Conference Proceedings Paper

Dually Flat Geometries in the State Space of Statistical Models

Jan Naudts

Universiteit Antwerpen, Universiteitsplein 1, 2610 Antwerpen, Belgium; jan.naudts@uantwerpen.be

Abstract: This paper investigates whether the dually flat geometries introduced by Amari are useful as a tool to study the manifold of thermodynamic equilibrium states. The mathematical setting is that of statistical models belonging to an exponential family. The metric tensor is derived from the relative entropy. Flat geometries are introduced and thermodynamic length is calculated. The ideal gas serves as an example.

Keywords: information geometry; dually flat geometries; thermodynamic length; exponential family; relative entropy; ideal gas

---

1. Introduction

The aim of the present contribution is to survey recent developments in information geometry and to point out their relevance for thermodynamics and statistical physics. Information geometry [1–4] is differential geometry used as an analytic tool in other domains such as statistics and information theory. Before information geometry emerged as a sub-discipline of mathematics, differential geometry was already applied to thermodynamics and statistical physics, starting with the works of Weinhold [5] and of Ruppeiner [6]. See also [7–13].

An essential feature of thermodynamics is its dual structure. Equilibrium states can be characterized either by intensive variables, or by corresponding extensive variables. Examples are inverse temperature β and external magnetic field h for the intensive variables, total energy U and total magnetization M for the extensive quantities. They are related via a pair of convex or concave functions, *in casu,* the entropy $S(U, M)$ and Massieu's function $\Phi(\beta, h)$. In statistical physics the latter is given by the logarithm of the partition sum. The relation between intensive and extensive variables is based on Legendre transforms.

In the literature [10] the question was debated whether the geometry of the set of equilibrium states can be derived from the metric of a Euclidean space in which the equilibrium states are embedded. Behind this question resides the mathematical result that any Riemannian manifold can be isometrically embedded in a Euclidean space. The present paper goes beyond Riemannian manifolds. To clarify this point let me refresh some basic concepts of differential geometry. Two objects determine the geometry of a differentiable manifold.

- The metric tensor g determines the length of vectors belonging to a tangent plane and the angles between them. In the present work g is defined starting from the relative entropy $D(p, p^\theta)$, which is minus the entropy of an arbitrary probability distribution p when the equilibrium distribution p^θ is taken as the reference measure. See the next section. Mathematicians call the quantity $D(p, p^\theta)$ a divergence and use it as a means of determining g, which is then the Fisher information metric.

- The metric tensor g does not suffice, one needs a notion of geodesics as well. Usually one works with Riemannian geodesics which are determined by the Christoffel symbols Γ^k_{ij} and the Levi-Civita connection. However, other connections are of interest as well. In particular, Amari [1] introduced a couple of flat connections. One of them is the trivial connection, given the canonical

coordinates of a model belonging to the exponential family. The other is its dual with respect to the Levi-Civita connection. These concepts are discussed in Section 3.

Throughout the text the notation $\partial_i \equiv \partial/\partial\theta^i$ is used. Einstein's summation convention is followed.

2. Exponential Families

2.1. A Statistical Model

Starting point of Statistical Physics is that the probability $p(x)$ of finding a classical system in the state x is given by the Boltzmann–Gibbs distribution. In statistics such a model is said to belong to the exponential family [14] because it can be written into the form

$$p^\theta(x) = \exp(\theta^k F_k(x) - \alpha(\theta)). \tag{1}$$

The probability distribution, written in this form, facilitates the calculation of certain expectation values. Indeed, from the normalization of $p^\theta(x)$ follows

$$0 = \partial_k \int dx\, p^\theta(x) = \int dx\, p^\theta(x) \left[F_k(x) - \partial_k \alpha(\theta) \right]. \tag{2}$$

This implies $\mathbb{E}_\theta F_k = \partial_k \alpha(\theta)$, where $\mathbb{E}_\theta F_k$ denotes the expectation of $F_k(x)$ for the given values of the parameters θ^k ($\mathbb{E}_\theta F_k = \langle F_k \rangle_\theta$ in physicist's notation).

Consider now the series expansion $p^{\theta+\epsilon}(x) = p^\theta(x) + \epsilon^k p^\theta(x) \left[F_k(x) - \mathbb{E}_\theta F_k \right] + \cdots$. This expression shows that the random variables $F_k(x) - \mathbb{E}_\theta F_k$ can be identified with vectors tangent to the manifold of model states at the given point p^θ.

2.2. The Metric Tensor

At this point one can proceed in many ways. In a more general context such as that of [15,16] the indicated way runs via the relative entropy, which for a model of the exponential family equals the Kullback–Leibler divergence,

$$D(p, p^\theta) = \int dx\, p(x) \log \frac{p(x)}{p^\theta(x)}. \tag{3}$$

The relative entropy is related to the Boltzmann–Gibbs–Shannon entropy

$$S(p) = -\int dx\, p(x) \log p(x) \tag{4}$$

by

$$\begin{aligned} D(p, p^\theta) &= S(p^\theta) - S(p) + \int dx\, [p^\theta(x) - p(x)] \log p^\theta(x) \\ &= S(p^\theta) - S(p) + \int dx\, [p^\theta(x) - p(x)] \theta^k F_k(x). \end{aligned} \tag{5}$$

The metric tensor g is defined by $\quad g_{ij}(\theta) = \partial_i \partial_j D(p, p^\theta) \Big|_{p=p^\theta}$.

A short calculation gives $\quad g_{ij}(\theta) = \int dx\, \frac{1}{p^\theta(x)} \left[\partial_i p^\theta(x) \right] \left[\partial_j p^\theta(x) \right]$.

The latter is a well-known expression for the Fisher information matrix.

2.3. The Statistical Gradient

An inner product of pairs of random variables U, V is defined by

$$\langle U, V \rangle_\theta \;=\; \int dx \, p^\theta(x) U(x) V(x). \tag{6}$$

One has for any $U(x)$

$$\partial_i \mathbb{E}_\theta U = \int dx \, \partial_i p^\theta(x) U(x) = \int dx \, p^\theta(x) \partial_i \log(p^\theta(x)) U(x) = \langle F_i - \mathbb{E}_\theta F_i, U - \mathbb{E}_\theta U \rangle_\theta. \tag{7}$$

The estimator $U - \mathbb{E}U$ is (identified with) a tangent vector and is called the statistical gradient of the expectation $\mathbb{E}_\theta U$. In particular one has

$$\partial_i \mathbb{E}_\theta F_j = \langle F_i - \mathbb{E}_\theta F_i, F_j - \mathbb{E}_\theta F_j \rangle_\theta = g_{ij}(\theta). \tag{8}$$

The tangent vectors $F_i - \mathbb{E}_\theta F_i$ span a Hilbert space. The elements of the metric tensor $g(\theta)$ determine the length of these basis vectors and the angles between them.

2.4. Dual Parameters

The Massieu function $\Phi(\theta)$ is the Legendre transform of the entropy. It is defined by

$$\Phi(\theta) = \max_p \{ S(p) + \theta^k \mathbb{E}_p F_k \}, \qquad \text{where} \qquad \mathbb{E}_p F_k = \int dx \, p(x) F_k(x). \tag{9}$$

The maximum is reached for $p = p^\theta$. Because it is a Legendre transform one automatically has

$$\partial_i \Phi(\theta) \;=\; \eta_k \quad \text{with} \quad \eta_k = \mathbb{E}_p F_k. \tag{10}$$

The η_k can be used instead of the θ^k to parametrize the model. From

$$\Phi(\theta) \;=\; S(p^\theta) + \theta^k \eta_k \qquad \text{follows that} \qquad \frac{\partial}{\partial \eta_i} S(p^\theta) = \frac{\partial \theta^j}{\partial \eta_i} \partial_j \left[\Phi(\theta) - \theta^k \eta_k \right] = -\theta^i. \tag{11}$$

3. Flat Geometries

Whether the model p^θ is flat or curved depends on the choice of connection. Up to now the literature on differential geometry in thermodynamics always choose for the Riemannian curvature, which is based on the Levi–Civita connection. However, the dually flat geometry of Amari [1–4] offers an alternative.

3.1. Dual Connections

The Levi–Civita connection is determined by the Christoffel symbols

$$\Gamma^k_{ij} \;=\; \frac{1}{2} g^{ks} \left(\partial_i g_{sj} + \partial_j g_{is} - \partial_s g_{ij} \right), \tag{12}$$

which are fully determined by the metric tensor g. Note that g^{kl} are the matrix entries of the inverse of the metric tensor g. Other connections are now considered. For the corresponding connection coefficients the symbols ω^k_{ij} are used instead of Γ^k_{ij}.

Two connections with coefficients ω and ω^* are said to be each other dual if they satisfy

$$\omega^l_{ki} g_{lj} + g_{il} \omega^{*l}_{kj} \;=\; \partial_k g_{ij}. \tag{13}$$

It is straightforward to verify in the present context that the Levi–Civita connection is the only connection which is torsion free (i.e. $\Gamma^k_{ij} = \Gamma^k_{ji}$) and self-dual.

Given an exponential family and its canonical coordinates θ^k, it is obvious to consider the trivial connection in which all connection coefficients vanish. The geodesic through two equilibrium states p^θ and $p^{\theta'}$ is then a straight line in parameter space. It satisfies

$$
\begin{aligned}
\log p^{(1-t)\theta+t\theta'}(x) &= [(1-t)\theta^k + t\theta'^k]F_k(x) - \alpha(\theta) \\
&= (1-t)\log p^\theta(x) + t\log p^{\theta'}(x).
\end{aligned}
\tag{14}
$$

It is known (see for instance Theorem 3.3 of [2]) that the dual of a flat connection is again a flat connection. This implies that there exists another flat connection, which is fully determined by the metric tensor g. From the definition (13) follows that its coefficients are given by

$$
\omega^k_{ij} = g^{kl}\partial_i g_{jl} = 2\Gamma^k_{ij}.
\tag{15}
$$

3.2. Euler–Lagrange Equations

Let us now calculate the geodesic for the connection (15). Consider the time-dependent orbit $p^{\gamma(t)}$. The connection coefficients appear in the Euler-Lagrange equations

$$
\ddot{\gamma}^k + \omega^k_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0.
\tag{16}
$$

Let us verify that the solutions are of the form

$$
\gamma^k(t) = \theta^k[(1-t)\eta^{(1)}_k + t\eta^{(2)}_k],
\tag{17}
$$

where $\eta^{(1)}$ and $\eta^{(2)}$ are constant and $\theta^k[\eta]$ is the inverse function of $\eta^k(\theta)$. From (17) follows

$$
\dot{\gamma}^k = \frac{\partial\theta^k}{\partial\eta_i}\left(\eta^{(2)}_i - \eta^{(1)}_i\right) = g^{ki}\left(\eta^{(2)}_i - \eta^{(1)}_i\right)
\tag{18}
$$

$$
\ddot{\gamma}^k = \left(\partial_j g^{ki}\right)\dot{\gamma}^j\left(\eta^{(2)}_i - \eta^{(1)}_i\right) = \left(\partial_j g^{ki}\right)g^{jl}\left(\eta^{(2)}_l - \eta^{(1)}_l\right)\left(\eta^{(2)}_i - \eta^{(1)}_i\right).
\tag{19}
$$

From the definition (15) follows $\partial_j g^{ki} = -g^{im}\omega^k_{jm}$. Hence one obtains

$$
\ddot{\gamma}^k = -g^{im}\omega^k_{jm}g^{jl}\left(\eta^{(2)}_l - \eta^{(1)}_l\right)\left(\eta^{(2)}_i - \eta^{(1)}_i\right) = -\omega^k_{jm}\dot{\gamma}^m\dot{\gamma}^j,
\tag{20}
$$

which is (16). One concludes that for this connection the geodesics follow straight lines in the space of the extensive parameters η_i.

3.3. Thermodynamic Length

The main advantage of a flat geometry is that it is easy to calculate path lengths. Fix two equilibrium states with intensive parameters $\theta_{(1)}$, respectively $\theta_{(2)}$. The squared length of an infinitesimal line segment is $(\mathrm{d}s)^2 = g_{ij}(\theta)\mathrm{d}\theta^i\mathrm{d}\theta^j$. The geodesic γ connecting the two points has coordinates

$$
\gamma^k(t) = (1-t)\theta^k_{(1)} + t\theta^k_{(2)}.
\tag{21}
$$

The length of the path between the two points equals

$$
L(\theta_{(1)}, \theta_{(2)}) = \int_\gamma \mathrm{d}s = \int_\gamma \sqrt{g_{ij}(\theta)\mathrm{d}\theta^i\mathrm{d}\theta^j} = \sqrt{\bar{g}_{ij}\left[\theta^i_{(2)} - \theta^i_{(1)}\right]\left[\theta^j_{(2)} - \theta^j_{(1)}\right]}
\tag{22}
$$

with

$$\overline{g}_{ij} = \left[\int_0^1 dt \sqrt{g_{ij}(\gamma(t))} \right]^2. \tag{23}$$

Similarly, let $\eta^{(1)}$ and $\eta^{(2)}$ be the extensive parameters corresponding with $\eta^{(1)}$, respectively $\eta^{(2)}$. The path linear in the η-space is determined by $\eta_k(\gamma(t)) = (1-t)\eta_k^{(1)} + t\eta_k^{(2)}$. The length along this path equals

$$L^*(\eta^{(1)}, \eta^{(2)}) = \sqrt{\overline{g}^{ij} \left[\eta_i^{(2)} - \eta_i^{(1)} \right] \left[\eta_j^{(2)} - \eta_j^{(1)} \right]} \quad \text{with} \quad \overline{g}^{ij} = \left[\int_0^1 dt \sqrt{g^{ij}(\gamma(t))} \right]^2. \tag{24}$$

3.4. Example

The Massieu function of the ideal gas in a volume V can be written as (see for instance [14])

$$\Phi(\beta, \mu) = \frac{V}{V_0} e^{\beta\mu} \left(\frac{\beta_0}{\beta} \right)^{3/2}. \tag{25}$$

Here, β is the inverse temperature and μ is the chemical potential.

The probability distribution p^θ belongs to the exponential family with $\theta^1 = \beta/\beta_0$ and $\theta^2 = \beta\mu$. A short calculation gives

$$\eta_1 = -\frac{3}{2\theta^1}\Phi \quad \text{and} \quad \eta_2 = \Phi \quad \text{and} \quad g(\theta) = \frac{1}{\theta^1}\Phi \begin{pmatrix} \frac{15}{4\theta^1} & -\frac{3}{2} \\ -\frac{3}{2} & \theta^1 \end{pmatrix}. \tag{26}$$

Translated in the usual symbols of physics the latter becomes (see for instance Section 1.6 of [14])

$$g(\theta) = pV \begin{pmatrix} 15\beta_0/4\beta & -\frac{3}{2} \\ -\frac{3}{2} & \beta/\beta_0 \end{pmatrix}, \tag{27}$$

where p is the pressure and is related to Massieu's function and the total number of particles N by $\beta pV = \Phi = N$. The connection ω, defined by (15), equals 2Γ. The Christoffel symbols are given by

$$\Gamma^1 = \begin{pmatrix} -5/2\theta^1 & 1/2 \\ 1/2 & 0 \end{pmatrix} \quad \text{and} \quad \Gamma^2 = \begin{pmatrix} -15/8[\theta^1]^2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{28}$$

A tedious calculation shows that all coefficients of the Riemann curvature tensor vanish. Hence the Riemannian geometry is also flat. This is an artifact due to the absence of interactions in the ideal gas model (see [6]).

Consider for instance an isotherm, this is, β is kept constant. Then a path of the form (21) changes the chemical potential μ from $\mu^{(1)}$ to $\mu^{(2)}$ in a linear way. The length of such a path equals

$$\begin{aligned} L(\theta_{(1)}, \theta_{(2)}) &= \beta|\mu^{(2)} - \mu^{(1)}| \int_0^1 dt \sqrt{g_{22}((1-t)\theta^1 + t\theta^2)} \\ &= \beta|\mu^{(2)} - \mu^{(1)}|\sqrt{\Phi(\beta, \mu^{(1)})} \int_0^1 dt \, e^{t\beta(\mu^{(2)} - \mu^{(1)})/2} \\ &= 2\left| \sqrt{\Phi(\beta, \mu^{(2)})} - \sqrt{\Phi(\beta, \mu^{(1)})} \right| = 2|\sqrt{N^{(2)}} - \sqrt{N^{(1)}}|. \end{aligned} \tag{29}$$

On the other hand, when the path is linear in η-space, then both Φ and Φ/θ^1 are changed in a linear way. For instance, keep $\eta_1 - \frac{3}{2\theta^1}\Phi \sim pV$ constant. Then the density $\rho = N/V$ is proportional to the inverse temperature β. The length along such a path is proportional the change in Massieu's function $\Phi = N$.

4. Discussion

The geometry of the manifold of thermodynamic equilibrium states is a long-standing topic of interest in thermodynamics and in statistical physics. The relation between curvature and interactions is being studied since the pioneering work of Ruppeiner [6]. The obvious context is the Riemannian geometry, which is based on the Levi-Civita connection. Recently, Amari [1] introduced dually flat geometries as an alternative tool to analyze statistical manifolds. The present work starts from the observation that the dual structure of thermodynamics is based on the same convex analysis as Amari's duality. The paper shows in the simple case of the ideal gas that these flat geometries have a thermodynamic meaning as well. Further research is needed to explore possible advantages of the present approach, in particular, when applied to interacting models.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Amari, S. *Differential Geometrical Methods in Statistics*; Springer: Berlin/Heidelberg, Germany, 1985.
2. Amari, S.; Nagaoka, H. Methods of information geometry, volume 191 of Translations of Mathematical Monographs. Available online: http://qh.eng.ua.edu/e_paper/e_book/Ameri_book_Methods_of_ Information_Geometry.pdf (accessed on 21 October 2016).
3. Amari, S. *Information Geometry and Its Applications.* Springer Japan KK: Tokyo, Japan, 2016.
4. Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L. *Information geometry*. In preparation.
5. Weinhold, F. Metric geometry of equilibrium thermodynamics. *J. Chem. Phys.* **1975**, *63*, 2479–2483.
6. Ruppeiner, G. Thermodynamics: A Riemannian geometric model. *Phys. Rev. A* **1979**, *20*, 1608–1613.
7. Salamon, P.; Nitzan, A.; Andresen, B.; Berry, S. Minimum entropy production and the optimization of heat engines. *Phys. Rev. A* **1980**, *21*, 2115–2129.
8. Gilmore, R. Length and curvature in the geometry of thermodynamics. *Phys. Rev. A* **1984**, *30*, 1994–1997.
9. Andresen, B.; Berry, R.S.; Gilmore, R.; Ihrig, E.; Salamon, P. Thermodynamic geometry and the metrics of Weinhold and Gilmore. *Phys. Rev. A* **1988**, *37*, 845–848.
10. Andresen, B.; Berry, R.S.; Ihrig, E.; Salamon, P. Inducing Weinhold's metric from Euclidean and Riemannian metrics. *Phys. Rev. A* **1988**, *37*, 849–851.
11. Andresen, B.; Niven, R.K. Jaynes' Maximum Entropy Principle, Riemannian Metrics and Generalised Least Action Bound. In *Complex Physical, Biophysical and Econophysical Systems: Lecture Notes in Complex Systems*; Dewar, R. L.; Detering, F.; Hackensack, F., Eds.; World Scientific: Singapore, Singapore, 2009, pp. 283–318.
12. Crooks, G.E. Measuring Thermodynamic Length. *Phys. Rev. Lett.* **2007**, *99*, 100602.
13. Kumar, P.; Mahapatra, S.; Phukon, P.; Sarkar, T. Geodesics in information geometry: Classical and quantum phase transitions. *Phys. Rev. E* **2012**, *86*, 051117.
14. Naudts, J. *Generalised Thermostatistics.* Springer-Verlag: London, UK, 2011.
15. Naudts, J.; Anthonis B. Data set models and exponential families in statistical physics and beyond. *Mod. Phys. Lett. B* **2012**, *26*, 1250062.
16. Anthonis, B. Extension of information geometry for modeling non-statistical systems. Phd Thesis, University of Antwerp, Antwerp, Belgium, 2014. arXiv:1501.00853.