic states* or *classical pure states*.

Dynamical evolution of the system is given by $6N$ Hamilton equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}\,, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}\,, \qquad i = 1, \dots, 3N, \tag{2.2}$$

where $H = H(q,p,t)$ is the Hamiltonian of the system and $t \in \mathbb{R}$ denotes time. Physical quantities $F_i(q,p,t)$ (*observables, dynamical variables*) are real-valued functions on $\Gamma \times \mathbb{R}$. For large N it is not possible to control all q's and p's, instead one is able to take into account only a small number of dynamical variables F_i, and measurements give mean values m_i of them. Thus one is forced to use statistical methods.

Statistical mechanics makes predictions about systems which are based on partial information obtained from incomplete measurements. By an incomplete measurement we mean a measurement giving only a few values of m_i. The central notion of classical statistical mechanics is the notion of probability (density) distribution $\rho\,(q,p,t)$ such that

$$\rho\,(q,p,t) \geq 0, \qquad \int_\Gamma \rho\,(q,p,t)\, d\Gamma = 1\,, \qquad d\Gamma = dq\,dp = dq^1 \dots dq^{3N} dp_1 \dots dp_{3N}\,; \tag{2.3}$$

ρ is called a *statistical state* or *probabilistic measure* on Γ.

The time evolution of ρ is given by Liouville's equation

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0\,, \tag{2.4}$$

where $\{\rho, H\}$ is the Poisson bracket of ρ and H. If $\rho = \rho(q,p)$, $H = H(q,p)$, then we say that the system is in statistical equilibrium and then obviously $\{\rho\,(q,p), H(q,p)\} = 0$.

Let us consider now an equilibrium system for which we are interested in r stochastic variables represented by r functions $F_1(q,p), \dots, F_r(q,p)$. Their mean values m_i are defined formally as

$$m_i = \langle F_i \rangle = \int_\Gamma \rho\,(q,p)\, F_i(q,p)\, d\Gamma\,, \qquad i = 1, \dots, r. \tag{2.5}$$

The problem is that $\rho\,(q,p)$ is usually not known. Instead, as was said above, from direct measurements we know only mean values m_i corresponding to these stochastic variables F_i. This partial knowledge does not allow for a unique reconstruction of $\rho\,(q,p)$ without additional assumptions. Therefore we can only define a set of probability distributions, the so-called *macrostate* $K_{F_1 \dots F_r}$ (with respect to $F_1, \dots, F_r$),

$$K_{F_1 \dots F_r} = \left\{ \rho\,(q,p)\,; \qquad \int_\Gamma \rho\,(q,p)\, F_i(q,p)\, d\Gamma = m_i\,, \quad i = 1, \dots, r \right\}. \tag{2.6}$$

The macrostate $K_{F_1 \dots F_r}$ is a convex set because if $\rho_1, \rho_2 \in K_{F_1 \dots F_r}$, then also their convex combination

$$\rho = \alpha_1 \rho_1 + \alpha_2 \rho_2 \,, \qquad \alpha_1 \geq 0, \quad \alpha_2 \geq 0, \quad \alpha_1 + \alpha_2 = 1 \,, \tag{2.7}$$

gives the same mean values m_i since

$$m_i(\alpha_1 \rho_1 + \alpha_2 \rho_2) = \alpha_1 m_i(\rho_1) + \alpha_2 m_i(\rho_2) = (\alpha_1 + \alpha_2) m_i = m_i \,. \tag{2.8}$$

The set $F_1, \dots, F_r$ has to be *thermodynamically regular*, i.e. F_i must be such that [3]:

- $F_1, \dots, F_r$ and $F_0 = 1$ are linearly independent,
- there exist real numbers $\theta^1, \dots, \theta^r$ such that

$$\int\limits_{\Gamma} \exp\left[-\sum_{i=1}^{r} \theta^i F_i(q,p) \right] d\Gamma < \infty \,.$$

The entropy of the macrostate $K_{F_1 \dots F_r}$ is defined as

$$S(K_{F_1 \dots F_r}) = \sup \left\{ S(\rho) \, ; \quad \rho \in K_{F_1 \dots F_r} \right\} \,, \tag{2.9}$$

where

$$S(\rho) = -k \int\limits_{\Gamma} \rho\,(q,p) \, \ln \rho\,(q,p) \, d\Gamma \,, \qquad k = 1.38 \times 10^{-23} J \cdot K^{-1} \,. \tag{2.10}$$

The above formula for entropy is the same as the formula for information and differs only by the multiplicative Boltzmann constant k which establishes the physical unit of entropy.

By a *representative density* (statistical state) we call

$$\rho^* \in K_{F_1 \dots F_r} \,, \quad \text{such that} \quad S(\rho^*) = S(K_{F_1 \dots F_r}) \,. \tag{2.11}$$

The representative state ρ^* is unique because of convexity of the macrostate. Thus we have come to the maximum information principle [1, 2].

Maximum information (uncertainty, entropy) principle: For a macrostate $K_{F_1 \dots F_r}$ generated by a thermodynamically regular set of r physical quantities $F_1, \dots, F_r$, the representative distribution ρ^* is the best probability measure compatible with the available (measured) data.

As a matter of fact this principle stands for a direct extension of Laplace's principle of insufficient reason.

To find explicit form of ρ^* for a thermodynamically regular macrostate $K_{F_1 \dots F_r}$ we look for the maximum of the entropy functional (2.10) subject to the constraints

$$\int\limits_{\Gamma} \rho\,(q,p) \, d\Gamma = 1, \qquad \int\limits_{\Gamma} \rho\,(q,p) \, F_i(q,p) \, d\Gamma = m_i, \quad i = 1, \dots, r \,. \tag{2.12}$$

To do this we look for the maximum of an unconstrained functional

$$L(\rho) = -k \int\limits_{\Gamma} \rho \, \ln \rho \, d\Gamma - k w' \int\limits_{\Gamma} \rho \, d\Gamma - k \sum_{i=1}^{r} \theta^i \int\limits_{\Gamma} \rho \, F_i \, d\Gamma \,, \tag{2.13}$$

where $w', \theta^1, \dots, \theta^r$ are unknown Lagrange multipliers. We obtain

$$\rho^*(q,p) = \exp\left[-1 - w' - \sum_{i=1}^{r} \theta^i F_i(q,p)\right] \equiv Z^{-1}(\theta^1, \dots, \theta^r) \exp\left[-\sum_i \theta^i F_i(q,p)\right], \tag{2.14}$$

where

$$Z^{-1} = \exp[-1 - w'] = \exp[-w], \qquad Z(\theta^1, \dots, \theta^r) = \int_\Gamma \exp\left[-\sum_i \theta^i F_i(q,p)\right] d\Gamma, \tag{2.15}$$

is the normalization factor (in physics called the partition function, or the sum of states) and $\theta^1, \dots, \theta^r$ are the so-called *statistical temperatures*. They are uniquely determined by the measured m_i by the formulae

$$m_i = -\frac{\partial \ln Z(\theta^1, \dots, \theta^r)}{\partial \theta^i}, \qquad i = 1, \dots, r. \tag{2.16}$$

The entropy of the macrostate $K_{F_1 \dots F_r}$ is then

$$S(K_{F_1 \dots F_r}) = S(\rho^*) = k \ln Z(\theta^1, \dots, \theta^r) + k \sum_{i=1}^{r} \theta^i \langle F_i \rangle. \tag{2.17}$$

From the above and from the rest of this paper we will see that of particular interest for thermodynamics and geometry is the partition function $Z(\theta^1, \dots, \theta^r)$. Having $Z(\theta^1, \dots, \theta^r)$ one is in principle able to calculate all the required thermodynamic and geometrical quantities by the differentiation of Z and by elementary algebraic operations on Z and its derivatives.

From now on we will restrict ourselves to representative distributions only and thus we shall drop the asterisk *. The best known example of ρ for $r = 1$ is the canonical (Gibbs) distribution generated by only one stochastic variable, by the Hamiltonian $F_1 = H(q,p)$, which in the standard notation [4, 5] reads as

$$\rho(q,p) = Z^{-1}(\beta) \exp[-\beta H(q,p)], \qquad Z(\beta) = \int_\Gamma \exp[-\beta H(q,p)]\, d\Gamma, \tag{2.18}$$

where

$$\theta^1 = \beta = (kT)^{-1}, \qquad d\Gamma = \frac{1}{N!\, h^{3N}} dq^1 \dots dq^{3N} dp_1 \dots dp_{3N}, \qquad h = 6.626 \times 10^{-34} J \cdot s; \tag{2.19}$$

$N!$ takes account of the fact that particles are indistinguishable and h is the Planck constant. For this ρ one obtains

$$U = \langle H \rangle = -\frac{\partial \ln Z(\beta)}{\partial \beta}, \qquad S(\rho) = k \ln Z(\beta) + k\beta U. \tag{2.20}$$

The canonical distribution will be not used in this paper because the parameter space is here 1-dimensional and thus is not interesting from the point of view of geometry.

Another example is the *grand canonical distribution* generated by two stochastic variables, by the N-particle Hamiltonian H_N and by the number of particles N,

$$\rho(q,p) = Z^{-1}(T,\mu) \exp\left[\frac{-H_N(q,p) + \mu N}{kT}\right], \qquad \theta^1 = \frac{1}{kT}, \quad \theta^2 = -\frac{\mu}{kT}, \tag{2.21}$$

The grand canonical distribution is used for open systems for which the number of particles N is not fixed but the chemical potential μ can be controlled. Here

$$Z(T,\mu) = \sum_{N=0}^{\infty} \int_{\Gamma_N} \exp\left[\frac{-H_N(q,p) + \mu N}{kT}\right] d\Gamma_N\,, \tag{2.22}$$

and

$$U = \langle H_N\rangle = -\frac{\partial \ln Z}{\partial \theta^1} = kT^2\,\frac{\partial \ln Z}{\partial T}\,, \qquad \langle N\rangle = -\frac{\partial \ln Z}{\partial \theta^2} = \frac{\partial \ln Z}{\partial(\mu/kT)}\,. \tag{2.23}$$

The third example is the so-called *Boguslavski* or *isobaric-isothermal* or $P-T$ distribution generated by H and the volume V,

$$\rho(q,p) = Z^{-1}(T,P)\exp\left[\frac{-H(q,p) - PV}{kT}\right]\,, \qquad \theta^1 = \frac{1}{kT}\,, \qquad \theta^2 = \frac{P}{kT}\,. \tag{2.24}$$

This distribution is used for systems for which volume V is not fixed (it can fluctuate) but the pressure P is controlled.

The mentioned before *microcanonical distribution* [4, 5]

$$\rho(q,p) = W^{-1}\,\delta\big(H(q,p) - E\big)\,, \tag{2.25}$$

where δ is the Dirac δ-function (distribution) is in this scheme derived from the functional

$$L(\rho) = -k\int_\Gamma \rho\,\ln\rho\, d\Gamma - kw\int_\Gamma \rho\, d\Gamma\,. \tag{2.26}$$

For this distribution

$$S = k\ln W\,, \tag{2.27}$$

and the normalization factor W denotes the number of microstates (elementary cells) in Γ for which $H(q,p) = E$. We will not discuss this case.

3 Information thermodynamics of quantum systems

The state space of a quantum system is a Hilbert space $\mathcal{H}$, in most cases infinite dimensional, whose elements (vectors) $|\psi\rangle \in \mathcal{H}$ are called wave functions or *pure states*; we assume that they are orthonormal, i.e.

$$\langle\psi_k|\psi_i\rangle \equiv \int \overline{\psi_k}(x)\psi_i(x)\,dx = \delta_{ki}\,. \tag{3.1}$$

($\overline{\psi_k}$ is complex conjugate of ψ_k). Evolution of pure states is given by the Schrödinger equation

$$i\hbar\,\frac{\partial}{\partial t}\,|\psi\rangle = H|\psi\rangle\,, \qquad \hbar = h/2\pi, \tag{3.2}$$

where H is the Hamilton operator of the considered system.

Usually quantum system is not in a pure state but rather in a *mixed state* (also called density operator, density matrix or quantum statistical state) [5]

$$\rho = \sum_i w_i P_i = \sum_i w_i |\psi_i\rangle\langle\psi_i|, \tag{3.3}$$

where $P_i = |\psi_i\rangle\langle\psi_i|$ denotes the projection operator onto the one-dimensional subspace of $\mathcal{H}$ spanned by $|\psi_i\rangle$ and w_i means the probability that the actual state is $|\psi_i\rangle$. It is required that the set of all pure states $|\psi_i\rangle$ is complete and that they are mutually orthonormal.

Because of $\mathrm{Tr}\, P_i = 1$, and hence $\mathrm{Tr}\, \rho := \sum_i \rho_{ii} = \sum_i w_i = 1$, we have

$$\rho \geq 0\,, \qquad \rho^\dagger = \rho\,, \qquad \mathrm{Tr}\, \rho = 1\,. \tag{3.4}$$

The time evolution of ρ is given by the von Neumann equation

$$i\hbar \frac{\partial \rho}{\partial t} + [\rho, H] = 0\,, \qquad [\rho, H] = \rho H - H\rho\,. \tag{3.5}$$

Entropy of mixed states (von Neumann (1927)) is defined as

$$S(\rho) := -k\; \mathrm{Tr}\, (\rho \ln \rho)\; (= -k \sum_i w_i \ln w_i). \tag{3.6}$$

Let us now go over to quantum statistical physics. To this end let us consider r physical quantities (observables) which are by assumption self-adjoint linear operators F_i on $\mathcal{H}$. From measurements we know mean values m_i of F_i which are related to the (in principle unknown) density matrix ρ by the formulae

$$m_i \equiv \langle F_i \rangle = \mathrm{Tr}\, (\rho F_i)\,, \qquad i = 1, \ldots, r\,. \tag{3.7}$$

The quantum macrostate is defined analogously to (2.6) as

$$K_{F_1 \ldots F_r} = \{\rho; \quad \mathrm{Tr}\, (\rho F_i) = m_i, \qquad i = 1, \ldots, r\}\,, \tag{3.8}$$

and the representative density matrix (also unique) takes the form

$$\rho = Z^{-1}(\theta^1, \ldots, \theta^r)\, \exp\left[-\sum_{i=1}^{r} \theta^i F_i\right]\,, \tag{3.9}$$

with the partition function

$$Z(\theta^1, \ldots, \theta^r) = \mathrm{Tr}\, \exp\left[-\sum_{i=1}^{r} \theta^i F_i\right] \tag{3.10}$$

and

$$m_i = -\frac{\partial \ln Z(\theta^1, \ldots, \theta^r)}{\partial \theta^i}\,, \qquad S(\rho) = k \ln Z + k \sum_{i=1}^{r} \theta^i \langle F_i \rangle\,. \tag{3.11}$$

The representative density matrix for arbitrary r shall be called a *generalized* (*Gibbs*) *density matrix*. Its special cases are the canonical, grand canonical, and the Boguslavski counterparts of the classical distributions.

The microcanonical density matrix is also analogously given by

$$\rho = \sum_{i=1}^{W} \frac{1}{W} |\psi_i\rangle\langle\psi_i|\,, \quad \text{where} \quad H|\psi_i\rangle = E|\psi_i\rangle\,, \quad E\text{–fixed.}$$

4 Riemannian structure of a family of probability distributions

Let us consider a parameterized family $\mathcal{S}^r = \{\rho(y,\theta)\}$ of probability distributions, where y denotes now all microscopic random variables (phase space variables for classical systems) and $\theta = (\theta^1, \ldots, \theta^r)$ denote thermodynamic parameters. Such a family is called a statistical or probabilistic model [6].

Let $\rho(y,\theta)$ and $\sigma(y,\theta)$ be two probability distributions from $\mathcal{S}^r$ and let they be mutually absolutely continuous. The *relative information* [3] of ρ and σ is defined as

$$I(\rho|\sigma) = \int \rho(y,\theta)\big(\ln\rho(y,\theta) - \ln\sigma(y,\theta)\big)\,dy = \int \rho(y,\theta)\ln\frac{\rho(y,\theta)}{\sigma(y,\theta)}\,dy\,. \tag{4.1}$$

The other names for $I(\rho|\sigma)$ are *Kullback information*, *Rényi-Kullback information*, *information gain*, *directed divergence* or *directed distance*. In general it is not symmetric, i.e. $I(\rho|\sigma) \neq I(\sigma|\rho)$ but

$$I(\rho|\sigma) \geq 0\,, \qquad I(\rho|\sigma) = 0 \quad \text{iff} \quad \rho(y,\theta) = \sigma(y,\theta). \tag{4.2}$$

The symmetrized counterpart of $I(f|g)$ is defined as

$$J(\rho,\sigma) = I(\rho|\sigma) + I(\sigma|\rho) = \int (\rho - \sigma)\,\ln\frac{\rho}{\sigma}\,dy \tag{4.3}$$

and is called *divergence*, *information distance* or again *information gain*.

Parallel to $\mathcal{S}^r$ let us introduce an r-dimensional parameter space $\mathcal{P}^r$, where $\mathcal{P}^r$ is a collection of all points $\mathcal{P}^r = \{\theta\} = \{(\theta^1, \ldots, \theta^r)\}$. By assumption $\mathcal{P}^r$ is an r-dimensional differentiable manifold with $\theta = (\theta^1, \ldots, \theta^r)$ playing the role of local coordinates. Usually $\mathcal{P}^r$ will be a subset of $\mathbb{R}^r$. In this paper, statistical model will be given by a family of probability distributions $\rho(y,\theta)$ having the same functional form but differing by the numerical values of the parameters θ. Therefore we will have a 1–1 correspondence between points of $\mathcal{S}^r$ and $\mathcal{P}^r$ and thus any geometrical structure defined on $\mathcal{S}^r$ can be considered as a structure on $\mathcal{P}^r$. This is important for physics because it allows to endow spaces of directly measurable parameters with geometrical structures. For instance, ρ and σ will be treated as two different points of $\mathcal{S}^r$ and at the same time as two different points of $\mathcal{P}^r$. With every tangent space $T_\theta\mathcal{P}$ will be associated a corresponding tangent space $T_\theta\mathcal{S}$ in such a way that every tangent basis vector $\partial_i \equiv \partial/\partial\theta^i$ in $T_\theta\mathcal{P}$ will have a representative $\partial_i\ell \equiv \partial_i\ell(y,\theta) \equiv \partial\ell(y,\theta)/\partial\theta^i$ in $T_\theta\mathcal{S}$, where $\ell(y,\theta) = \ln\rho(y,\theta)$, see [6, 7].

To define a Riemannian structure on $\mathcal{S}^r$, and thus on $\mathcal{P}^r$, let us take two adjacent points in $\mathcal{S}^r$, $\rho(y,\theta)$ and $\sigma(y,\theta+\Delta\theta) \equiv \rho(y,\theta+\Delta\theta)$. Moreover, let us assume that $\rho(y,\theta)$ satisfy the following regularity conditions [3]:

- the partial derivatives $\partial\ln\rho/\partial\theta^i$, $\partial^2\ln\rho/\partial\theta^i\partial\theta^j$ and $\partial^3\ln\rho/\partial\theta^i\partial\theta^j\partial\theta^k$ exist in all points of the intervals $A^i = (\theta^i, \theta^i + \Delta\theta^i)$, for all $i,j,k = 1,2,\ldots,r$,
- for arbitrary $\theta^i \in A^i$ and for all i,j,k one has
$$\left|\frac{\partial\rho}{\partial\theta^i}\right| < F(y)\,, \qquad \left|\frac{\partial^2\rho}{\partial\theta^i\partial\theta^j}\right| < G(y)\,, \qquad \left|\frac{\partial^3\ln\rho}{\partial\theta^i\partial\theta^j\partial\theta^k}\right| < H(y)\,,$$
where $F(y)$ and $G(y)$ are some functions integrable on the whole space of variables y, and $\int\rho(y,\theta)H(y)\,dy < M$, with M not depending on θ,
- $\displaystyle\int\frac{\partial\rho}{\partial\theta^i}\,dy = 0, \qquad \int\frac{\partial^2\rho}{\partial\theta^i\partial\theta^j}\,dy = 0$ (from normalization of ρ).

To simplify the notation let us write

$$\rho(y,\theta)=\rho(\theta), \qquad \sigma(y,\theta+\Delta\theta)\equiv\rho(y,\theta+\Delta\theta)=\rho(\theta+\Delta\theta) \tag{4.4}$$

and

$$J\left(\rho(\theta),\rho(\theta+\Delta\theta)\right)=J(\theta,\theta+\Delta\theta)\,. \tag{4.5}$$

Then using the properties of J inherited from those of I, and the above regularity conditions, after expanding $J(\theta,\theta+\Delta\theta)$ into power series with respect to θ (to the second order) we get

$$J(\theta,\theta+\Delta\theta)=g_{ij}(\theta)\Delta\theta^i\Delta\theta^j\,, \tag{4.6}$$

where

$$g_{ij}(\theta)=\langle\partial_i\ell\,\partial_j\ell\rangle\equiv\left\langle\frac{\partial\ln\rho}{\partial\theta^i}\,\frac{\partial\ln\rho}{\partial\theta^j}\right\rangle=\int\rho(y,\theta)\,\frac{\partial\ln\rho(y,\theta)}{\partial\theta^i}\,\frac{\partial\ln\rho(y,\theta)}{\partial\theta^j}\,dy \tag{4.7}$$

and $\langle\,\rangle$ again means expectation value. It can be shown that also

$$I(\theta|\theta+\Delta\theta)=\frac{1}{2}\,g_{ij}\Delta\theta^i\Delta\theta^j\,. \tag{4.8}$$

From the construction above it is obvious that $g_{ij}(\theta)$ are components of a covariant tensor and that the quadratic form $dl^2=g_{ij}(\theta)d\theta^i d\theta^j$ is symmetric and positive definite. Thus we see that $g_{ij}(\theta)$ treated as a Riemann metric tensor on $\mathcal{S}^r$ is equivalent to the Fisher information matrix [8, 6] for $\mathcal{S}^r$. Actually it was Rao [9] who first proposed to interpret dl^2 as the metric form on $\mathcal{S}^r$. To find a distance between two points of $\mathcal{S}^r$, i.e. between two statistical hypothesis, he integrated dl along a geodesics passing through these two points.

5 Riemannian geometry of a set of classical generalized Gibbs distributions and of the associated parameter spaces

As a statistical model $\mathcal{S}^r$ of a classical physical system let us consider now a family of generalized r-parameter Gibbs distribution functions

$$\rho(y,\theta)=Z^{-1}(\theta)\,e^{-\theta^iF_i(y)}, \qquad i=1,\dots,r\,. \tag{5.1}$$

Associated to $\mathcal{S}^r$ is a parameter space $\mathcal{P}^r$. Again $F_i:\Gamma\longrightarrow\mathbb{R}^1$ are linearly (but not statistically) independent stochastic variables and $\theta=(\theta^1,\dots,\theta^r)$ are macroscopic parameters (statistical temperatures). The normalization factor $Z(\theta)$ is

$$Z(\theta)=\int_\Gamma e^{-\theta^iF_i(y)}\,dy \tag{5.2}$$

and

$$m_i=\langle F_i\rangle=\int_\Gamma\rho(y,\theta)\,F_i(y)\,dy=-\frac{\partial\ln Z}{\partial\theta^i}\,. \tag{5.3}$$

According to (4.6) and (4.7) the square infinitesimal distance on $\mathcal{P}^r$ is

$$dl^2=J(\theta+d\theta|\theta)=2I(\theta+d\theta|\theta)=g_{ij}(\theta)d\theta^id\theta^j\,, \tag{5.4}$$

with

$$
\begin{aligned}
g_{ij}(\theta) &= \frac{\partial^2 I}{\partial\theta^i\partial\theta^j} = \left\langle \frac{\partial \ln\rho}{\partial\theta^i}\frac{\partial \ln\rho}{\partial\theta^j} \right\rangle = -\left\langle \frac{\partial^2 \ln\rho}{\partial\theta^i\partial\theta^j} \right\rangle = \frac{\partial^2 \ln Z(\theta)}{\partial\theta^i\partial\theta^j} \\
&= -\frac{\partial m_i}{\partial\theta^j} = -\frac{\partial m_j}{\partial\theta^i} = \langle (F_i(y)-m_i)(F_j(y)-m_j)\rangle \, .
\end{aligned}
\tag{5.5}
$$

From the last term we see that the components g_{ij} are actually given by covariances of the stochastic variables F_i and F_j and thus have an obvious physical interpretation in terms of the theory of fluctuations.

The Christoffel symbols

$$
\Gamma_{ijk} = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial\theta^j} + \frac{\partial g_{ij}}{\partial\theta^k} - \frac{\partial g_{jk}}{\partial\theta^i}\right)
\tag{5.6}
$$

reduce for our metric to

$$
\Gamma_{ijk} = \frac{1}{2}\frac{\partial g_{ij}}{\partial\theta^k} = \frac{1}{2}\frac{\partial^3 \ln Z}{\partial\theta^i\partial\theta^j\partial\theta^k} = -\frac{1}{2}\langle (F_i-m_i)(F_j-m_j)(F_k-m_k)\rangle \, .
\tag{5.7}
$$

The components of the curvature tensor

$$
R_{ijkl} = \frac{1}{2}\left[\frac{\partial^2 g_{jk}}{\partial\theta^i\partial\theta^l} - \frac{\partial^2 g_{ik}}{\partial\theta^j\partial\theta^l} + \frac{\partial^2 g_{il}}{\partial\theta^j\partial\theta^k} - \frac{\partial^2 g_{jl}}{\partial\theta^i\partial\theta^k}\right] + g^{mn}(\Gamma_{mil}\Gamma_{njk} - \Gamma_{mik}\Gamma_{njl})
\tag{5.8}
$$

($g^{ij}g_{jk} = \delta^i_k$) reduce subsequently to

$$
R_{ijkl} = g^{mn}(\Gamma_{mil}\Gamma_{njk} - \Gamma_{mik}\Gamma_{njl}) \, .
\tag{5.9}
$$

One sees that R_{ijkl} are functions of the second and third derivatives of $\ln Z$ and thus the functions of the second and third moments of F_i. Fortunately, due to the special for of g_{ij}, they do not depend on the fourth derivatives of $\ln Z$ which are not equivalent to the fourth moments of F_i. Because the curvature tensor has many symmetries, the number of its independent components is equal to $\frac{1}{12}r^2(r^2-1)$. Therefore, for $r=2$ it has only one independent component, say R_{1212}.

The scalar curvature $R = g^{jk}R^i{}_{jik}$ for $r = 2$ can be expressed in a simple form, very convenient for numerical calculations,

$$
R = \frac{2}{g}R_{1212} = \frac{-2}{g^2}\begin{vmatrix} g_{11} & g_{12} & g_{22} \\ \frac{\partial g_{11}}{\partial\theta^1} & \frac{\partial g_{12}}{\partial\theta^1} & \frac{\partial g_{22}}{\partial\theta^1} \\ \frac{\partial g_{11}}{\partial\theta^2} & \frac{\partial g_{12}}{\partial\theta^2} & \frac{\partial g_{22}}{\partial\theta^2} \end{vmatrix} \, , \qquad g = \det(g_{ij}).
\tag{5.10}
$$

6 Examples

In this section we shall illustrate the general scheme developed above. To have a meaningful Riemannian geometry we need at least two parameters, i.e. we need $r \geq 2$. The simplest example is the ideal gas composed of N identical particles of mass m for which Hamiltonian H has the form

$$
H = \sum_{i=1}^{N} \frac{{}^{2}_{i}}{2m} \, ,
\tag{6.1}
$$

and we use the Boguslavski (the $P-T$, or pressure–temperature) distribution generated by H and the volume V,

$$\rho(q,p;\alpha,\beta)=Z^{-1}(\alpha,\beta)e^{-\beta H(q,p)-\alpha V}, \qquad \theta^1=\beta=\frac{1}{kT}, \quad \theta^2=\alpha=\frac{P}{kT}. \tag{6.2}$$

The partition function $Z(\alpha,\beta)$ is [10]

$$Z(\alpha,\beta)=\frac{1}{N!h^{3N}}\int_0^\infty dV\, e^{-\alpha V}\int e^{-\beta H}dq\,dp=\left(\frac{2\pi m}{h^2\beta}\right)^{3N/2}\alpha^{-(N+1)}, \tag{6.3}$$

and for large N, $\ln Z=(3N/2)\ln(2\pi m/h^2\beta)-N\ln\alpha$. Consequently (for $h=1$)

$$U=\langle H\rangle=-\frac{\partial\ln Z}{\partial\beta}=\frac{3}{2}N\beta^{-1}, \qquad \langle V\rangle=-\frac{\partial\ln Z}{\partial\alpha}=N\alpha^{-1}, \tag{6.4}$$

and

$$\begin{aligned}
g_{11} &= \frac{\partial^2\ln Z}{\partial\beta^2}=\langle H^2\rangle-\langle H\rangle^2=\frac{3}{2}N\beta^{-2},\\
g_{12} &= \frac{\partial^2\ln Z}{\partial\beta\partial\alpha}=\langle HV\rangle-\langle H\rangle\langle V\rangle=0,\\
g_{22} &= \frac{\partial^2\ln Z}{\partial\alpha^2}=\langle V^2\rangle-\langle V\rangle^2=N\alpha^{-2}.
\end{aligned} \tag{6.5}$$

The coordinates α, β for $\mathcal{P}^2$ are orthogonal. The scalar curvature R calculated according to (5.10) is equal to zero. This can be also seen if the metric form

$$dl^2=\frac{3N}{2}\beta^{-2}d\beta^2+N\alpha^{-2}d\alpha^2 \tag{6.6}$$

we express in the new logarithmic coordinates

$$\beta'=\sqrt{3N/2}\ln\beta, \qquad \alpha'=\sqrt{N}\ln\alpha, \tag{6.7}$$

in which $dl^2=(d\beta')^2+(d\alpha')^2$ takes the Euclidean form. Then geodesics are straight lines; physically they represent politropic processes. Ideal gas is the only known case for which $R=0$. This can be related to the fact that this is the only case for which there are no phase transitions and all states are stable.

One of the simplest classical systems of interacting particles is the so-called van der Waals gas for which the Hamiltonian has the form

$$H=\sum_{i=1}^{N}\frac{{}^2_i}{2m}+\sum_{i,j}\left((d/r_{ij})^{12}-(d/r_{ij})^6\right), \tag{6.8}$$

where r_{ij} denotes the distance between two particles ($i\neq j$) and d is a constant. By changing the second part of H responsible for interactions between particles we can have different models of gases. It appears that for all such systems R is positive for all stable states and diverges to infinity along the so-called spinodal line and at the critical point [10, 11]. It is not altogether clear what is the physical meaning of the curvature $R(\theta)$. However, these examples suggest that R^{-1} can be treated as a measure of thermodynamic stability of the systems.

Next we shall consider one of the simplest quantum systems by which we mean here a 1D (one-dimensional) Ising model (a magnetic system). For quantum systems the information gain is defined as

$$I(\rho|\sigma) = \mathrm{Tr}\,[\rho(\ln\rho - \ln\sigma)]\,. \tag{6.9}$$

However, in the quantum case one has to be cautious because not all expressions in (5.5) are really equal. Also, if F_i do not commute then to calculate information distance one has to use the Wilcox formula [12]

$$\frac{de^{A(s)}}{ds} = \int_0^1 e^{(1-\lambda)A}\,\frac{dA}{ds}\,e^{\lambda A}\,d\lambda = \int_0^1 e^{\lambda A}\,\frac{dA}{ds}\,e^{(1-\lambda)A}\,d\lambda \tag{6.10}$$

for parameter differentiation of exponential operators. Using this formula one obtains [13]

$$\begin{aligned}\frac{1}{2}\frac{\partial^2 J}{\partial\theta^i\partial\theta^j} &= \int_0^1 d\lambda\mathrm{Tr}\left[\rho\,e^{\lambda\theta^l F_l}\,\frac{\partial\ln\rho}{\partial\theta^i}\,e^{-\lambda\theta^l F_l}\,\frac{\partial\ln\rho}{\partial\theta^j}\right] \\ &= \int_0^1 d\lambda\mathrm{Tr}\left[\rho\,e^{\lambda\theta^l F_l}\,(F_i - \langle F_i\rangle)\,e^{-\lambda\theta^l F_l}\,(F_j - \langle F_j\rangle)\right].\end{aligned} \tag{6.11}$$

and further

$$\frac{1}{2}\frac{\partial^2 J}{\partial\theta^i\partial\theta^j} = -\left\langle\frac{\partial^2\ln\rho}{\partial\theta^i\partial\theta^j}\right\rangle = \frac{\partial^2\ln Z}{\partial\theta^i\partial\theta^j}\,,\qquad Z(\theta) = \mathrm{Tr}\,e^{-\theta^i F_i}. \tag{6.12}$$

In the quantum case the components of the metric tensor are, analogously to the classical case, given by covariances of two operators, F_i and F_j, where the covariances are now equal to

$$\mathrm{cov}\,(F_i, F_j) = \int_0^1 d\lambda\mathrm{Tr}\left[\rho\,e^{\lambda\theta^l F_l}\,(F_i - \langle F_i\rangle)\,e^{-\lambda\theta^l F_l}\,(F_j - \langle F_j\rangle)\right]. \tag{6.13}$$

It is reasonable to call these expressions covariances because for commuting F_i they reduce to

$$\mathrm{cov}\,(F_i, F_j) = \langle(F_i - \langle F_i\rangle)\,(F_j - \langle F_j\rangle)\rangle\,. \tag{6.14}$$

To compute the Christoffel symbols and the curvature tensor we need the third order derivatives of $\ln Z$. Using once more the Wilcox formula we obtain a complicated expression

$$\begin{aligned}\frac{\partial^3\ln Z}{\partial\theta^k\partial\theta^j\partial\theta^i} &= -\int_0^1 d\lambda\mathrm{Tr}\left\{\rho\left(\left(\int_0^1 d\mu e^{-\mu A}\,(F_k - \langle F_k\rangle)\,e^{\mu A}e^{-\lambda A}\,(F_i - \langle F_i\rangle)\,e^{\lambda A}\,(F_j - \langle F_j\rangle)\right.\right.\right. \\ &\left.\left.\left.+\int_0^\lambda d\mu\left[e^{-\lambda A}\,(F_i - \langle F_i\rangle)\,e^{\lambda A}, e^{-\mu A}\,(F_k - \langle F_k\rangle)\,e^{\mu A}\right](F_j - \langle F_j\rangle)\right)\right)\right\},\end{aligned} \tag{6.15}$$

where $A = -\theta^l F_l$ and $[\,,\,]$ again denotes commutator. These are the third moments because for commuting F_i they reduce to

$$\frac{\partial^3\ln Z}{\partial\theta^i\partial\theta^j\partial\theta^k} = -\,\langle(F_i - \langle F_i\rangle)\,(F_j - \langle F_j\rangle)\,(F_k - \langle F_k\rangle)\rangle\,. \tag{6.16}$$

Let us consider now a 1D Ising model (spin lattice) [5, 14] composed on N spins s_i in an external magnetic field h. Let $s_i = 1$ or $s_i = -1$, $i = 1, \ldots, N$, and as usually let the spin chain be closed, i.e. $s_{N+i} = s_i$. The Hamiltonian of this system is equal to

$$H = -J\sum_{k=1}^{N} s_k s_{k+1} - h\sum_{k=1}^{N} s_k = F_1 + hF_2\,, \tag{6.17}$$

where J is the coupling constant characterizing interaction of the nearest-neighbour spins. The term hF_2 represents interaction of spins with the external magnetic field h and its mean value is equal to $-Mh$, where M denotes magnetization of the lattice. The probability distribution for this system is

$$\rho = Z^{-1}(\beta, \beta h)e^{-\beta H} = Z^{-1}(\beta, \beta h)e^{-\beta F_1 - \beta h F_2}, \qquad \theta^1 = \beta = \frac{1}{kT}, \quad \theta^2 = \beta h = \frac{h}{kT}. \quad (6.18)$$

In the limit of large N one obtains [14]

$$Z(\beta, \beta h) = e^{N\beta J}\left\{\cosh(\beta h) + \left[\cosh^2(\beta h) - 2e^{-2\beta J}\sinh(2\beta J)\right]^{1/2}\right\}^N. \quad (6.19)$$

Using the formulae from previous sections for the scalar curvature we receive (after dividing by N) [13]

$$R = A^{-1}\cosh y + 1 = \cosh y \left(\sinh^2 y + e^{-4x}\right)^{-1/2} + 1, \quad (6.20)$$

where

$$x = \beta J, \qquad y = \beta h, \qquad A(x, y) = \left(\sinh^2 y + e^{-4x}\right)^{1/2}. \quad (6.21)$$

One immediately sees that R is a positive function of x and y. Moreover, it is symmetric in y which means that it is independent of the orientation of h. However, R behaves differently for positive x ($J > 0$, ferromagnetism) and negative x ($J < 0$, antiferromagnetism).

For physics it is interesting to analyze the behaviour of R with respect to the temperature T and the external field h because these parameters can be controlled easily. For instance, for T finite and $h \to \infty$, we have $R \to \coth y + 1 \to 2$, whereas for T finite and $h \to 0$, one has $R \to e^{2x} + 1 = e^{2J/kT} + 1$. If $h \to 0$ and $T \to 0$ then also $R \to e^{2x} + 1 = e^{2J/kT} + 1$. In turn, if $T \to \infty$, then $R \to 2$ for either sign of J. On the other hand, if $h = 0$ and $T \to 0$, then $R \to \infty$ for $J > 0$, and $R \to 1$ for $J < 0$. For a perfect paramagnetic substance $J = 0$, so $R \to 2$ for $h \to 0$. Therefore $R - 2$ can be taken as a joint measure of fluctuations caused by interacting spins, i.e. $R - 2$ measures deviation of the considered system from ideal paramagnetism.

7 Concluding remarks and possible generalizations

The methods of differential geometry are well established in mathematical statistics. They are not so well rooted in the formalism of statistical physics and in thermodynamics although it seems that they may have more important implications in this field. Especially the curvature R is a new thermodynamic function of the more standard thermodynamic parameters. The first law of thermodynamics concerns only first-order derivatives of $\ln Z$, and thus only first moments of stochastic variables F_i. The second law of thermodynamics imposes some conditions on the second moments of these variables, or more generally on their covariances, and thus on second derivatives of $\ln Z$. The curvature tensor and the scalar curvature R are functions of the second and third moments of F_i. Imposition of any condition on R would be equivalent to introducing a new law of thermodynamics.

Another choice is not to go beyond the standard thermodynamics but to find an interpretation for R as a new thermodynamic quantity. From experiments we know that in one-phase systems, far away from critical points and lines of coexisting phases, fluctuations are small and states of systems are stable. Fluctuations become large and important in multi-phase systems, and especially in the vicinity of critical points. As a result, in the vicinity of critical points systems become extremely unstable. Comparison of these facts with the computed numerical values of R leads to the conclusion that R (or $R-a$, a — a constant, for magnetic systems) could be treated as a joint measure of fluctuations occurring in the vicinity of a given state, whereas R^{-1} (or $(R-a)^{-1}$) could be treated as a measure of stability of the state [13, 15] (the bigger R, the less stable system). To support this interpretation let us remind that $R=0$ for ideal gas, $R \geq 0$ for all classical systems. For quantum systems, $R \geq 0$ for ideal boson gas and $R \leq 0$ for ideal fermion gas (not computed here) [15]; $R \geq 0$ for 1D Ising system and $R \geq 0$ for the mean field model [13]. Because $R=0$ for ideal gas (noninteracting particles) we actually infer that R measures this part of fluctuations which comes from interactions only, and hence measures 'nonideality' of the systems.

The ideal fermion gas is so far the only known physical system for which R is negative. This does not contradict the above interpretation of R because this system occupies quite a unique position among all physical systems. It is known that fluctuations in an ideal fermion gas are actually smaller than in the classical ideal gas because of negative spatial correlations occurring for fermionic systems. This is caused by the statistical effect of repulsion between particles (the Pauli principle) [4].

As a possible generalizations we would like to propose to compute R for:

- parameterized families of probability distributions of different types (not necessarily exponential distributions),
- generalized parameterized distributions for which θ^i would also include temperatures of higher order (such probability distributions could be obtained from the maximum information principle based not only on the knowledge of first moments m_i but also on the knowledge of higher-order moments,
- statistical models describing irreversible phenomena such as relaxation, diffusion and transport processes with dissipation. Then $I(\rho|\sigma)$ would mean the *production of entropy*,
- statistical models having applications in communication theory, pattern recognition, also in biology and ecology, in theoretical linguistics, and so on.

Yet another possibility could be based not on the use of Riemannian metric but rather on a connection (or a family of connections) [6, 7].

Acknowledgements

I would like to thank the organizers of this conference, and in particular to Professors M. Ohya and N. Watanabe, for invitation, hospitality and support.

References

[1] E. T. Jaynes: *Phys. Rev.* **106**, 620 (1957).

[2] R. S. Ingarden and K. Urbanik: *Bull. Acad. Polon. Sci., Ser. Sci. Phys. Astron.* **IX**, 313 (1961).

[3] S. Kullback: *Information Theory and Statistics*, Wiley, New York 1959.

[4] R. K. Pathria: *Statistical Mechanics*, Pergamon, Oxford 1972.

[5] R. Balian: *From Microphysics to Macrophysics*, Springer, Berlin 1991.

[6] S. Amari: *Differential-Geometrical Methods in Statistics*, LNS **28**, Springer, Berlin 1985.

[7] M. K. Murray and J. W. Rice: *Differential Geometry and Statistics*, Chapman & Hall, London 1993.

[8] R. A. Fisher: *Phil. Trans. Roy. Soc.* A222, 309 (1921).

[9] C. R. Rao: *Bull. Calcutta Math. Soc.* **37**, 81 (1945).

[10] H. Janyszek: *J. Phys.* A *Math. Gen.* **23**, 477–490 (1990).

[11] D. Brody and N. Rivier: *Phys. Rev.* E **51**, 1006–1011 (1995).

[12] R. M. Wilcox: *J. Math. Phys.* **4**, 884 (1967).

[13] H. Janyszek and R. Mrugała: *Phys. Rev.* A **39**, 6515 (1989).

[14] K. Huang: *Statistical Mechanics*, John Wiley and Sons, New York 1963.

[15] H. Janyszek and R. Mrugała: *J. Phys.* A **23**, 467 (1990).