数理解析研究所講究録
1196巻 2001年 32-51

CLASSICAL AND QUANTUM INFO-MANIFOLDS

R. F. Streater

Dept. of Maths., King's College London, Strand, WC2R 2LS

1 Estimation; the Cramer-Rao inequality

Let $\rho_\eta(x)$ be a probability density, depending on a parameter $\eta \in R$. The *Fisher information* of ρ_η is defined to be [8]

$$G := \int \rho_\eta(x) \left(\frac{\partial \log \rho_\eta(x)}{\partial \eta} \right)^2 dx. \tag{1}$$

We note that this is the variance of the random variable $Y = \partial \log \rho_\eta / \partial \eta$, which has mean zero. G is associated with the *family* $\mathcal{M} = \{\rho_\eta\}$ of distributions, rather than any one of them. This concept arises in the theory of estimation as follows. Let X be a random variable whose distribution is believed or hoped to be one of those in $\mathcal{M}$. We estimate the value of η by measuring X independently m times, getting the data $x_1, \ldots, x_m$. An *estimator* f is a function of $(x_1, \ldots, x_m)$ that is used for this estimate. So X is a function of m independent copies of X, and so is a random variable. To be useful, the estimator must be independent of η, which we do not (yet) know. We say that an estimator is *unbiased* if its mean is the desired parameter; it is usual to take f as a function of X and to regard $f(x_i)$, $i = 1, \ldots, m$ as samples of f. Then the condition that f is unbiased becomes

$$\rho_\eta . f := \int \rho_\eta(x) f(x) dx = \eta. \tag{2}$$

We use the notation $\rho.f$ for the expectation of f in the state ρ. A good estimator should also have only a small chance of being far from the correct value, which is its mean if it is unbiased. This chance is measured by the variance. Fisher [8] stated, and Rao [24] and Cramer proved, that the variance V of an unbiased estimator f obeys the inequality $V \geq G^{-1}$. For the proof, differentiate eq. (2) w. r. t. η to get

$$\int \frac{\partial \rho_\eta(x)}{\partial \eta} f(x) dx = 1, \tag{3}$$

which can be written as

$$\int Y(x)(f(x)-\eta)\rho_\eta(x)\,dx = \int \left(\frac{\partial \log \rho}{\partial \eta}\right)(f(x)-\eta)\,\rho_\eta(x)\,dx = 1. \tag{4}$$

We note that this is the correlation of Y and f, so the covariance matrix becomes

$$\begin{pmatrix} G & 1 \\ 1 & V \end{pmatrix}. \tag{5}$$

This is positive semi-definite, giving the result. □

If we do N independent measurements of the estimator, and average them, we improve the inequality to $V \geq G^{-1}/N$. This inequality expresses that, given the family ρ_η, there is a limit to the reliability with which we can estimate η. Fisher termed V/G^{-1} the *efficiency* of the estimator f. Equality in the Schwarz inequality occurs if and only if the two functions are proportional. Let $-\partial\xi/\partial\eta$ denote the factor of proportionality. Then the optimal estimator occurs when

$$\log \rho_\eta(x) = -\int \partial\xi/\partial\eta(f(x)-\eta)\,d\eta. \tag{6}$$

Doing the integral, and adjusting the integration constant by normalisation, leads to

$$\rho_\eta(x) = Z^{-1}\exp\{-\xi f(x)\} \tag{7}$$

which is the 'exponential family'.

This can be generalised to any n-parameter manifold $\mathcal{M} = \{\rho_\eta\}$ of distributions, $\eta = (\eta_1, \ldots, \eta_n)$ with $\eta \in R^n$. Suppose we have unbiased estimators $(f_1, \ldots, f_n)$, with covariance matrix V. Fisher introduced the *information matrix*

$$G^{ij} = \int \rho_\eta(x) \frac{\partial \log \rho_\eta(x)}{\partial \eta_i} \frac{\partial \log \rho_\eta(x)}{\partial \eta_j} dx. \tag{8}$$

We note that $Y^j := \partial \log \rho/\partial\eta_j$ is a random variable with zero mean, and that G^{ij} is its covariance matrix. Rao remarked that G^{ij} provides a

Riemannian metric for $\mathcal{M}$. We now derive the analogue of the inequality when $n > 1$. Put $V_{ij} = \rho_\eta.[(f_i - \eta_i)(f_j - \eta_j)]$, the covariance matrix of $\{f_i\}$. Differentiate the condition for being unbiased,

$$\int \rho_\eta(x) f_i(x)\, dx = \eta_i \tag{9}$$

with respect to η_j, and rearrange as above, to get

$$\int \rho_\eta(x) Y^i(x)(f_j(x) - \eta_j)\, dx = \delta_{ij}. \tag{10}$$

This is the correlation between Y^i and f_j. The covariance matrix of the $2n$ random variables Y^i, f_j therefore is

$$\begin{pmatrix} G & I \\ I & V \end{pmatrix}. \tag{11}$$

This is therefore a positive semi-definite matrix. If it is not definite, it has zero as an eigenvalue, which leads to $GV = I$, and the manifold must be the exponential family, as before. If it is definite, so is its inverse, which is found to be

$$\begin{pmatrix} \left(G - V^{-1}\right)^{-1} & -G^{-1}\left(V - G^{-1}\right)^{-1} \\ -V^{-1}\left(G - V^{-1}\right)^{-1} & \left(V - G^{-1}\right)^{-1} \end{pmatrix}. \tag{12}$$

It follows that the leading submatrices $(G - V^{-1})^{-1}$ and $(V - G^{-1})^{-1}$ are positive definite, and thus so are their inverses. It follows that we get the matrix inequality $V \geq G^{-1}$.

2 Entropy methods, exponential families

Gibbs knew that the state of maximum entropy, given the mean energy, is the canonical state. More generally, let Ω be a countable sample space, and let Σ denote the set of probabilities (or *states*) on Ω. Let $f_1, \ldots, f_n$ be n linearly independent random variables, whose means we can measure. We want to find the 'best' choice for the the state, given these means. The least prejudiced choice of ρ (Jaynes) is to maximise the entropy S subject to the $n + 1$ constraints given by normalisation and the means of f_j, $j = 1, \ldots, n$. We use λ, ξ^j as Lagrange multipliers; then we must maximise

$$-\sum_{\omega\in\Omega} \rho(\omega) \log \rho(\omega) - \lambda \sum_{\omega} \rho(\omega) - \sum_{j=1}^{n} \xi^j \rho(\omega) f_j(\omega)$$

by varying $\rho(\omega)$ subject to no constraints. We get

$$\rho_\xi(\omega) = Z^{-1} \exp -\{\sum_j \xi^j f_j(\omega)\} \text{ where } Z = \sum_{\omega} \exp\{-\sum_j \xi^j f_j(\omega)\}. \quad (13)$$

These make up the *exponential manifold* M determined by $\mathcal{F}$:=Span$\{f_1, \ldots, f_n\}$ and parametrised by $\xi^1, \ldots, \xi^n$; these are called the canonical coordinates on $\mathcal{M}$, which has dimension n. At least one, say f_1, must be bounded below, to ensure $Z < \infty$ holds for some ξ.

The ξ^j are determined by the given expectation values by the conditions $\rho_\xi . f_j = \eta_j$, $j = 1, \ldots, n$. The η_j are thus also coordinates for the manifold (the *mixture* coords.) It is easy to show that

$$\eta_j = -\frac{\partial \Psi}{\partial \xi^j}, \qquad j = 1, \ldots, n; \qquad V_{jk} = -\frac{\partial \eta_j}{\partial \xi^k}, \qquad j, k = 1, \ldots, m, \quad (14)$$

where $\Psi = \log Z$, and that Ψ is a convex function of ξ^j. The Legendre dual to Ψ is $\Psi - \sum \xi^i \eta_i$ and this is the entropy $S = -\rho . \log \rho$. The dual relations are

$$\xi^j = \frac{\partial S}{\partial \eta_j} \qquad G^{jk} = -\frac{\partial \xi^j}{\partial \eta_k}. \quad (15)$$

By the rule for Jacobians, V and G are mutual inverses. Therefore, the method of maximum entropy leads to the exponential family, which allows the optimisation of the Cramer-Rao bound, and gives us estimators of 100% efficiency.

3 Manifolds modelled by Orlicz spaces

Pistone and Sempi [23] have developed a version of information geometry, which does not depend on a choice of $\mathcal{F}$, the span of a finite number of estimators. Let (Ω, μ) be measure space and let $\mathcal{M}$ be the set of all probability measures that are equivalent to μ; such a measure is determined by its Radon-Nikodym derivative ρ relative to μ. The topology on $\mathcal{M}$ is not given by the L^1-distance, but by an Orlicz norm.

Given $\rho \in \mathcal{M}$, the *Cramer class* at ρ is the set of all random variables X on (Ω, μ) such that the moment-generating function

$$\widehat{X}_\rho(t) := \int e^{-tX} \rho d\mu \tag{16}$$

is finite in a 'hood of the origin. This is enough to ensure that it is analytic in an interval about $t = 0$. The Cramer class C_ρ at a point ρ in $\mathcal{M}$ is furnished with the Luxemburg norm

$$\|X\|_\rho = \inf \left\{ r > 0 : E_\rho \left[\cosh \left(\frac{u}{r} \right) - 1 \right] \leq 1 \right\}. \tag{17}$$

The Cramer class C at ρ is an Orlicz space, and so is a Banach space with this norm. The centred Cramer class $C(0)$ is defined as the subset of C at ρ with zero mean in the state ρ; this is a closed subspace. A sufficiently small ball in the quotient Banach space $C/C(0)$ then parametrises a 'hood of ρ, and can be identified with the tangent space at ρ; namely, the 'hood contains those points σ of $\mathcal{M}$ such that

$$\sigma = Z^{-1} e^{-X} \rho \qquad \text{for some } X \in C. \tag{18}$$

where Z is a normalising factor. Pistone and Sempi show that the bilinear form

$$G(X, Y) = E_\rho [XY] \tag{19}$$

is a Riemannian metric on the tangent space C/C_0, thus generalising the Fisher-Rao theory.

This theory is called *non-parametric* estimation theory, because we do not limit the distributions to those specified by a finite number of parameters, but allow any 'shape' for the density ρ. It is this construction that we take over to the quantum case, except that the spectrum is discrete and the distributions are not always equivalent.

4 Efron, Dawid and Amari

A Riemannian metric G, eq. (15) gives us a notion of parallel transport, namely, that given by the Levi-Civita affine connection. Recall that an affine map, U (acting on the right) from one vector space $\mathcal{T}_1$ to another, $\mathcal{T}_2$, is one that obeys

$$(\lambda XU+(1-\lambda)YU) = \lambda XU+(1-\lambda)YU, \text{ for all } X,Y \in \mathcal{T}_1 \text{ and all } \lambda \in [0,1]. \tag{20}$$

The same definition works on an *affine space*, that is, a convex subset of a vector space. This leads to the concept of an affine connection.

Let $\mathcal{M}$ be a manifold and denote by T_ρ the tangent space at $\rho \in \mathcal{M}$. Consider an affine map $U_\gamma(\rho,\sigma) : T_\rho \to T_\sigma$ defined for each pair of points ρ,σ and each continuous path γ in the manifold starting at ρ and ending at σ. Let ρ,σ and τ be any three points and γ_1 a path from ρ to σ, and γ_2 any path from σ to τ.

Definition 1 *We say that U is an* affine connection, *if $U_\emptyset = Id$ and*

$$U_{\gamma_1\cup\gamma_2} = U_{\gamma_1} \circ U_{\gamma_2}. \tag{21}$$

Let X be a tangent vector at ρ; we call XU_{γ_1} the parallel transport of X to σ, along the path γ_1.

We also require U to be smooth in ρ in a 'hood of the point ρ, when we identify a ball in the tangent space with part of the manifold by the exponential map. In physics it is usually the differential of U along a specified direction that is called 'affine connection'. Equivalently, a connection defines a covariant derivative of a vector field on the manifold:

$$\nabla_Y X := d/dt\; XU_\gamma(\rho,\gamma(t))|_{t=0} \tag{22}$$

where $\{\gamma(t)\}$, $0 \le t \le 1$ is any path from ρ to σ, which starts at ρ in the direction $Y \in T_\rho$. This is designed to convert vector fields to tensor fields. Conversely, a covariant derivative defines a connection. This concept allows us to specify that two tangent vectors to the manifold at points ρ and σ are *parallel* if the parallel transport (along a specified curve) of one from ρ to σ is proportional to the other. A *geodesic* is a self-parallel curve on $\mathcal{M}$: the tangent vectors to the curve at different points are parallel, when transported along the curve. Geodesics relative to the Levi-Civita connection are lines of minimal length, as measured by the metric.

Estimation theory might be considered geometrically as follows. For theoretical reasons, we expect the distribution of a random variable to lie on a submanifold $\mathcal{M}_0 \subseteq \mathcal{M}$ of states. The data give us a histogram, which is a distribution, but not a pretty one. We seek the point on $\mathcal{M}_0$ that is 'closest' to the data. Suppose that the sample space is Ω, with $|\Omega| < \infty$. Let us place all positive distributions, including the experimental one, in a common manifold, $\mathcal{M}$. This manifold will have the Riemannian structure, G, provided by the Fisher metric. We then draw the geodesic curve through the data point that has shortest distance to the sub-manifold $\mathcal{M}_0$; where it cuts $\mathcal{M}_0$ is our estimate for the state. This procedure, however, does not always lead to unbiased estimators. Efron [7] and Dawid [6] noticed that the Levi-Civita connection is not the only useful one, and that there are others that might be used in estimation theory. First, the ordinary mixtures of densities ρ_1, ρ_2 leads to

$$\rho = \lambda\rho_1 + (1-\lambda)\rho_2, \qquad 0 < \lambda < 1. \tag{23}$$

Done locally, this leads to a connection on the manifold, now called the (-1)-Amari connection: two tangents are parallel if they are proportional as functions on the sample space. This differs from the parallelism given by the Levi-Civita connection. We need to use (-1)-geodesics to give unbiased estimates for f.

There is another obvious convex structure, that obtained from the linear structure of the space of centred random variables, also known as the scores. Take $\rho_0 \in \mathcal{M}$ and write $f_0 = -\log \rho_0$. Consider a perturbation ρ_X of ρ_0, which we write as

$$\rho_X = Z_X^{-1} e^{-f_0 - X}. \tag{24}$$

The random variable X is not uniquely defined by ρ_X, since by adding a constant to X, we can adjust the partition function to give the same ρ_X. Among all these equivalent X we can choose the *score* which has zero expectation in the state ρ_0: $\rho_0.X = 0$. We can define a sort of mixture of two such perturbed states, ρ_X and ρ_Y by

$$\text{`}\lambda\rho_X + (1-\lambda)\rho_Y\text{'} := \rho_{\lambda X + (1-\lambda)Y}. \tag{25}$$

This is a convex structure on the space of states, and differs from that given in eq. (23). It leads to an affine connection, now called the $(+1)$-Amari connection. How do these connections relate to the metric?

Definition 2 *Let G be a Riemannian metric on the manifold $\mathcal{M}$. A connection $\gamma \mapsto U_\gamma$ is called a metric connection if*

$$G_\sigma(XU_\gamma, YU_\gamma) = G_\rho(X, Y) \tag{26}$$

for all tangent vectors X, Y and all paths γ from ρ to σ.

The Levi-Civita connection is a metric connection, but the $(\pm)$ Amari connections are not; they are, however, dual relative to the Rao-Fisher metric; let γ be a path connecting ρ with σ; then for all X, Y:

$$G_\sigma(XU^+(\rho,\sigma), YU^-(\rho,\sigma)) = G_\rho(X,Y). \tag{27}$$

Let $\nabla^\pm$ be the two covariant derivatives obtained from the connections $U^\pm$. Amari [1] defines intermediate covariant derivatives

$$\nabla^\alpha = \frac{1}{2}(1+\alpha)\nabla^+ + \frac{1}{2}(1-\alpha)\nabla^-. \tag{28}$$

These uniquely define connections, $U^{(\alpha)}$, whose dual relative to G is $U^{(-\alpha)}$. The Levi-Civita covariant derivative is the case $\alpha = 0$, which is self-dual and therefore metric, as is known. Amari shows that $\nabla^{(\pm)}$ define flat connections without torsion. Flat means that the transport is independent of the path, and 'no torsion' means that U takes the origin of T_ρ to the origin of T_ρ around any loop; it is linear, and not a general affine map. In that case there are affine coordinates, that is, global coordinates in which the respective convex structure is obtained by simply mixing coordinates linearly. Amari shows that for $\alpha \neq \pm 1$, ∇^α is not flat, but that the manifold is a sphere in the Banach space ℓ^p, $p = -\alpha/2 + 1/2$. In particular, the case $\alpha = 0$ leads to the unit sphere in the Hilbert space L^2, and the Levi-Civita parallel transport is vector translation in this space, followed by projection back onto the sphere. The resulting affine connection is not flat, because the shpere is not flat. The metric distance between measures is the Hellinger distance, and the natural coordinates are the square-roots of the densities, imitating the wave-functions of quantum mechanics. Similar results were obtained in infinite dimensions in [9, 10].

In estimation theory, the method of maximum entropy for unbiased estimators makes use of the ∇^- connection. This is true also in the dynamics of neural nets, dense liquids, Onsager theory, Brownian particles in a potential and the Soret and Dufour effects [25]; the micro-state after a small time is replaced by a macrostate, which is the same as the max-entropy estimation of the state by one on the manifold generated by exponentials of the macrovariables (or, slow variables). The (intractible) microdynamics is continuously projected in a rolling construction onto the (easier) manifold of exponential states. This idea was proposed by Kossakowski [17], Ingarden, et al. [16], and beautifully expounded by Balian, et al. [3]. The resulting non-linear dynamics can be described thus: after each time-step of the linear dynamics of the system, Nature makes the best estimate of the state among those lying on the manifold.

5 The finite quantum info manifold

Chentsov [5] asked whether the Fisher-Rao metric was unique. Any manifold has a large number of different metrics on it; apart from those that differ just by a constant factor, one can multiply a metric by a space-dependent factor. There are many others. Chentsov therefore imposed conditions on the metric. He saw the metric (and the Fisher metric in particular) as a measure of the distinguishability of two states. He argued that if this is to be true, then the distance between two states must be *reduced* by any stochastic map; for, a stochastic map must 'muddy the waters', reducing our ability to distinguish states. He therefore considered the class of metrics G that are reduced by any stochastic map on the random variables.

Definition 3 *A stochastic map is a linear map on the algebra of random variables that preserves positivity and takes 1 to itself.*

Chentsov was able to prove that the Fisher-Rao metric is unique, among all metrics, being the only one (up to a constant multiple) that is reduced by any stochastic map. It is therefore uniquely defined up to this factor within the category of commutative function algebras, with stochastic maps as morphisms.

In quantum mechanics, instead of the abelian algebra of random variables we use the algebra of matrices M_n. Measures on Ω are replaced by 'states', that is, $n \times n$ density matrices. For convenience we limit discussion to the interior of the set of states; these are positive-definite matrices of trace 1, which are faithful states and invertible matrices. We take this set to be the manifold $\mathcal{M}$; it is a genuine manifold, and not one of the non-commutative manifolds without points that occur in Connes's theory. The natural morphisms of the quantum info manifold are the completely positive maps that preserve the identity. Chentsov found some good candidates for different monotone metrics, hinting that uniqueness of the metric is not true for quantum mechanics. In fact, this is so; Petz completed the analysis after Chentsov died; see [20, 14].

As in the classical case, there are several affine structures on this manifold. The first comes from the mixing of the states, and is called the -1-affine structure. Coordinates for a state ρ in a hood of ρ_0 are provided by $\rho - \rho_0$, a small traceless matrix. The whole tangent space at ρ is thus identified with the set of traceless matrices, and this is a vector space with the usual rules for adding matrices. Obviously, the manifold is flat relative to this affine structure.

The $+1$-affine structure is constructed as follows. Since a state $\rho_0 \in \mathcal{M}$ is faithful we can write $H_0 := -\log \rho_0$ and any ρ near $\rho_0 \in \mathcal{M}$ as

$$\rho = Z_X^{-1} \exp -(H_0 + X) \tag{29}$$

for some Hermitian matrix X, which is ambiguous up to a multiple of the identity. We choose to fix X by requiring $\rho_0.X = 0$, and call X the 'score' of ρ. Then the tangent space at ρ can be identified with the set of scores, and the $+1$-linear structure is given by matrix addition of the scores. Corresponding to these two affine structures, there are two affine connections, whose covariant derivatives are denoted $\nabla^{(\pm)}$. Following Hasegawa [13], one can also form interpolating affine structures from eq. (28).

As an example of a metric on $\mathcal{M}$, let $\rho \in \mathcal{M}$, and for X, Y in T_ρ define the *GNS* metric by

$$G_\rho(X,Y) = \mathrm{Re}\,\mathrm{Tr}[\rho XY]. \tag{30}$$

This metric is reduced by all cp stochastic maps F; that is, it obeys

$$G_{F^*\rho}(XF, XF)) \leq G_\rho(X,X), \tag{31}$$

in accordance with Chentsov's idea. G is just the real part of the scalar product in the Gelfand-Naimark-Segal construction, and is positive definite since ρ is faithful. This has been adopted by Helstrom and others [15, 28, 19] in the theory of quantum estimation theory. However, Nagaoka [18] has noted that if we take this metric, then the $(+1)$ and the (-1) affine connections are not dual; the dual to the (-1) affine connection, relative to this metric, is not flat and has torsion. This failure of duality is confirmed in [14].

In estimation theory we naturally seek a quantum analogue of the Cramer-Rao inequality. Given a family $\mathcal{M}$ of density operators, parame-trized by a real parameter η, we seek an estimator X whose mean we can measure in the true state ρ_η. To be unbiased, we require $\mathrm{Tr}\,\rho_\eta X = \eta$, which, as in the classical case gives

$$\mathrm{Tr}\left\{\rho_\eta\rho_\eta^{-1}\frac{\partial\rho_\eta}{\partial\eta}(X-\eta)\right\} = 1. \tag{32}$$

It is tempting to regard $L_r = \rho^{-1}\partial\rho/\partial\eta$ as a quantum analogue of the Fisher info; it has zero mean, and the above equation says that its covariance with $X-\eta$ is equal to 1. The Schwarz inequality then leads to $\mathcal{V}(X) \geq [\rho_\eta.(L_r^*L_r)]^{-1}$, where we use $\rho.X$ to denote $\mathrm{Tr}[\rho X]$. For several estimators, the method used earlier gives this as a matrix inequality.

However, ρ and its derivative do not (in general) commute, so Y is not Hermitian, and is not popular as a measure of quantum information. Helstrom, and Petz and Toth [21] get round this by using the idea of a *logarithmic derivative.* Let g be a real or complex scalar product on the space of matrices; we say that a matrix L is the g-logarithmic derivative of the family ρ_η if for any matrix X,

$$\frac{\partial\rho_\eta.X}{\partial\eta} = g(L^*,X). \tag{33}$$

The symmetric logarithmic derivative uses the real part of the *GNS* metric for g, so that

$$\frac{\partial}{\partial\eta}\mathrm{Tr}(\rho_\eta X) = \frac{1}{2}\mathrm{Tr}[\rho_\eta(L_sX+XL_s)]. \tag{34}$$

Another metric in Chentsov's allowed class is the Bogoliubov-Kubo-Mori metric; let X and Y have zero mean in the state ρ. Then put

$$g_\rho(X,Y) = \int_0^1 \mathrm{Tr}\left[\rho^\alpha X \rho^{1-\alpha} Y\right] d\alpha. \tag{35}$$

This is one of the family of scalar products found by Petz to obey the Chentsov property. The corresponding logarithmic derivative, L_B, is defined such that

$$\frac{\partial}{\partial\eta}\rho_\eta.X = \int_0^1 \rho_\eta^\lambda L_B \rho_\eta^{1-\lambda} X\, d\lambda \tag{36}$$

and is given explicitly by

$$L_B = \int_0^\infty (\lambda+\rho_\eta)^{-1}\frac{\partial\rho_\eta}{\partial\eta}(\lambda+\rho_\eta)^{-1} d\lambda. \tag{37}$$

Each metric leads to a Cramer-Rao inequality, also in matrix form for several estimators, and some of these are stronger than others [21, 22].

The *BKM* metric has other desirable properties, apart from entering in Kubo's 'theory of linear response'. For the metric g, the connections with covariant derivatives $\nabla^{(\pm\alpha)}$ are dual, and there are affine coordinates for ∇^α, namely, it is the unit sphere in the (finite-dim.) Banach space $\mathcal{C}_p$, the Schatten class with norm $\|X\|_p = (\mathrm{Tr}|X|^p)^{1/p}$. The case $p = 1/2$, or $\alpha = 0$, leads to the Hilbert space of Hilbert-Schmidt operators, which has been used in [4]. More, the Massieu function $\log Z$ is the generating function for all the connected Kubo functions, and in particular, the mean is the first derivative, and the metric is the second, as in eq. (14). The entropy is again the Legendre transform of the Massieu function, and the reciprocal relations of eq. (15) hold. It follows that the Cramer-Rao inequality for the *BKM*-metric is achieved exactly for the exponential family, agreeing with the method of maximum entropy. In [12] we show that the *BKM* metric is the only Chentsov metric for which the $\pm$-affine structures are mutually dual.

6 Araki's expansionals and the analytic manifold

Araki [2] has considered the case where ρ is a *KMS* state on a W^*-algebra. He then perturbed the state by adding bounded operators to the *KMS* Hamiltonian; the perturbed *KMS* state has a convergent Kubo-Mori perturbation expansion, which defines an analytic function in the Banach space of bounded perturbations. We [26] try to follow this for unbounded perturbations.

Let Σ be the set of density operators on $\mathcal{H}$, and let $\operatorname{int}\Sigma$ be its interior, the faithful states. We shall deal only with systems described by $\rho \in \operatorname{int}\Sigma$; this means that for a free Schrödinger particle, or system of such, we are limited to systems inside a finite volume of real space. Then we would expect the entropy to be finite. The following class of states turns out to be tractable. Let $p \in (0,1)$ and let $\mathcal{C}_p$, denote the set of operators C such that $|C|^p$ is of trace class. This is like the Schatten class, except that we are in the bad case, $0 < p < 1$, for which $C \mapsto (\mathrm{Tr}[|C|^p])^{1/p}$ is only a quasi-norm. Let

$$\mathcal{C}_< = \bigcup_{0<p<1} \mathcal{C}_p. \tag{38}$$

One can show that the entropy

$$S(\rho) := -\mathrm{Tr}[\rho \log \rho] \tag{39}$$

is finite for all states in $\mathcal{C}_<$. We take the underlying set of the quantum info manifold to be

$$\mathcal{M} = \mathcal{C}_< \cap \operatorname{int}\Sigma. \tag{40}$$

We shall cover $\mathcal{M}$ with balls, each belonging to a Banach space, and shall show that we have a Banach manifold when $\mathcal{M}$ is furnished with the topology induced by the norms; for this, the main problem is to ensure that various Banach norms are equivalent.

Let $\rho_0 \in \mathcal{M}$ and write $H_0 = -\log \rho_0 + cI$. We choose c so that $H_0 \geq I$, and we write $R_0 = H_0^{-1}$ for the resolvent at 0. We define a 'hood of ρ_0 to be the set of states of the form

$$\rho_V = Z_V^{-1} \exp - (H_0 + V)\,, \tag{41}$$

where V is a sufficiently small H_0-bounded form perturbation of H_0. The necessary and sufficient condition to be Kato-bounded is that

$$\|V\|_0 := \|R_0^{1/2} V R_0^{1/2}\|_\infty < \infty. \tag{42}$$

The set of such V make up a Banach space, $\mathcal{T}(0)$, with (42) as norm. The first result is that $\rho_V \in \mathcal{M}$ for V inside a small ball in $\mathcal{T}(0)$. For the proof, let a be the form-bound of V, and let q_V be the form of $H_0 + V$. Then we have for some $b \geq 0$,

$$-bI + (1-a)q_0 \leq q_V \leq bI + (1+a)q_0. \tag{43}$$

Let L be any finite dimensional subspace of $\operatorname{Dom} q_0$, and put

$$\lambda(q, L) = \sup\{q(\psi, \psi) : \|\psi\| = 1, \;\; \psi \in L\}. \tag{44}$$

Then the ordered eigenvalues of q are given by

$$\lambda(q, n) = \inf\{\lambda(q, L) : \dim L = n\}. \tag{45}$$

¿From (43) we have for each L,

$$-b + (1-a)\lambda(q_0, n) \leq \lambda(q_V, L). \tag{46}$$

Since $\lambda(q_0, n) \to \infty$ with n, the spectrum of H_V is purely discrete. Thus

$$\exp \beta\,(b - (1-a)\lambda(q_0, n)) \geq \exp -\beta\lambda(q_V, n). \tag{47}$$

Summing over n gives the traces

$$\mathrm{Tr} e^{-\beta H_V} \leq e^{\beta(b-(1-a)H_0)}$$

which is of trace class for some $\beta < 1$ if a is small enough.

We now consider [27] the special case when V is an H_0-bounded as an operator; the condition for this is $\|R_0 V\| < \infty$. Then V is also form-bounded, since

$$\|R_0^{1/2} V R_0^{1/2}\|_\infty \leq \|R_0 V\|_\infty < \infty. \tag{48}$$

In this case we can use the larger norm to provide a topology. This is not equivalent to the topology we get using the norm (42); we are moving from ρ_0 in a direction more regular than the general direction in the tangent space, and this allows us to furnish this slice of the manifold with a stronger topology. The state defined by V is given by

$$\rho_V := Z_V^{-1} \exp -(H_0 + V). \tag{49}$$

Thus, V and $V + cI$ give rise to the same state; near ρ_0 the regular directions in $\mathcal{M}$ are thus parametrised by the quotient space

$$\widehat{\mathcal{T}} = \mathcal{T}/\{cI\}. \tag{50}$$

We may therefore use the *score*, $V - \rho_0.V$, as coordinates for the 'regular' manifold, now using just the operator bounded perturbations. We show that these are displacements of the state in *analytic directions*; in [11] we find a more general class of analytic directions, which together make up the 'analytic' manifold. This is an attempt to find the quantum analogue of the Cramer class. We shall come to this later.

The norms $\|R_0 V\|_\infty$ on overlapping regions are equivalent. For, around ρ_V we perturb with X such that $\|R_V X\|_\infty < \infty$, and

$$\|R_V X\|_\infty = \|R_V H_0 R_0 X\|_\infty \leq \|R_V H_0\|.\|R_0 V\|_\infty, \tag{51}$$

and the converse inequality holds similarly. We define the $(+)$-affine connection by transporting the score $V - \operatorname{Tr} \rho V$ at the point ρ to the score $V - \operatorname{Tr} \sigma V$ at σ. This connection is flat and torsion-free, since it patently does not depend on the path between ρ and σ. The $(-)$-connection can be defined in $\mathcal{M}$ since each $\mathcal{C}_p$ is a vector space. It is likely, but not proved, that the $(-)$-mixture of states is continuous in the topology we have defined here.

A case between operator bounded and form bounded is ϵ-bounded:

$$\|V\|_\epsilon := \|R_0^{1/2-\epsilon} V R_0^{1/2+\epsilon}\|_\infty < \infty, 0 \leq \epsilon \leq 1/2. \tag{52}$$

This is the analogue of the Cramer class, since we prove that Z is an analytic function of V in this case.

Araki proved that if V is bounded, the Kubo-Mori expansion converges:

$$\log Z_V = \sum_{n=0}^{\infty} (n!)^{-1} \int_0^1 \prod d\alpha_i \delta(\sum \alpha_i - 1) K_n \tag{53}$$

where

$$K_n := \operatorname{Tr} (\rho^{\alpha_1} V \ldots \rho^{\alpha_n} V) . \tag{54}$$

We prove (with Grasselli) that the series converges also for ϵ- bounded perturbations, and that the $\|V\|_\epsilon$ are equivalent on overlapping regions. We now give an outline of the method.

We need an economical estimate for the n-Kubo function. If V were bounded, we could use the Hölder inequality for traces, with $p_i = 1/\alpha_i$ using that $\sum \alpha_i = 1$:

$$|\mathrm{Tr}\,[\rho^{\alpha_1} V_1 \dots \rho^{\alpha_n} V_n]\,| \leq \mathrm{Tr}\,\rho \|V_1\|_\infty \dots \|V_n\|_\infty . \tag{55}$$

We do better, since there is $\beta < 1$ such that ρ^β is of trace class, so we can replace ρ by ρ^β. We can thus borrow $\rho^{(1-\beta)\alpha_j}$ to help bound the potentials. Also, as $\sum \alpha_j = 1$, the region of integration is the (overlapping) union of regions S_j where $\alpha_j \geq 1/n$. By cyclicity, we may take $j = n$. We then write $\rho^{\alpha_j} V_j$ as

$$\dots \left[\rho^{\alpha_j \beta}\right] \left[H^{1-\delta_{j-1}+\delta_j} \rho^{(1-\beta)\alpha_j}\right] \left[R^{\delta_j} V_j R^{1-\delta_j}\right] \dots \tag{56}$$

The dots are factors taken with other terms. We bound the middle [...] by the spectral theorem, arranging the parameters δ_j so that we get an integrable function of α_j in S_n, $1 \leq j \leq n-1$. We bound the final [...] using the ϵ-boundedness of V, by a suitable choice of the δ_j. We end up with a factorial bound on the n-point function, so the series converges as a geometric series.

The manifold can be furnished by a real-analytic structure, by asserting that the ring of germs of analytic functions on the manifold consists of functions that are analytic in these analytic directions. The mixture coordinates η are examples of analytic functions; we say that we have an *analytic parametrisation* of the manifold by η. It remains to prove that the ξ are analytic functions of η, before we can say that η are analytic coordinates.

7 Singular perturbations

Every point of our manifold has some directions in its tangent space that remain within $\mathcal{M}$ but are not analytic directions. Consider the anharmonic oscillator,

$$H = (p^2 + q^2)/2 + \lambda q^{2n}, \qquad \lambda > 0. \tag{57}$$

It is known that $\exp -\beta H$ is of trace-class for all $\beta > 0$, so these states are in $\mathcal{M}$. It is also known that there is a singularity at $\lambda = 0$. Our result shows that if we start at $\lambda > 0$ then there is a region around this state where the manifold has analytic directions. Obviously, any point in $\mathcal{M}$ has many analytic directions: the bounded perturbations, provide many such. The metric is finite in a much wider class of directions: if ρ^β is of trace-class, and V is a form such that $\rho^\delta V$ is bounded for $\delta = (1-\beta)/2$, the a regularised *BKM* metric in the V-direction is finite at ρ.

The natural class of states, the analogue of the Orlicz space of [23], is the set $\mathcal{M}_{\max}$ of states of finite entropy. The natural class of states σ in a 'hood of a state ρ of finite entropy consists of states of finite entropy whose entropy relative to ρ is also finite. This 'hood will consist of many non-analytic perturbations of ρ. It is known that the -1-mixture (the usual mixture) of states of finite entropy has finite entropy, so $\mathcal{M}_{\max}$ has the -1-affine structure. Here is a simple proof.

Theorem 1

$$\begin{aligned} S(\lambda\rho + (1-\lambda)\sigma) &\leq \lambda S(\rho) + (1-\lambda)S(\sigma) \\ &+ \lambda \log(1/\lambda) + (1-\lambda)\log(1/(1-\lambda)). (58) \end{aligned}$$

Proof.
$-\log x$ is an operator monotone decreasing function. Since $\lambda\rho + (1-\lambda)\sigma \geq \lambda\rho$, we have

$$-\log(\lambda\rho + (1-\lambda)\sigma) \leq -\log(\lambda\rho).$$

Hence

$$-\lambda\rho.\log(\lambda\rho + (1-\lambda)\sigma) \leq -\lambda\rho.\log(\lambda\rho).$$

Similarly

$$-(1-\lambda)\log(\lambda\rho + (1-\lambda)\sigma) \leq -(1-\lambda)\sigma\log((1-\lambda)\sigma).$$

Adding, gives

$$\begin{aligned} S(\lambda\rho + (1-\lambda)\sigma) &\leq -\lambda\rho.(\lambda\rho) - (1-\lambda)\sigma.\log((1-\lambda)\sigma) \\ &= \lambda S(\rho) + (1-\lambda)S(\sigma) + \lambda\log(1/\lambda) \\ &\quad +(1-\lambda)\log(1/(1-\lambda)) < \infty. \end{aligned}$$

So the space $\mathcal{M}_{\max}$ of density matrices of finite entropy is a (-1)-affine space.

In [26] we propose a Luxemburg norm for the tangent space at a point $\rho \in \mathcal{M}_{\max}$. We expect that a 'hood of a point ρ will consist of all states $\sigma \in \mathcal{M}_{\max}$ having finite relative entropy, thus: $S(\sigma|\rho) := \rho.(\log\rho - \log\sigma) < \infty$.

Acknowledgements
It is a pleasure to thank M. Ohya for the invitation to the conference, H. Araki for discussions, and H. Hasegawa for arranging the trip.

References

1. Amari, S.-I., **Differential Geometric Methods in Statistics**, *Lecture Notes in Statistics*, **28**, 1985. Springer-Verlag.
2. Araki, H., *Publ. RIMS*, **9**, 165-209, Kyoto, 1968.
3. Balian, R., Y. Alhassid and H. Reinhardt, 'Dissipation in many-body systems: a geometrical approach based on information theory', *Phys. Reports*, **131**, 1-146, 1986.
4. Brody, D. C., and L. P. Hughston, *Phys. Lett.* **77**, 2851-, 1996.
5. Chentsov, N. N., **Statistical Decision and Optimal Inference**, *Nauka*, Moscow, 1972; in Russian. English version, *Amer Math Soc. Translations*, **53**, 1982.
6. Dawid, A., 'Discussion of a paper by Bradley Efron', *Ann. Stat.*, **3**, 1231-1234, 1975. 'Further comments on a paper by Bradley Efron', *Ann. Stat.*, **5**, 1249, 1977.
7. Efron, B. 'Defining the curvature of a statistical problem', *Ann. Stat.*, **3**, 1189-1242, 1975. 'The geometry of exponential families', *Ann. Stat.*, **5**, 457-458, 1977.
8. Fisher, R. A., 'Theory of statistical estimation', *Proc. Camb. Phil. Soc.*, **22**, 700-725, 1925.
9. Gibilisco, P., and G. Pistone, 'Connections on non-parametric statistical manifolds by Orlicz space geometry', *Infinite-dimensional Anal., Quantum Prob., and Related Topics*, **1**, 325-347, 1998.

10. Gibilisco, P., and T. Isola, 'Connections on statistical manifolds of density operators by geometry of non-commutative L^p-spaces, *Infinite-dimensional Analysis, Quantum Probability and Related Topics*, **2**, 169-178, 1999.

11. Grasselli, M., and R. F. Streater, 'The quantum info manifold for epsilon-bounded forms', *Reports on Math. Phys.*, **46**, 325-335, 2000; Los Alamos Archive Math-ph/9910031.

12. 'The Uniqueness of the Chentsov Metric', to appear in *Inf. Dim. Anal.Quant. Prob.*

13. Hasegawa, H. *Reps. on Math. Phys*, **33**, 87-, 1993. 'Noncommutative extension of the information geometry', pp 327-337 in **Quantum Communication and Measurement**, eds. V. P. Belavkin, O. Hirota and R. L. Hudson, Plenum Press, N. Y. 1995.

14. Hasegawa, H., and D. Petz, 'Non-commutative extension of information geometry II', 109-118 in **Quantum Communication, Computing and Measurement**, Eds. O. Hirota et al., Plenum Press, N. Y. 1997.

15. Helstrom, C. W., **Quantum Detection and Estimation Theory**, Academic Press, N. Y., 1976.

16. Ingarden, R., Y. Sato, K. Sagura, and T. Kawaguchi, 'Information thermodynamics and differential geometry' *Tensor*, **33**, 347-353, 1979.

17. Kossakowski, A., 'On the quantum informational thermodynamics', *Bull. acad. polonaise des sciences*, **17**, 263-267, 1969.

18. Nagaoka, H., 'Differential aspects of quantum state estimation and relative entropy', in **Quantum Communication and Measurement**, eds. V. P. Belavkin et al., Plenum Press, 1995.

19. Ohya, M. and D. Petz, **Quantum Entropy and its Use**, Springer-Verlag, 1993.

20. Petz, D. 'Monotone Metrics on Matrix Spaces', Lin. Alg. Appl. **244**, 81-96, 1996.

21. Petz, D., and G. Toth, 'The Bogoliubov inner product in quantum statistics', *Lett. in Math. Phys.*, **27**, 205-216, 1993.

22. Petz, D., and C. Sudar, 'Geometries of quantum states', *J. Mathematical Phys.*, **37**, 2662-2673, 1996.

23. Pistone, G., and C. Sempi, 'Infinite-dimensional geometric structure on the space of all probability measures equivalent to a given one', *Annals of Statistics*, **33**, 1543-1561, 1995.

24. Rao, C. R., 'Information and accuracy attainable in the estimation of statistical parameters', *Bull. Calcutta Math. Soc.*, **37**, 81-91, 1945.

25. Streater, R. F., 'Gas of Brownian particles in a potential', *J. Stat. Phys.*, **88**, 447-, 1997. 'Information geometry and reduced quantum description', *Reports on Math. Phys.*, **38**, 419-436, 1996. 'A model of dense liquids', *Banach Center Publications*, **43**, 381-393, Warsaw, 1998. 'Onsager relations in statistical dynamics' *Open Systems and Info. Dyn*, **6**, 87-100, 1999. 'The Soret and Dufour effects in statistical dynamics', *Proc. Roy. Soc.*, **456**, 205-221, 1999.

26. Streater, R. F. 'The information manifold for relatively bounded potentials', to appear in the Bogoliubov Memorial Volume, ed. A. A. Slavnov, Steklov Institute, Moscow; 2000. Los Alamos Archive Math-ph 9910035.

27. Streater, R. F.,'The analytic quantum info manifold', to appear in **Stochastic Processes, Physics and Geometry**, eds. F. Gesztesy, S. Paycha and H. Holden; Canad. Math, Soc., 2000. Los Alamos Archive Math-ph/9910036.

28. Uhlmann, A., 'The metric of Bures and the geometric phase' 267-274, in **Groups and Related Topics**, eds. R. Gielerak et al., Kluwer, 1992. 'Density operators as an arena for differential geometry', *Reps. Math. Phys.*, **33**, 253-263, 1993.