Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Monotone invariants and embeddings of statistical manifolds

by

Hông Vân Lê

Preprint no.: 43 2003

Monotone invariants and embeddings of statistical manifolds

Hông Vân Lê
Max-Planck-Institute for Mathematics in Sciences
Inselstraße 22-26
D-04103 Leipzig

May 5, 2003

Abstract

In this note we prove certain necessary and sufficient conditions for the existence of embedding of statistical manifolds. In particular we prove that any C^3-bounded smooth statistical manifold can be embedded into a space of probability measures on the finite set and any statistical manifold can be embedded into the space of probability measure on the infinite set. As a result we get positive answers to the question of Amari on the existence of embedding of exponential families and to the Lauritzen question on realization of statistical manifolds as statistical models.

CONTENTS

1 Introduction

In the program of geometrization of mathematical statistics one likes to understand geometric structures on families of probability distributions. These structures must not only be natural in mathematical categorical languages, but they must also have a clear statistical meaning. Such structures are the Fisher metrics and the Centsov-Amari connections which are of fundamental role in the information geometry - the established domain in the geometrization program of mathematical statistics.

The Fisher metric was invented by Rao [R1] and has been systematically studied by Chentsov [C], Morozova-Centsov [M-C], Amari [Am-N] and others ([L], [R2], [Ay], [J], ect.) in the field of geometric aspects of statistics and information theory. The Fisher metric is a Riemannian metric on a family M of probability measures which can be considered as a differentiable manifold. Such a family is called **a statistical model.** Centsov and Amari independently also dicovered a natural structure on statistical models, namely a 1-parameter family of invariant connections which includes the Levi-Civita connection of the Fisher metric. This family of invariant connections is defined by a 3-symmetric tensor T together with the Levi-Civita connection of the Fisher metric. Thus we shall called a Riemannian manfold (M, g) with a 3-symmetric tensor T **a statistical manifold**. Since two 3-symmetric tensors T and $k \cdot T$, $k \neq 0$, define the same family of Centsov-Amari connections, we shall say that two statistical manifolds (M, g, T) and (M, g, kT) are conformal equivalent.

A natural and important question in the mathematical statistics is to understand, if a given family M of probability distributions can be considered as a subfamily of another given one N. In our geometric language it can be formulated as a problem of embedding of a statistical manifold (M, g, T) into another one (N, g', T').This problem includes the Lauritzen question [L], if any statistical manifold is a statistical model. It also concerns the following important problem posed by Amari [Am], if any finite dimensional statistical model can be embedded into the space Cap^N of probability distributions of the sample space Ω^N of N elementary events for some finite N. First we note that if we consider statistical models equipped only with the Fisher metric, then the answer is positive. Actually it is a simple consequence of the Nash embdeding theorem (see Proposition 5.1). But the problem is more complicated, if we consider also the Fisher metric together with a

3-symmetric tensor T. We shall construct a class of invariants of statistical manifolds which present obstructions to embedding of given statistical manifolds. These invariants measure certain relations between the metric tensor g and the 3-symmetric tensor T. In particular using these inviariants we show that no statistical manifold which is conformal equivalent to the space Cap^N can be C^1-embedded into the product of m copies of the normal Gaussian manifolds for any $N > 3$ and any finite m. (This example points out the difference between our embedding problem with the embedding of Riemannian manifolds, where the large size of the target manifold is a sufficient condition for the existence of an embedding.) We also prove that any smooth statistical manifold M^m can be embedded to a the space Cap^N for $N = 4(m+1)[(2m(m+1)-1)m(m+1)+2+m)+(m+2)(m+3)]$ (see Theorem 5.2). If M^m is compact (or C^3-bounded), then we can lower N by dividing factor $(m+1)$. Thus we can say that any statistical manifold is a statistical model. As a consequence we get a new proof of Matumoto theorem on the existence of the contrast function for a statistical manifold.

Acknowledgement. I thank Nyhat Ay and Jürgen Jost for introducing me into the information geometry and many helpful discussions on the objects. My thanks are also due to Qing Han and Kaoru Ono for stimulating discussions.

2 Fisher metric, Centsov-Amari connections and potential functions.

In this section we recall the notions of statistical models (and statistical manifolds), their Fisher metric and the Centsov-Amari connections on these spaces. We show the existence of hiarachy of different (weak) potential functions for statistical models (resp. statistical manifolds). At the end of the section we discuss the problem, if a given statistical manifold is a statistical model.

The notion of statistical models and their Fisher (information) metric was introduced by Rao [R1].

A statistical model is a (possibly immersed) submanifold M (a family) in the space $Cap(\Omega)$ of all probability measures on a sample space Ω which is usually (and in our note) assumed to be a differentiable manifold. A statistical model is usually considered with the **Fisher metric** $g^F(x)$ defined as follows. For any $V, W \in T_xM$ we put

$$g^F(V,W)_x = E_x((\partial_V \ln p(x,\omega))(\partial_W \ln p(x,\omega)) =$$

$$\int_\Omega (\partial_V \ln p(x,\omega))(\partial_W \ln p(x,\omega))p(x,\omega), \tag{2.1}$$

where E_x denotes the expectation w.r.t the measure $p(x,.)$. For simplicity we shall consider families of Borel measures on a finite dimensional manifold Ω, and if $\dim \Omega \geq 1$ then our family must consist of mutually absolutely continous measures. In these cases we represent the measure $p(x,\omega)$ as a nonnegative density function on a discrete sample space Ω^N of N elementary events, or as $p(x,\omega)d\omega$ with $p(x,\omega)$ considered as an almost everywhere positive density function and with $d\omega$ being a specified Borel measure. The function under integral in (2.1) is well defined, if $p(x,\omega)$ is positive. Thus we have two conditions on $p(x,\omega)$, namely

$$p(x,\omega) > 0 \, \forall (x,\omega) \in M \times \Omega, \tag{2.1.a}$$

$$\int_\Omega p(x,\omega)\, d\omega = 1 \, \forall x \in M. \tag{2.1.b}$$

Since the density function $p(x,\omega)$ defines uniquely the mapping $M \to Cap(\Omega)$ we shall call $p(x,\omega)$ **a probability potential of the metric** g_F. We shall see in Proposition 2.2 that for a given Riemannian metric g_F on a smooth manifold M there exist many probability potentials $f(x,\omega)$ for g_F even if we fix the space $(\Omega, d\omega)$.

Some time it is useful to consider functions $p(x,\omega)$ which satisfy (2.1) and (2.1.a) but not neccesary (2.1.b). In this case the Riemannian metric g^F will be called **weak Fisher metric** and the function $p(x,\omega)$ will be called **a weak probability potential** of g^F.

If $\Omega = \Omega_N$ - the set of N elementary events, then (2.1) becomes

$$g^F(V,W)(x) = \sum_{i=1}^N p_i(x) \frac{\partial_V p_i(x)}{p_i(x)} \frac{\partial_W p_i(x)}{p_i(x)} =$$

$$= \sum_{i=1}^N \frac{1}{p_i(x)} \partial_V p_i(x) \partial_W p_i(x). \tag{2.1.1}$$

Before considering some examples of Fisher metrics and weak probability potentials which shall be used later in this note, we list some simple properties of the Fisher metric as well as of weak Fisher metrics.

2.2. Proposition. (see also [A-N], [C], [J]). *a) We denote by V and W two vector fields on M. The Fisher metric (2.1) can be computed by the following formula:*

$$(2.2.a) \qquad g^F(V,W)_x = 4\int_\Omega (\partial_V \sqrt{p(x,\omega)})(\partial_W \sqrt{p(x,\omega)})d\omega.$$

The formula holds also for weak Fisher metrics.

b) For a statistical model with parameters (x_i) in an open domain of $\mathbf{R}^n$ we can compute its Fisher metric via the following formula

$$(2.2.b) \qquad g^F_{ij}(x) = \int_\Omega \frac{\partial^2 \ln p(x,\omega)}{\partial_i \partial_j} p(x,\omega)\, d\omega.$$

c) If $\mathcal{D}$ is a diffeomorphism of Ω then the function $p(x,\mathcal{D}(\omega))$ is also a probability potential of the Fisher metric g^F, or generally, $p(x,\mathcal{D}(\omega))$ is a weak potential for weak Fisher metric g^F.

d) Suppose that (M,g) is a Riemannian manifold which admits a weak probability potential $P(x,\omega)$ of g and $(N, g_{|N})$ is a Riemannian submanifold of M. Then the restriction of P_g to $N\times\Omega$ is a weak probability potential for the induced metric $g_{|N}$.

Proof. The first formula (2.2.a) and the third statement follow from the rule of integration under the change of variables (see [A-N], [C], [J].) Thus they also hold for weak probability potentials. The second formula (2.2.b) is a consequence of (2.1.b) and integration by part.

The last statement follows immediately from the definition (2.1). □

2.3. Examples.

2.3.1. The standard Euclidean metric g^0 on the positive quadrant $\mathbf{R}^N_+(x_i > 0)$ admits a weak probability potential $\{p_i(x) = \frac{1}{4}x_i^2, i = \overline{1,N}.\}$. Indeed we have from (2.2.a)

$$g^F(V,W)_{(x_1,\cdots,x_N)} = \sum_{i=1}^N (\partial_V x_i)(\partial_W x_i) = \sum_{i=1}^N V^i W^i = g^0(V,W)$$

2.3.2. We denote by $(Cap^N)_+$ the space of positive probability distributions on Ω^N (see also [A-N],[J], [C]) Then $((Cap^N)_+, g^F)$ can be identified with the positive quadrant $S^{N-1}_+(2)$ of the sphere $S^{N-1}(2)$ of radius 2 with the standard metric of constant curvature as follows. By definition we have

$$Cap^N_+ := \{(p_1, \cdots, p_n) | p_i > 0 \text{ for } i = \overline{1,n} \,\&\, \sum p_i = 1\}.$$

We define the embedding map

$$f : Cap^N_+ \to S^{N-1}(2)$$

$$(p_1, \cdots, p_N) \mapsto (q_1 = 2\sqrt{p_1}, \cdots, q_N = 2\sqrt{p_N}).$$

The Fisher metric in the new coordinates (q_i) is

$$g_{(q_1,\cdots,q_N)}(V,W) = \sum_{i=1}^{N} (\partial_V q_i)(\partial_W q_i).$$

Comparing this with (2.3.1) and using (2.2.d) we see immediately that the Fisher metric on $(Cap^N)_+ = S^{N-1}_+(2)$ is the standard metric of constant postive curvature on the sphere $S^{N-1}(2)$. This metric is of course can be extended smoothly on the whole space Cap^N.

2.3.3. The exponential family (see [A-N], [J], [L]): $M = \mathbf{R}^k(\theta)$ has the probability potential

$$p(\theta, \omega) = \exp[C(\omega) + \theta^i f_i(\omega) - \phi(\theta)],$$

here

$$\phi(\theta) = \ln \int_\Omega \exp[C(\omega) + \sum_{i=1}^{n} \theta^i f_i(\omega)] d\omega.$$

The exponential family includes the following Gaussian models.

2.3.3.a. The normal distribution (also the univariate Gaussian model): $M^2 = \mathbf{R}^2(\mu, \sigma)$, $(\Omega, d\omega) = (\mathbf{R}(x), dx)$.

$$p(\mu,\sigma)(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$$

Here we put $\omega = x$ and new coordinates

$$\theta^1 = \frac{\mu}{\sigma^2}, \theta^2 = -\frac{1}{2\sigma^2}.$$

In these coordinates (x,θ) we have

$$C(x)=0,\ f_1(x)=x,\ f_2(x)=x^2,$$

$$\phi(\theta)=\frac{\mu^2}{2\sigma^2}+\ln(\sqrt{2\pi}\,\sigma)=-\frac{(\theta^1)^2}{4\theta^2}+\frac{1}{2}\ln\frac{-\pi}{\theta^2}.$$

2.3.3.b. The inverse Gaussian model: $M=\mathbf{R}^2_+(\chi,\psi)$, $(\Omega,d\omega)=(\mathbf{R}_+,\,dx)$.

$$p(\chi,\psi,x)=\sqrt{\frac{\chi}{2\pi}}x^{3/2}\exp(\sqrt{\chi\psi}-\frac{1}{2}(\frac{\chi}{x}+\chi x)).$$

Using (2.2.b) we get the following expression for the Fisher metric of an exponential family (2.3.3)

$$g_{ij}(x)=\frac{\partial^2\phi(x)}{\partial_i\partial_j}.$$

In coordinates (μ,σ) of the normal distribution we have ([L])

$$g(\partial_\mu,\partial_\mu)=\frac{1}{\sigma^2},$$
$$g(\partial_\mu,\partial_\sigma)=0,$$
$$g(\partial_\sigma,\partial_\sigma)=\frac{2}{\sigma^2}.$$

2.3.4. Poissonsverteilung: $M=\mathbf{R}^+(\xi)$, $\Omega=\{0,1,2,\cdots\}\ni x$.

$$p(x,\xi)=e^{-\xi}\frac{\xi^x}{x!}$$

2.4. Divergence potential. (see [A-N], [R2]) A (weak) Fisher metric (2.1) on M can be derived from a divergence function ρ on $M\times M$, i.e. a function ρ with the following properties

(2.4.1) $\qquad \rho(x,y)\ge 0$ with equality iff $x=y$

A divergence function ρ is called **a divergence potential** for a metric g, if $g(x)$ coincides with the the restriction $Hess(\rho)_{|(x,x)}$ on the first factor $i_1(TM)$

in the orthogonal decomposition $T_{(x,x)}(M,M) = (T_xM,0) \oplus (0,T_xM) = (i_1(T_xM)) \oplus (i_2(T_xM))$.

$$g(X,Y)_x = Hess(\rho)(i_1(X), i_1(Y)). \tag{2.4.2}$$

An example of a divergence potential for a Fisher metric is the Jensen function $J_H^{\lambda,\mu}(x,y)$ of the entropy function $H(x)$ on M or a Kullback relative entropy function $K(x,y)$ on $M \times M$. We recall that the (Shannon's) entropy function $H(x)$, the Jensen function $J(x,y)$ and the Kullback relative entropy function K on $Cap(\Omega)$ are defined as follows

$$H(x) := -\int_\Omega \ln p(x,\omega)\partial(x,\omega)\, d\omega,$$

$$J_H^{\lambda,\mu}(x,y) = H(\lambda x + \mu y) - \lambda H(x) - \mu H(y) \text{ for } \lambda \cdot \mu \geq 0, \lambda + \mu = 1,$$
$$K(x,y) := \int_\Omega [\ln p(y,\omega) - \ln p(x,\omega)]\, p(y,\omega)\, d\omega.$$

To see that $H^{\lambda,\mu}$ and K are divergence potentials for the associated Fisher metric we apply formula (2.2.b.

We summarize these relations in the following diagram

$$\begin{array}{l} \\ weak \\ \\ Fisher \\ \\ tensor \\ \\ \end{array} \left(\begin{array}{r} \longleftarrow \qquad \{\textsf{weak probability potential } f(x,\omega) : f > 0\} \\ \cup \\ \leftarrow \{\textsf{probability potential } f(x,\omega) : f > 0 \,\&\, \int_\Omega f(x,\omega) = 1\} \\ \downarrow \textsf{(relative) entropy} \\ \{\textsf{divergence potential } \rho(x,y) : \rho \geq 0 \,\&\, \rho(x,x) = 0\} \\ \downarrow \textsf{taking Hessian at the diagonal} \\ - - - - - - - - - - - - - - - - \rightarrow \{\textsf{Riemannian metrics}\}. \end{array} \right.$$

2.5. Centsov-Amari connections. Let $p(x,\omega)$ be a probability potential for a Riemannian metric g. We define a symmetric 3-tensor T on M as follows

$$T(X,Y,Z) = \int (\partial_X \ln p(x,\omega))(\partial_Y \ln p(x,\omega))(\partial_Z \ln p(x,\omega))\, p(x,\omega) = \tag{2.5.1}$$

$$= \int_\Omega \frac{1}{p^2}(\partial_X p(x,\omega))(\partial_Y p(x,\omega))(\partial_Z p(x,\omega))\, d\omega =$$
$$= 8\int_\Omega \frac{1}{\sqrt{p}}(\partial_X \sqrt{p(x,\omega)})(\partial_Y \sqrt{p(x,\omega)})(\partial_Z \sqrt{p(x,\omega)})\, d\omega.$$

We denote by ∇^F the Levi-Civita connection of the (weak) Fisher metric g^F. We define

$$< \nabla^t_X Y, Z > := < \nabla^F_X Y, Z > + t \cdot T(X, Y, Z). \tag{2.5.2}$$

The connections ∇^t are called the Centsov-Amari connections.

2.6. A statistical manifold (M, g, T) is a Riemannian manifold (M, g) equipped with a 3-symmetric tensor S (see [L]). The symmetry of the tensor T is motivated by the formula for T in (2.5.1). If $p(x, \omega)$ is a (weak) probability potential for the metric g, and T is defined from $p(x, \omega)$ by (2.5.1), we shall say that $p(x, \omega)$ is **probability potential** for (M, g, T). We shall say that two statistical manifolds (M, g, T) and (M', g', T') are **conformal equivalent**, if there is number $t \neq 0$ and a diffeomeorphism $f : M \to M'$ such that $f(g') = g, f(T') = tT$. This condition is sufficient and neccesary in order two manifolds M and M' are not only isometric but they have the same family of Centsov-Amari connections defined by (2.5.2). The symmetry of T implies that all these connections are torsion-free. Furthermore it is easy to check that for any statistical manifold (M, g, T) the connection ∇^t and the connection ∇^{-t} are dual in the sense

$$\nabla_Z < X, Y > = < \nabla^t_Z X, Y > + < X, \nabla^{-t}_Z Y >$$

for any vector fields X, Y, Z on M. Thus any statistical manifold (M, g, T) has a natural **dual structure** $(<, >, \nabla^1, \nabla^{-1})$ as defined by Amari [A-N]. Conversely suppose that we are given a 3-tensor T and the connections ∇^t defined by (2.5.2). Then the condition that ∇^t are torsion free together with the duality condition of ∇^1 and ∇^{-1} implies that T is 3-symmetric. If we are only interested in the geodesics of the connections ∇^t, then we can symmetrize T (if T is not 3-symmetric) in two variables to get new torsion free connections with the same geodesics (see Prop. 7.9. Chapter III in [K-N]).

A submanifold N in a statistical manifold (M, g, T) with the induced Riemannian metric $g_{|N}$ and induced tensor $T_{|N}$ is called **statistical submanifold** of (M, g, T). Clearly if $f(x, \omega)$ is a (weak) probability potential for (M, g, T), then its restriction to any submanifold $N \subset M$ is a (weak) probability potential for the induced statistical structure.

Now let us compute the tensor T for the Riemannian manifolds with (weak) Fisher metrics in examples 2.3.1, 2.3.2, 2.3.3.

- The standard Euclidean metric in (2.3.1) with the weak potential function $\{\frac{1}{2}x_i^2, i = 1, N\}$. A direct computation shows that

(2.6.1)
$$T_{ijk}(x_1, \cdots, x_N) = \delta_{ijk}\frac{1}{x_i}.$$

- The exponential family in (2.3.2). Our computation here can be done by noticing that the connection ∇^1 is flat, i.e. its torsion and curvature tensors vanish (see [A-N], [J], [L]). Moreover in the coordinates (θ_i), for which (2.3.2) holds, all the Christoff symbols Γ^1_{ijk} vanish. In other words $\{\theta_i\}$ are affine coordinates for ∇^1 and $\partial\theta_i$ are parallel vector fields. Thus our tensor T coincides with the Christoff symbols Γ^0_{ijk} of the Fisher metric.

(2.6.2)
$$T_{ijk} = \Gamma^0_{ijk} = \frac{1}{2}(g_{ij,k} + g_{jk,i} - g_{ij,k}) = \frac{1}{2}\partial_i\partial_j\partial_k\phi.$$

Applying the formula (2.6.2) to the normal distribution 2.3.3.a (the univariate Gaussian manifold) we get (see [L])

(2.6.3)
$$T(\partial_\mu, \partial_\mu, \partial_\mu) = 0 = T(\partial_\mu, \partial_\sigma, \partial_\sigma),$$

(2.6.4)
$$T(\partial_\mu, \partial_\mu, \partial_\sigma) = \frac{2}{\sigma^3}, \; T(\partial_\sigma, \partial_\sigma, \partial_\sigma) = \frac{8}{\sigma^3}.$$

2.6.5. Remark. ([A-N, [Ma]) Any divergence function $\rho(x, y)$ on $M \times M$ defines a tensor T on M via the following formula
$(2.6.5.a)$
$$T(X, Y, Z)_x = -\partial_{i_2(Z)}Hess(\rho)(i_1(X), i_1(Y))_{(x,x)} + \partial_{i_1(Z)}Hess(\rho)(i_2(X), i_2(Y))_{(x,x)}.$$

Here i_1, i_2 is defined as in 2.4 and we use the same notation $i_j(X), i_j(Y)$ for any local vector field on $M \times M$ which coincides with $i_j(X), i_j(Y)$ at (x, x). It is easy to check that the LHS of (2.6.5) is a tensor [E,M]. If g and T are defined by the same divergence function $\rho(x, y)$ via (2.4.2) and (2.6.5.a) we shall call $\rho(x, y)$ **a divergence potential for the statistical manifold** (M, g, T). It is a known fact that the Kullback relative entropy funtion is a divergence potential for the associated statistical model. In a general a tensor T defined from a divergence function $\rho(x, y)$ will defined a family of dual connections ∇^t (see above) which are not neccesary torsion free. That is the case with the divergence function in quantum statistics (see section 6). The non-symmetry of a tensor T is a clear obstruction for embedding the corresponding statistical manifold into a classical (Lauritzen) statistical manifold.

2.7. Statistical models and statistical manifolds.

Since any probability function $p(x, \omega)$ defines a map $M \to Cap(\Omega)$, we shall say that a statistical manifold (M, g, T) is a statistical model, if there probability potential $p(x, \omega)$ for g and T. In particular, if a statistical manifold (M, g, T) is a statistical model, then it must admit a divergence potential.

2.7.1 Theorem. ([Ma]) *For any statistical manifold (M, g, T) there exists a divergence potential ρ for g and for T.*

In the section 5 we shall prove that any bounded statistical manifold (in particular any compact statistical manifold) can be embedded in the space Cap_+^N for some finite N. We can also prove that any statistical manifold can be embedded into the space Cap_+^∞. Thus any statistical manifold is a statistical model. This gives a positive answer to Lauritzen question [L]. In this way we also get a new proof of the Matsumoto theorem 2.7.1 [Ma].

3 Embeddings of linear statistical spaces.

An Euclidean space $(\mathbf{R}^n, g^0)$ equipped with a 3 -symmetric tensor T will be called **a linear statistical spaces.** We observe that the problem of equivalence of linear statistical spaces is the problem of equivalence of 3-symmetric tensors T under the action of the orthogonal group $O(n)$. This is a problem of the classical theory of invariants, and in principle it can be completely solved. In this section we discuss the geometry of these invariants and we show several neccesary and sufficient conditions for the existence of embedding of one linear statistical space into another linear statistical space.

First we note that by dimension argument the space $\mathcal{R}^n$ which consists of the following 3-symmetric tensors

(3.1)

$$T^v(x, y, z) = < v, x >< y, z > + < v, y >< x, z > + < v, z >< x, y >,$$

is an irreducible component of the $SO(n)$-action on $S^3(\mathbf{R}^n)$. It is known (see e.g. [O-V]) that if $n \geq 3$, the action of $SO(n)$ on $S^3(\mathbf{R}^n) \otimes \mathbf{C}$ has two irreducible components

$$S^3(\mathbf{R}^n)\otimes\mathbf{C} = (\mathcal{R}(3\pi_1)\otimes\mathbf{C})\oplus(\mathcal{R}^n\otimes\mathbf{C}), \tag{3.2}$$

where π_1 denotes the first fundamental weight corresponding to the identity representation of $SO(n)$. By dimension argument, we conclude that the space $S^3(\mathbf{R}^n)$ of the $SO(n)$ action has also two irreducible (real) components:

$$S^3(\mathbf{R}^n) = \mathcal{R}(3\pi_1)\oplus\mathcal{R}^n. \tag{3.2.1}$$

3.3. Remark. We can characterize the space $\mathcal{R}^n$, $n \geq 3$, as the space of all 3-symmetric tensors T such that

$$T(x,y,z) = 0, \text{ if } <x,y>= 0 \,\&\, <y,z>= 0 \,\&\, <z,x>= 0. \tag{3.3.a}$$

(It suffices to take into account the fact that any T^v satisfies (3.3.a).)

To compute the orthogonal projection of a 3-symmetric tensor T on the space $\mathcal{R}^n$ in the decomposition (3.2.1) we can use the following Lemma. We denote by π_2 the orthogonal projection form $S^3(\mathbf{R}^n)$ to $\mathcal{R}^n$

3.4. Lemma. *We have*

$$\pi_2(S) = \frac{1}{n+2}T^{Tr(S)}. \tag{3.4.1}$$

Here we identify the 1-form $Tr(S)$ with a vector in $\mathbf{R}^n$ by using the Euclidean metric g^0.

Proof. We note that the component $\mathcal{R}(3\pi_1)$ in (3.1) is the image of the symmetrization map

$$S : (\mathbf{R}^n)\otimes(S^2_0(\mathbf{R}^n)) \to S^3(\mathbf{R}^n),$$

where $S^2_0(\mathbf{R}^n)$ denote the set of traceless symmetric bilinear forms on $\mathbf{R}^n$. Clearly the 1-form $Tr(S)$ vanishes, whenever $S \in \mathcal{R}(3\pi_1)$. Alternatively we notice that the linear map defined in RHS of (3.4.1) is $SO(n)$ equivariant map. The coefficient $1/(n+2)$ is obtained by computing $Tr(T^v)$. □

Thus we shall call any tensor $T \in \mathcal{R}^n$ of **trace type.**

We note that

$$\dim S^3(\mathbf{R}^n) = C_n^3 + 2C_n^2 + n = \frac{n(n+1)(n+2)}{6}. \tag{3.4.2}$$

Thus the dimension of the quotient $S^3(\mathbf{R}^n)/SO(n)$ is at least $C_n^3+C_n^2+n$. A direct computation shows that the dimension of the orbit $SO(n)([\sum_{i=1}^n a_i v_i^3])$ is $C_n^2 = \dim SO(n)$, if $\Pi a_i \neq 0$. Here $\{v_i\}$ is an orthonormal basis in $\mathbf{R}^n$. Hence the dimension of $S^3(\mathbf{R}^n)/O(n) = C_n^3 + C_n^2 + n$. This dimension is exactly the number of complete invariants of pairs of a positive definite bilinear form g and a 3-symmetric tensor T.

Since the dimension of $G_k(\mathbf{R}^n) = k(n-k)$, it follows that generically it is impossible to embed a linear statistical space (R^k, g^0, T) into a given statistical linear space (R^n, g^0, T), unless $k(n-k) \geq C_k^3 + C_k^2 + k$. Clearly the dimension condition is not sufficient as the following proposition shows.

3.5. Proposition. *A linear statistical space $(\mathbf{R}^k, g^0, T)$ can be embedded into a linear statistical space $(\mathbf{R}^N, g^0, T^v)$, if and only if $N \geq k$ and T is also a trace type: $T = T^w$ with $|w| \leq |v|$.*

Proof. The neccesary condition follows from (3.1) which implies that the restriction of T^v on $\mathbf{R}^k$ equals $T^{\bar{v}}$, where $\bar{v}$ is the orthogonal projection of v to $\mathbf{R}^k$. Conversely, if $|w| \leq |v|$ we can find an orthogonal transformation, such that w equals the orthogonal projection of v on $\mathbf{R}^k$. □

There are several invariants of a 3-symmetric tensor T which behave well under linear embeddings. First we note that the metric g extends canonically on the space $S^3(\mathbf{R}^n)$. We then can define the absolute norm

$$||T|| := \sqrt{< T, T >}.$$

Next we define **comasses** of a 3-symmetric tensor T as follows

$$\mathcal{M}^3(T) := \max_{|x|=1,|y|=1,|z|=1} T(x,y,z),$$

$$\mathcal{M}^2(T) := \max_{|x|=1,|y|=1} T(x,y,y),$$

$$\mathcal{M}^1(T) := \max_{|x|=1} T(x,x,x).$$

Clearly we have

$$0 \leq \mathcal{M}^1(T) \leq \mathcal{M}^2(T) \leq \mathcal{M}^3(T) \leq ||T||.$$

3.6. Proposition. *The comasses are norms of T, i.e, if $\mathcal{M}^1(T) = 0$, then T vanishes. They are monotone invariants of T in a sense, that if T is a restriction of 3-symmetric tensor $\bar{T}$ on $\mathbf{R}^N$, then*

$$(3.6.1). \qquad ||T|| \le ||\bar{T}||,\ \mathcal{M}^i(T) \le \mathcal{M}^i(\bar{T}),\ \forall i = 1,2,3.$$

Proof. To prove the first statement we use the identity

$$-12T(x,y,z) = T(x+y+z, x+y+z, x+y+z) + T(x+y-z, x+y-z, x+y-z) +$$

$$+T(x-y+z, x-y+z, x-y+z) + T(-x+y+z, -x+y+z, -z+y+z) -$$

$$2(T(x,x,x) + -T(y,y,y) + T(z,z,z)).$$

The second statement follows immediately from the definition. □

Now for a space $(\mathbf{R}^n, g^0, T)$ and for $1 \le k \le n$ we put

$$\lambda_k(T) := \min_{\mathbf{R}^k \subset \mathbf{R}^n} \mathcal{M}^1(T_{|\mathbf{R}^k}).$$

We can easily check that if $\bar{T}$ is a restriction of T to a subspace $\mathbf{R}^m \subset \mathbf{R}^n$, then

$$\lambda_k(\bar{T}) \ge \lambda_k(T) \ge 0 \text{ for all } k \le m.$$

Thus $\lambda_k(T)$ is a monotone invariant of linear statistical manifolds. These invariants are related by the following inequalities

$$\mathcal{M}^1(T) = \lambda_n(T) \ge \lambda_{n-1}(T) \cdots \ge \lambda_2(T) \ge \lambda_1(T) = 0.$$

The last equality follows from the fact, that the function $T(x,x,x)$ is anti-symmetric on $S^{n-1}(|x| = 1) \subset \mathbf{R}^n$ and S^{n-1} is connected. We observe that if T is of trace type, then $\lambda_{n-1}(T) = \cdots = \lambda_1(T) = 0$.

We are going to give a lower bound on the monotone invariant λ_{n-1} of a linear statistical space of certain type. The equality $\lambda_{n-1}(\mathbf{R}^n, g^0, T) \ge A$ means that no hyperplan with the norm $\mathcal{M}^1$ strcitly less than A can be embedded in $(\mathbf{R}^{n-1}, g^0, T)$.

3.7. Lemma.*a) Let $T = \sum_{i=1}^n (N - \varepsilon_i)(x^i)^3$ be a 3-symmetric tensor on $\mathbf{R}^n$ with $n \ge 4$, $N \ge 4$ and $|\varepsilon_i| \le 1/4$. Then we have*

$$\lambda_{n-1}(T) \ge \frac{N}{\sqrt{10}} - 1/4.$$

b) Let $T = N\sum_{i=1}^{n}(x^i)^3$ and H a hyperplan in $\mathbf{R}^n$ which is orthogonal to $(kn, 1, 1, \cdots, 1)$ and $n \geq 5$, $k \geq 3$. Then we have

$$\lambda_{n-2}(T_{|H}) \geq \frac{N}{5} - 1.$$

c) Let $x = ((1-\varepsilon), \frac{1}{kn}, \cdots, \frac{1}{kn}) \in S^n(1) \subset \mathbf{R}^{n+1}$, where $n \geq 4, k \geq n$. We denote by H the tangential plan T_xS^n and we denote by T^0 the standard 3-symmetric tensor on $\mathbf{R}^{n+1}$ (see (2.6.1)). Then we have

$$\lambda_{n-1}(T^0_{|H}) \geq \frac{kn}{5} - 1.$$

Proof. a) We denote by $v_H = (v^1, \cdots, v^n)$ the unit vector which is orthogonal to H. Clearly v_H is defined uniquely up to sign. It suffices to show that there is a unit vector w such that

$$< w, v_H >= 0, \tag{3.7.1}$$

$$T(w, w, w) \geq \frac{N}{\sqrt{10}} - 1/4. \tag{3.7.2}$$

We consider two cases.

Case 1. We assume that not all the coordinates v^i are of the same sign. Without loss of generality we assume that $v^1 \leq 0$, $v^2 > 0$. Then we choose

$$w := A(|v^2|, |v^1|, 0, \cdots, 0), \text{ where}$$
$$A := (|v^1|^2 + |v^2|^2)^{-1}.$$

Clearly w is a unit vector and $w \in H$. It is easy to check that

$$T(w, w, w) \geq (N - 1/4)A^3(|v^1|^3 + |v^2|^3) \geq \frac{N - 1/4}{2} > \frac{N}{\sqrt{10}} - 1/4.$$

So in this case the equality (3.7.2) holds.

Case 2. We assume that all the coordinates v^i are the same sign. Without loss of generality we assume that $v^1 \geq v^2 \geq \cdots \geq v_n \geq 0$. If $v_{n-1} = v_n = 0$, then we can apply our argument in case 1 to get a unit vector w which satisfies (3.7.1) and (3.7.2)(namely our choosen vector w has zero coordinates

w^i, if $i \leq n-1$). So now, using the condition that $n \geq 4$, we shall assume that $v_1 \geq v_2 \geq v_3 > 0$. We set

$$w := (-a, -a, \lambda \cdot a, 0, \cdots, 0), \text{ where}$$

$$\lambda := \frac{v^1 + v^2}{v^3} \geq 2, \text{ and}$$

$$a > 0 \text{ such that } a^2(\lambda^2 + 2) = 1.$$

Clearly w is a unit vector and $w \in H$. A straighforward computation gives us

$$T(w, w, w) \geq a^3(\lambda^3(N - 1/4) - 2(N + 1/4)) \geq$$

$$\frac{1}{\sqrt{(\lambda^2+2)^3}}((N - 1/4)(\lambda^3 - 2) - 1) >$$

$$(N - 1/4)\frac{\lambda^3 - 3}{\sqrt{(\lambda^2+2)^3}}. \tag{3.7.3}$$

Clearly it suffices to prove the following

3.7.4. Lemma. *If $\lambda \geq 2$, then we have*

$$10(\lambda^3 - 3)^2 > (\lambda^2 + 2)^3.$$

Proof of Lemma 3.7.4. We write

$$\lambda^3 - 3 = (\lambda - \sqrt[3]{3})(\lambda^2 + \lambda\sqrt[3]{3} + \sqrt[3]{9}).$$

Using the following obvious inequality for $\lambda \geq 2$

$$(\lambda^2 + \lambda\sqrt[3]{3} + \sqrt[3]{9})^2 > (\lambda^2 + 2)^2,$$

it suffices to prove for $\lambda \geq 2$

$$10(\lambda - \sqrt[3]{3})^2 \geq (\lambda^2 + 2). \tag{3.7.5}$$

We rewrite (3.7.5) in the following way

(3.7.6). $$((\sqrt{10} - 1)\lambda - \sqrt[3]{3}) \cdot ((\sqrt{10} + 1)\lambda - \sqrt[3]{3}) \geq 2.$$

It is easy to see that the function of λ on the LHS of (3.7.6) is monotone. Thus it suffices to check the equality for $\lambda = 2$. This completes our proof.

b) For a given vector $v = (v^1, \cdots, v^n)$ it suffices to find a unit vector $w = (w^1, \cdots, w^n)$ with the following property

$$(kn)w^1 + \sum_{i=2}^{n-1} w^i = 0 = \sum_{i=1}^{n} v^i w^i,$$

$$T(w, w, w) \geq (N/5) - 1.$$

Since $kn \neq 0$, without loss of generality we can assume that $v^1 = 0$.

We shall find w by using perturbation of the proof of case a. We shall normalize v by the condition $\sum (v^i)^2 = 1$.

Case 1. Not all the coordinates v^i are of the same sign, so we assume that $v^1 = 0, v^2 \leq 0, v^3 > 0$. We assume that $(0, w^2, w^3, 0 \cdots, 0)$ is the solution to the case 1 in 3.7.a w.r.t $\mathbf{R}^n(x^2, \cdots, x^n)$. In particular we have $(w^2)^2 + (w^3)^2 = 1$ and $w^2 > 0, w^3 > 0$. We choose ε_i from the following equations

$$w := (-\varepsilon_1, (1 - \varepsilon_2)w^2, (1 - \varepsilon_2)w^3),$$

$$-kn\varepsilon_1 + (1 - \varepsilon_2)(w^2 + w^3) = 0, \tag{3.7.7}$$

$$\varepsilon_1^2 = (2\varepsilon_2 - \varepsilon_2^2). \tag{3.7.8}$$

From (3.7.7) we get

$$\varepsilon_1 = \frac{(1 - \varepsilon_2)(w^2 + w^3)}{kn}. \tag{3.7.7$'$}$$

Substituting this into (3.7.8) we get

$$\left(\frac{(w^2 + w^3)^2}{(kn)^2} + 1\right)\varepsilon_2^2 - \left(2 + \frac{2(w^2 + w^3)^2}{(kn)^2}\right)\varepsilon_2 + \left(\frac{w^2 + w^3}{kn}\right)^2 = 0. \tag{3.7.9}$$

Clearly one of solution ε_2 of (3.7.9) is

$$\varepsilon_2 = 1 + \left(\frac{w^2 + w^3}{kn}\right)^2 - \sqrt{\left(1 + \frac{(w^2 + w^3)^2}{(kn)^2}\right.} \leq \frac{1}{2n^2}\left(1 + \frac{1}{2n^2}\right), \tag{3.7.10}$$

since $0 < w^1 + w^2 \leq 2$ and $k \geq 3$. We also have from (3.7.7)

$$\varepsilon_1 < \frac{1}{kn}.$$

Now it is easy to check that

$$T(w,w,w) \geq (\frac{N}{\sqrt{10}} - 1/4)(1 - \frac{1}{3n})^3 - \frac{N}{(3n)^3} \geq \frac{N}{5} - 1.$$

Case 2. We also apply our perturbation method above. So now we shall assume that $v^1 = 0$ and $v_2 \geq v_3 \geq v_4 > 0$, and $(0, -a, -a, \lambda a, \cdots, 0)$ is the solution corresponding to the case 2 in 3.7.a. We set

$$w := (\varepsilon_1, -a(1-\varepsilon_2), -a(1-\varepsilon_2), \lambda \cdot a(1-\varepsilon_2), 0, \cdots, 0).$$

$$\lambda := \frac{v^1 + v^2}{v^3} \geq 2, \text{ and}$$

$$a > 0 \text{ such that } a^2(\lambda^2 + 2) = 1.$$

Our perturbations terms ε_i satisfy the following equations

$$-kn\varepsilon_1 + (1-\varepsilon_2)(\lambda - 2)a = 0, \tag{3.7.11}$$

$$\varepsilon_1^2 = (2\varepsilon_2 - \varepsilon_2^2). \tag{3.7.12}$$

From (3.7.11) we get

$$\varepsilon_1 = \frac{(1-\varepsilon_2)(\lambda-2)a}{kn}. \tag{3.7.13}$$

Now substituting (3.7.13) into (3.7.12) we get one of solution ε_2

$$\varepsilon_2 = (1 + 2A^2) - \sqrt{(1+2A^2)^2 - A^2(1+A^2)},$$

where

$$A = \frac{(\lambda-2)a}{kn} < \frac{1}{kn}.$$

Since $0 < (\lambda - 2)a < 1$ we have $\varepsilon_2 < 1/(kn)$ and $\varepsilon_1 < 1/(kn)$ and

$$T(w,w,w) > \frac{N}{5} - 1.$$

□

c) The condition here is only δ(small)-differed from the condition in the statement b. Namely

$$T^0_{|H} = (\frac{1}{1-\varepsilon}(x^1)^3 + (kn)\sum_{i=2}^{n+1}(x^i)^3)_{|H}.$$

Now we shall use the same solution w in the statement b for estimate in our case c. We still keep the notation T for the tensor in the statement b. A straighforward computation show that for w choosen in case 1(b) we have now

$$T^0(w) \ge T(w) \ge \frac{kn}{5} - 1.$$

For the solution w in case 2 (b) we have $\varepsilon_1 < 1/(kn)$ and $\varepsilon_2 < 1/(kn)$. Hence

$$T^0(w) > \frac{kn}{5} - 1.$$

□

3.7. 14. Proposition. *Lemma 3.7.a holds also for $n = 3$ but not for $n = 2$, Lemma 3.7.b holds also for $n = 4$, but not for $n = 3$, and Lemma 3.7.c holds also for $n = 3$ but not for $n = 2$.*

Proof of Proposition 3.7.14. We notice that the condition $n = 4$ we use in the proof of Lemma 3.7.a only in the case 2, in order to make an assumption that the vector $(v_1, v_2, v_3, \cdot)$ which is orthogonal to our given hyperplan has all coordinates v_1, v_2, v_2 not zero. Thus this proof is also valid for $n = 3$ provided our hyperplan is orthogonal to such a vector. Now we observe that funtion $\lambda_{n-1}(T_{|H})$, where H is a hyperplan in $\mathbf{R}^n$ is a continous function on $\mathbf{R}P^{n-1}$. The our estimate on $\lambda_{n-1}(T_{|H})$ also holds for hyperplane whose coordinates contain a zero. This shows that Lemma 3.7.a also holds for $n = 3$. It cannot be valid for $n = 2$, because $\lambda_1(T_{|H^2})$ always vanishes. We use the same argument to prove the remained statements of Proposition 3.7.14. □

There are several obvious monotone invariants of T which are not norms.

$$A^1(T) := \max_{|x|=|y|=|z|=1, <x,y>=<y,z>=<z,x>=0} T(x, y, z)$$

is well-defined for $n \geq 3$.

$$A^2(T) := \max_{|x|=|y|=1, <x,y>=0} T(x, y, y),$$

is well-defined for $n \geq 2$. From Remark 3.3 we see easily

$$\ker A^1 = \mathcal{R}^n.$$

On the other hand we have

$$\ker A^2 \subset \mathcal{R}(3\pi_1).$$

Thus A^1 and A^2 are different invariants.

We can also use decomposition (3.2.1) to define a less obvious monotone invariant.

3.8. Lemma. *We denote by π_1 the first component of T in decomposition (3.2.1). Then $||T||_1 := ||\pi_1(T)||$ is a monotone invariant of T.*

Proof. Let $\mathbf{R}^k$ be a subspace of $\mathbf{R}^n$. We denote by $\pi_k^n T$ the restriction of T to $\mathbf{R}^k$. Clearly

$$\pi_k^n(T) = \pi_k^n(\pi_1 T) + \pi_k^n(\pi_2 T).$$

We have noticed in Proposition 3.5 that the restriction of the trace form $\pi_2 T$ to any subspace is also a trace form. Thus $\pi_k^n(\pi_2)$ is an element in $\mathcal{R}^k \subset S^3(\mathbf{R}^k)$. Hence we have

$$\pi_1(\pi_k^n T) = \pi_1(\pi_k^n(\pi_1 T)). \tag{3.8.1}$$

Since all the projections π_1, π_k^n decrease the norm $||.||$ we get

$$||\pi_k^n T||_1 = ||\pi_1(\pi_k^n T)|| = ||\pi_1(\pi_k^n(\pi_1 T))|| \leq ||\pi_1(T)|| = ||T||_1.$$

□

3.9. Proposition. *For $n = 1$ all the comasses concide with the absolute norm. A neccesary and sufficient condition, such that a statistical line $(\mathbf{R}, g^0, T)$ can be embedded into $(\mathbf{R}^N, g^0, T')$ is that $\mathcal{M}^1(T) \leq \mathcal{M}^1(T')$.*

Proof. It suffices to show that we can embedd $(\mathbf{R}, g^0, T)$ into $(\mathbf{R}^N, g^0, T')$ if we have $\mathcal{M}^1(T) \leq \mathcal{M}^1(T')$. We note that $T'(v, v, v)$ defines a anti-symmetric function on the sphere $S^{N-1}(|v| = 1) \subset \mathbf{R}^N$. Thus there is a point $v \in S^{N-1}$ such that $T'(v, v, v) = \mathcal{M}^1(T)$. Clearly the line $v \otimes \mathbf{R}$ defines the required embedding. □

Let us consider the embedding problem for 2-dimensional linear statistical spaces. It is easy to see that

$$S^3(\mathbf{R}^2) = \mathbf{R}^2 \oplus \mathbf{R}^2.$$

Thus the quotient $S^3(\mathbf{R}^2)/SO(2)$ equals $(\mathbf{R}^2 \oplus \mathbf{R}^2)/S^1$. Geometrically there are several ways to see this. In the first way we denote components of $T \in S^3(\mathbf{R}^2)$ via $T_{111}, T_{112}, T_{122}, T_{222}$.

3.10. Lemma. *There exists an oriented orthonormal basic in* $\mathbf{R}^2$ *such that* $T_{111} = \mathcal{M}^1(T) > 0, T_{112} = 0$ *for all non-vanishing* T. *These numbers* $(T_{111}, T_{122}, T_{222})$ *are called canonical coordinates of* T. *Two tensors* T *and* T' *are equivalent, if and only if they have the same canonical coordinates.*

Proof. We choose an oriented orthonormal basis (v_1, v_2) by taking as v_1 a point on $S^1(|x| = 1)$, where the function $T(x, x, x)$ reaches the maximum. The first variation formula shows that in this case $T_{112} = 0$. This shows the existence of canonical coordinates. Clearly if two tensors have the same canonical coordinates then they are equivalent. Next if two tensors T and T' are equivalent, then their norms $\mathcal{M}^1$ are the same. We need to take care the case when there are several points x, at which $T(x, x, x)$ reaches the maximum. In any case they have the same first coordinates. Next we note that

$$< Tr(T), Tr(T) >= (T_{111} + T_{122})^2 + T_{222}^2,$$
$$||T||^2 = T_{111}^2 + T_{122}^2 + T_{222}^2.$$

Thus if two tensors are equivalent and have the same first coordinates they must have the same third coordinates T_{122} and the last one is the same up to sign. The condition on orientation tells us that the sign must by +. This proves the second statement. □

3.11. Remark. First we note that, if $T_{111} = \mathcal{M}^1(T)$, then $T_{122} = HessT_{v_1}(v_2, v_2) \leq 0$. Thus not any 3 numbers $(T_{111}, T_{122}, T_{222})$ can be served as canonical coordinates of some 3-symmetric tensor T on $\mathbf{R}^2$. Next we observe that $v = (x^1 v_1 + x^2 v_2)$ is a critical point of the function $T(v, v, v)$ on the circle $S^1((x^1)^2 + (x^2)^2 = 1))$, if and only if there is a number λ such that

$$T_{111}(x^1)^2 + T_{122}(x^2)^2 = \lambda x^1, \qquad (3.11.a)$$

$$(3.11.b) \qquad 2T_{122}x^1x^2 + T_{222}(x^2)^2 = \lambda x^2.$$

The equation (3.11.b) has 2 solutions, namely $x^2 = 0$ and $2T_{122}x^1 + cx^2 - \lambda = 0$. Clearly $(x^1 = \pm 1, x^2 = 0)$ are solutions to the system (3.11.a, b). These solutions correspond to the maximum and minimum of $T(v, v, v)$. Now we suppose that $T_{222} \neq 0$. Then substituting $x^2 = (\lambda - 2T_{122}x^1)/c$ into (3.11.a) and the equation $|v| = 1$ we get two quadratic equations for x^1 and λ

$$(3.11.c) \qquad (x^1)^2 + \frac{(\lambda - 2T_{122}x^1)^2}{c^2} = 1,$$

$$(3.11.d). \qquad T_{111}(x^1)^2 + \frac{T_{122}}{c^2}(\lambda - 2T_{122}(x^1)^2) = \lambda\, x^1.$$

It is easy to see that these equations have at most 4 solutions (x_1, λ). If T_{122} is positive, using the fact that $T(v_1, v_1, v_1)$ is a local minimum, we see easily that in this case the function $T(v, v, v)$ has exactly 6 critical points.

The second way to see invariants of 3-symmetric tensors on $\mathbf{R}^2$ is to decompose $S^3(\mathbf{R}^2)$ into the trace type component and the second component which is generated by $Sym(\Re(x\bar{y}\bar{z})), Sym(\Im(x\bar{y}\bar{z}))$. Here we identify $\mathbf{R}^2$ with $\mathbf{C}$, so x, y, z are complex numbers. (We note that this decomposition is not orthogonal.)

3.12. Proposition. *We can always embed the 2-dimensional statistical space* $(\mathbf{R}^2, g^0, 0)$ *into any linear statistical space* $(\mathbf{R}^n, g^0, T)$*, if* $n \geq 7$.

Proof. It suffices to prove for $n = 7$. We denote by $\mathcal{O}(T)$ the set of of all unit vectors $v \in S^6$ such that $T(v, v, v) = 0$. Clearly $\mathcal{O}(T)$ is a set of dimension 5 in S^6. Since T is anti-symmetric, there exists a connected component $\mathcal{O}^0(T)$ of $\mathcal{O}(T)$ which is invariant under the anti-symmetry involution. Now we consider the following function f on $\mathcal{O}^0(T)$. For each $v \in \mathcal{O}^0(T)$ we denote by A^v the bilinear symmetric 2-form on the space $T_x\mathcal{O}^0(T)$ considered as a subspace in $\mathbf{R}^n$:

$$A^v(y, z) = T(v, y, z).$$

Then we define $f(v)$ equal to $\det(A^v)$. Since $\mathcal{O}(T)$ has dimension 5, the function $f(v)$ is anti-symmetric on $\mathcal{O}^0(T)$. Hence the set $\mathcal{O}^0_0(T)$ of all $v \in \mathcal{O}^0(T)$ with $f(v) = 0$ has dimension 4 and it contains a connected component which is also invariant under the anti-symmetric involution. For the simplicity we denote this connected component also by $\mathcal{O}_0^0(T)$. Now we consider the following two cases.

Case 1. We assume that there is a point $v \in \mathcal{O}_0^0(T)$ such that the nullity of A^v is at least 2. Then there are two linear independent vectors $y, z \in T_v$ such that the restriction of A^v on the plan $\mathbf{R}^2(y, z)$ vanishes. Since the set $\mathcal{O}^0(T)$ is connected and anti-symmetric and of codimension 1 in S^{n-1}, the plan $\mathbf{R}(y, z)$ has a non-empty intersection with $\mathcal{O}^0(T)$ at a point w. Then the restriction of T on the plan $\mathbf{R}^2(v, w)$ is vanished, because

$$T(v, v, v) = T(w, w, w) = 0$$

$$T(v, w, w) = 0 \,(\text{since } A^v(w, w) = 0),$$

$$T(v, v, w) = 0 \,(\text{since } w \in T_v\mathcal{O}^0(T)).$$

Case 2. We assume that the nullity of A^v on $\mathcal{O}_0^0(T)$ is constantly 1. Using the anti-symmetric property of A^v we conclude that the restriction of A^v to the the plan $\mathbf{R}^4(v)$ which is orthogonal to the kernel of A^v has index constantly 2. Thus there exists a vector z which is orthogonal to the kernel y of A^v such that $A^v(z, z) = 0$. Clearly the restriction of A^v to the plan $\mathbf{R}^2(y, z)$ is vanished. Now we can repeat the argument in the case 1 to get a vector w such that the restriction of T to $\mathbf{R}^2(v, w)$ vanished. $\square$

Now let us consider embedding of linear statistical spaces in "standard" spaces.

3.13. Theorem. *a) Any statistical space $(\mathbf{R}^n, g^0, T)$ can be embedded in the statistical space $(\mathbf{R}^{n(n+1)}, g^0, T' = 2||T|| \sum_{i=1}^{N(n)} x_i^3)$, where x_i are the canonical Euclidean coordinates on $\mathbf{R}^{n(n+1)}$.*
b) The trivial space $(\mathbf{R}^n, g^0, 0)$ can be embedded into $(\mathbf{R}^{2n}, g^0, \sum_{i=1}^{2n}(dx^i)^3)$ for all n.

Proof. a) We prove by induction. The statement for $n = 1$ follows from Proposition 3.8. Suppose that the statement is valid till $n = k$.

3.14. Lemma. *Suppose that $T \in S^3(\mathbf{R}^{k+1})$. Then there are orthonormal coordinates $x_1, \cdots, x_k$ such that*

$$T = x_1 \sum_{i=1}^{k+1} a_i x_i^2 + \sum_{1<i,j,k} a_{ijk} x_i x_j x_k. \tag{3.14.1}$$

Proof of Lemma 3.14. We choose v_1 as the unit vector in $\mathbf{R}^{k+1}$, where the function $T(v,v,v)$ reach the maximum on the unit sphere S^k. The first variation formula show that $T(v_1,v_1,w)=0$ for all w which is orthogonal to v_1. Denote by $\mathbf{R}^k$ the orthogonal complement to $\mathbf{R}\cdot v_1$. Now we consider a bilinear symmetric form A on $\mathbf{R}^k$ defined as follows

$$A(x,y)=S(v_1,x,y).$$

There is an orthonormal basis on $\mathbf{R}^k$, where we can write $A(x,y)=\sum_{i=2}^{k+1} a_i x_i^2$. Clearly in this orthonormal basis we can write T in form of (3.14.1). □

Continuation of the proof of Theorem 3.13.a We shall show explicitely that that any statistical space $(\mathbf{R}^2, g^0, T=a_2x_1(x_2)^2)$ can be embedded in $(\mathbf{R}^4, g^0, \sum_{i=1}^4 (y_i)^3)$, if $0\le|a_2|\le 1/2$. We denote by $A(v_i)\in\mathbf{R}^4$ the image of the orthonormal basis vector v_i in $\mathbf{R}^2$. We let

$$A(v_1)=\pm(\frac{1}{2},\frac{1}{2},-\frac{1}{2},-\frac{1}{2}) \tag{3.15.1}$$

$$A(v_2)=(\sqrt{\frac{1+2a_2}{2}},-\sqrt{\frac{1+2a_2}{2}},\sqrt{\frac{1-2a_2}{2}},-\sqrt{\frac{1-2a_2}{2}}). \tag{3.15.2}$$

Here we take the sign $+$ in (3.16.1) if $a_2>0$ and we take the sing $-$, if $a_2<0$. Clearly A define the required embedding.

This together with Proposition 3.8 and the induction assumption complete the proof of Theorem 3.13. a.

Proof of Theorem 3.13. b. We decompose the embedding $f:(\mathbf{R}^n, g^0, 0)$ to $(\mathbf{R}^{2n}, g^0, \sum_{i=1}^{2n}(x^i)^3)$ as follows

$$f(x_1,\cdots,x_n)=(f^1(x_1),\cdots,f^n(x_n))$$

where f^i embeds the line $(\mathbf{R},(dx^i)^2,0)$ into $(\mathbf{R}^2,(dx^{2i-1})^2+(dx^{2i})^2,(dx^{2i-1})^3+(dx^{2i})^3)$. Clearly f is the required embedding. □

3.16. Further remarks. Let us define the **nullity** of T as the maximal dimension of a subspace, where T vanishes. Clearly the nullity of T is the maximal number k such that $\lambda_k(T)=0$. We also define the **rank** of T as the minimal number k such that T can obtained by taking the pull back via the orthogonal projection von the $\mathbf{R}^n$ to its subspace of dimension k. Clearly $N(T)+R(T)\ge n$. We conjecture that the nullity of a linear statistical space $(\mathbf{R}^{2^k+1}, g^0, T)$ is at least k for any T.

4 Monotone invariants and obstructions to embeddings of statistical manifolds

In this section we shall consider several classes of invariants of statistical manifolds. These classes behave well under embeddings, and therefore they give us obstructions of embedding of one statistical manifold into another one.

An assignment $(M, g, T) \mapsto A$ of each statistical manifold (M, g, T) to an object A is called **invariant** of statistical manifolds, if this assigment depends only on the equivalent class of (M, g, T).

We now define a subclass of invariants of statistical manifolds which behaves well under embeddings. Let $K(M, e)$ denote the category of statistical manifolds M with morphisms being embeddings. Functors of this category are called **monotone invariants** of statistical manifolds. Clearly any monotone invariant is an invariant of statistical manifolds.

It is easy to construct invariants of statistical manifolds from invariant of linear statistical spaces. We can also construct monotone invariants of statistical manifolds from monotone invariants of linear statistical spaces.

4.1. Examples.

a) First we show an example of an important invariant of statistical manifolds but not monotone invariants. The trace $Tr\, T$ shall be called **the trace form** of a statistical manifold (M, g, T). Clearly the trace form is invariant but not monotone invariant of statistical manifolds. Nevertheless according to Proposition 3.5 we get that any statistical submanifold of a statistical manifold of a trace type is also of trace type. Thus the trace type is a monotone invariant. In particular we can not embed the statistical space Cap^N and the normal Gaussian space into any statistical space of trace type. On the other hand, unlike the linear case, we cannot embed a statistical manifold of trace type into another one of trace type, even if the norm condition is satisfied. For example if the trace form is closed (or exact), then the trace form of its submanifolds is also closed (resp. exact). Hence within a class of statistical manifolds of trace type we get a new mononotone invariants which can be expressed via the closedness and the cohomology class of the corresponding trace form.

Furthermore we note that the class of 3-symmetric tensors of trace form is a subclass of all **decomposable tensors** T^3 which are a symmetric product of 1-forms and symmetric 2-forms. Any statistical submanifold of a statistical manifold with a decomposable tensor T has also the (induced) decomposable tensor. Thus the decomposability is also a monotone invariant. The Gaussian normal 2-dimensional manifold is an example of decomposable type but not of trace type.

b) We define for any statistical manifold (M, g, T) the following number

$$rank(T) = \sup rank(T(x))$$

$$||T||_0 = \sup_{x \in M} ||T(x)||.$$

$$\mathcal{M}^1(T)_0 = \sup_{x \in M} \mathcal{M}^1(T(x)).$$

$$||T||_{1,0} = \sup_{x \in M} ||T(x)||_1.$$

Clearly these three numbers are monotone invariants of statistical manifolds. The second and third numbers are norms, and the last one vanishes, if and only if (M, g, T) is of trace type. (Actually these invariants are invariant of immersions.)

4.2. Proposition. *Any statistical manifold which is conformal equivalent to the space Cap^N cannot be imbedded into the direct product of m copies of the normal Gaussian statistical manifold 2.3.3.a for any $N \geq 3$ and finite m.*

Proof. Using (2.6.1) and Lemma 3.7 we conclude that $\mathcal{M}^1(Cap^N) = \infty$. Thus any statistical manifold which is conformal equivalent to CaP^N has also the infinite invariant $\mathcal{M}^1$. On the other hand we compute easily from 2.3.3 and (2.6.3), (2.6.4) that the norm $\mathcal{M}^1$ of the Gaussian normal manifold and any direct product of a finite copies of it is finite, namely the norm $\mathcal{M}^1(x)$ is contant on the Gaussian manifold M^2 and $\mathcal{M}^1(x) = \sqrt{2}$. (By the observation above we can actually replace the "imbedded" statement by a stronger "immersed" statement.) □

4.3. Diameters of statistical manifolds. For a positive number $\rho > 0$ and a statistical manifold (M, g, T) we set

$$d_\rho(M, g, T) := \sup\{l \in R^+ \cup \infty \mid \exists \text{ an embedding of } ([0, l], dx^2, \rho(dx)^3) \text{ to } (M, g, T).\}$$

We shall call $d_\rho(M, g, T)$ **the diameter with weight** ρ of (M, g, T). Clearly d_ρ are monotone invariants for all ρ.

To estimate the diameter with weight ρ of a given statistical manifold (M, g, T) we can proceed as follows. For each point $x \in M$ we denote by $D_\rho(x)$ the set of all unit tangential vector $v \in T_xM$ such that $T(v, v, v) = \rho$. We denote by $D^i_\rho(x)$ the connected components of $D^i_\rho(x)$. We say that a unite vector v in T_xM is **ρ-charateristic** with weight $c(x)$, if there exists i such that we have

$$c(x) = \min_{w \in D^i_\rho(x)} < v, w >> 0.$$

The definition of diameter d_ρ uses the embedding. We can also use the immersion to defined **immersion diameter** d^{im}_ρ of statistical manifolds. Clearly immersion diameters are also monotone invariants of statstical manifolds and moreover

$$d_\rho(M, g, T) \le d^{im}_\rho(M, g, T).$$

In many cases it is easier to compute immersion diameters. We suspect that in most cases these two diameters coincide.

We shall say that a point $x \in M$ is **ρ-regular**, if there is an open neighborhood $U_\varepsilon(x) \subset M$ such that $D_\rho(U_\varepsilon) = U_\varepsilon \times D_\rho(x)$. It is easy to see that the set of ρ-regular points is open and dense in M for any given ρ.

4.4. Proposition. *The diameter d_ρ of (M^m, g, T) is infinite, if $m \ge 3$ and there exists a number $\varepsilon > 0$ such that one of the following 2 conditions holds:*
a)There exists a $(\rho + \varepsilon)$-regular point $x \in M$ such that the convex hull $Cov(D^i_{\rho+\varepsilon}(x))$ of one of connected components $D^i_{\rho+\varepsilon}(x)$ contains the origin point $0 \in T_xM^m$ as it interior point.
b) (M^m, g, T) has a complete Riemannian submanifold $(N, \bar{g})$ such that there exists a smooth section $x \mapsto (D_{\rho+\varepsilon}(x) \cap TN)$ over N.

Proof. The statement under the first condition a) is based on the fundamental Lemma of the convex integration technique of Gromov [Gr]. Namely we shall use the Gromov Lemma [2.4.1.A, Gr] in order to prove the following statement

4.5. Lemma. *Under the condition in Proposition 4.4.1 there exists a small neighborhood $U_\delta(x)$ in M and an embedded oriented curve $S^1 \subset U_\delta(x)$ such that for all point $s(t) \in S^1$ we have $\mathcal{M}^1(T_{s(t)}S^1) \ge \rho + (\varepsilon/2)$.*

Proof of Lemma 4.5. We denote by Exp the exponential map $T_xM \to M$ and $DExp$ the differential of the exponential map restricted to $S^{m-1} \times T_xM \subset T(T_xM)$ to TM. Here we denote by S^{m-1} the unit sphere in T_xM. The space T_xM is a linear statistical space, so we denote be $\mathcal{M}^1_x$ the induced norm-function on $S^{m-1} \times T_xM$:

$$\mathcal{M}^1_x(l) = T_x(l, l, l).$$

Since $Dexp$ is a continous function whose restriction to $S^{m-1} \times 0$ is the identity, there exists a ball $B(0, \delta)$ with center in $0 \in T_xM$ such that

$$\mathcal{M}^1(DExp(l)) - \mathcal{M}^1_x(l)) < \varepsilon/4 \tag{4.5.1}$$

for all $l \in S^{m-1} \times B(\delta) \subset T(T_xM)$. We can assume that$\delta$ is so small such that $DExp$ is a homeomorphism on $S^{m-1} \times B(0, \delta)$.

Now we apply the Gromov Lemma [2.4.1.A, Gr] to get a oriented curve $S^1(t)$ in the linear space T_xM such that

$$T(\frac{(\partial/\partial t)S^1(t)}{|(\partial/\partial t)S^1(t)|}) = \rho + \varepsilon \tag{4.5.2}$$

for all t. Next we observe that for all $\alpha > 0$ the curve $\alpha \cdot S^1(t)$ has the same norm as $S^1(t)$, i.e.

$$\mathcal{M}^1_x(T_{|(\alpha \cdot S^1)}(t)) = \mathcal{M}^1_x(T_{|(S^1)}(t)) = \rho + \varepsilon.$$

Thus we can assume now that our curve $S^1(t)$ which satisfies (4.5.2) lies in the ball $B(0, \delta)$. By our choice of δ (see (4.5.1)) we get from (4.5.2)

$$\rho + \frac{3}{4}\varepsilon \le \mathcal{M}^1(Exp(S^1(t))) \le \rho + \frac{5}{4}\varepsilon, \tag{4.5.3}$$

for all t. This curve $Exp(S^1(t))$ is an immersed curve. To get an embedded curve we perturb it a little such that the condition of Lemma 4.5 is satisfied. $\square$

Now let us continue the proof of Proposition 4.4.a. We denote by $S^1(t)$ the embedded curve in Lemma 4.5. Next by choosing a tubular neghborhood of $S^1(t)$ we can get a (small, thin) (oriented) embedded solid torus $T^3(t, s, r) = S^1(t) \times S^1(s) \times [0, R]$ in M^m such that our embedded curve is exactly the mean curve $S^1(t) \times \{0\} \times \{0\}$ on the solid torus. We can choose this torus T^3 so thin, such that for all s, t, r we have

$$\mathcal{M}^1(T^2_r(t, s)) \ge \rho + \frac{\varepsilon}{4}. \tag{4.5.4}$$

Using (4.5.4) we choose a smooth unit vector field $V(t,s)$ on the torus $T^3(t,s,r)$ which tangential to each torus $T^2_r(t,s)$ such that $T(V,V,V)=\rho$. The integral curve of this vector field is either a cicle or an curve of infinite length. If there exists an integral curve of infinite length then this curve is our desired curve for the Proposition 4.5. Assume now that all the integral curves are cicles. Then there exist an embedding $S^1(t)\times[0,\mu]\times[0,\mu]$ such that for all $(s,r)\in[0,\mu]\times[0,\mu]$ the circle $S^1(t)\times\{s\}\times\{r\}$ is an integral curve of V. Now we perturb V in a neighborhood $[0,\alpha]\times[0,\mu]\times[0,\alpha]$ with a very small α such that the perturbed unit vectorfield V' satisfies $T(V',V',V')=\rho$ and the integral curve of vector field V' is not any more periodic. This completes the proof of the first part in Proposition 4.4.

Using the same argument we can prove the second part b) of Proposition 4.4. First we get the existence of an embedded curve $S^1(t)$ of arbitrary length on M such that $\mathcal{M}^1(T_{|S^1}(t))\geq\rho+(1/4)\varepsilon$. Now we consider a torus tubular neighborhood of this curve in M and apply the same argument in the first part, namely we get on each torus $T^2(t,s)$ an integral curve whose unit tangential vector $V=(\partial/\partial t)S^1(t;s,r)$ satisfies the condition:

$$T(V,V,V)=\rho.$$

If there exists an inifinite integral curve then we are done. If not, that is all integral curve are circles, then we apply the perturbation method in the proof of the first part and get our desired curve. $\square$

4.6. Remark. Using the arugment in the proof of Proposition 4.4 we get the following monotonicity for diameters of a statistical manifolds. If $\dim M^m=m\geq 3$ then

$$d_\rho(M,g,T)\leq d_{\rho'}(M,g,T) \text{ if } \rho\geq\rho.$$

We can also get an upper estimate of diameters of a statistical manifold by using characterstic vectors. Instead of giving a formal generalization we shall consider a simple example of the ball $(B^n(r),g^0,\sum_{i=1}^n(dx^i)^3)$ in the linear statistical space. First we can easy compute that (e.g. via the first variational formula as we do it in Remark 3.11) $\mathcal{M}^1(B^n(r))=1$, moreover, if v is a unit vector in $\mathbf{R}^n$ such that $T(v,v,v)=1$, then v must be one of the coordinate vectors ∂x^i. Since we can easily compute all the critical value of the function $T(v,v,v)$ on the unit sphere S^{n-1} we get that for for any value c such that $(2)^{-1/2} < c \le 1$ the set $D_{1-c}(x)$ has exactly n connected components, each of them is diffeomorphic to S^{n-2}. We fix such a value c. We denote by $D^i_{1-c}(x)$ the component whose i-th coordinate has the maximal value. Clearly the unit coordinate vector ∂x^i is $(\rho - c)$-characterstic with the weight $\sin\alpha(c)$ which is the "larger" solution of the equation

$$\sin^3\alpha(c) + \cos^3\alpha(c) = 1 - c, \text{ and } 0 \le \alpha(c) \le \pi/4. \tag{4.6.1}$$

4.7. Lemma. *For $2^{-1/2} < c \le 1$ the immersion diameter d^{im}_{1-c} of the ball $B^n(r)$ of radius r in the standard linear statistical space $(\mathbf{R}^n, g^0, \sum_{i=1}^n (dx^i)^3)$ is less than $r \cdot (\sin^{-1}\alpha(c))$ where $\sin\alpha(c)$ is the larger solution of the equation (4.6). The immersion diameter $d^{im}\rho$ of $(\mathbf{R}^n, g^0, \sum_i (dx^i)^3)$ is zero, if $\rho > 1$.*

Proof. We have compute that the coordinate vector $\partial x^i = (0, \cdots, 1_i, 0)$ are $(1-c)$ characteristic vector with weight $\sin\alpha(c)$. Thus the projection of any unit tangential vector v of any immersed curve $[0,\alpha], (dx)^2, (1 - eps)(dx)^3)$ on the coordinate line $(0, \cdots, \mathbf{R}, \cdots, 0)$ has a length greater than or equal to $\sin\alpha(c)$. Thus after a time $t = r \cdot \sin^{-1}\alpha(c)$ the integral curve must quit the ball. Since we rescale the time to equal the length of curve, we obtain immediately the first statement.

The second statement follows from an easy computation of the norm $\mathcal{M}^1$ for standard linear statistical spaces. □

4.8. Proposition. *For a given ρ the diameter with weight ρ of Cap^N is equal to infinity if $N \ge 3$.*

Proof. Using Proposition 3.7.14 and Lemma 3.7.c we can find a point $x \in Cap^N$ such that

$$\lambda_{N-2}(T_x Cap^N) > 3\rho.$$

Since λ_i is a smooth function of tensors T, we can find a very small neighborhood U of a ρ-regular point-x such that all point y in U is also ρ-regular and moreover

$$\lambda_{N-2}(T_y Cap^N) > 2\rho, \text{ if } y \in U. \tag{4.8.1}$$

Now we embed a small torus $T^{N-2}(\delta)$ in this neighborhood U. From (4.8.1) it follows that for all $x \in S^{N-2}(\delta)$ we have

$$\mathcal{M}^1(T_x T^{N-2}(\delta)) > \rho + \varepsilon_1. \tag{4.8.2}$$

We denote

$$D_{\rho^+}(x) = \cup_{\rho' \geq \rho} D_{\rho'}(x).$$

We shall find a smooth section of $D_{\rho^+}(x) \cap T_x T^{N-2}(\delta)$ over $T^{N-2}(\delta)$. Next we observe that if $V(x) \in D_{\rho^+}(x)$ then $g(x) \cdot V(x) \in D_{\rho^+}(x)$. Thus using the unity partition functions it suffices to find for each $x \in T^{N-2}$ a small neighoboorhood $U(x) \subset T^{N-2}$ and a smooth section of $D_{\rho^+} \cap T(T^{N-2})$ over $U(x)$. The local existence od such a section follows from (4.8.2) and fom the continuity of the function $\mathcal{M}^1(x)$ on $T^{N-2}(\delta)$. Once we have a smooth section of D_{ρ^+} on $T^{N-2}(\delta)$ we can get an integral curve of of infinite length on $T^{N-2}(\delta)$. Afeter that we can perturbe this curve as we did in our proof of Proposition 4.4 in order to get an embedded curve which is an integral curve of the distribution D_ρ on Cap^N. □

4.9. Remark. We can construct other monotone invariants by looking at the embedding of non-constant 1-dimensional statistical manifolds. It seems that the most important among these new invariants come from of 1-dimensional statistical manifolds, whose asymptotical growth $\mathcal{M}^1(x)$ is polynomial or exponential.

5 Existence of embeddings into Cap^N.

We prove in this section the following theorems.

5.1.Proposition. *Any C^k-Riemannian manifold $(M^n, <,>_g)$, $k = 1$ or $3 \leq k \leq \infty$, posseses a probability potential, more precisely there exists a finite number of N positive functions $f_i : M \to \mathbf{R}, i = \overline{1, N}$, such that $\forall x \in M$ we have $\sum_{i=1}^N f_i(x) = 1$, and moreover for all $V, W \in T_x M$ the following equality is satisfied*

$$(5.1.1) \qquad < V, W >_g (x) = \sum_{i=1}^N (\partial_V \ln f_i(x))(\partial_W \ln f_i(x)) f_i(x).$$

Theorem 5.1 says that any Riemannian manifold can be considered as a family of probability distributions on the sample space Ω_N of N elementary events with the natural Fisher (information) metric on it. Refrasing we can say that any Riemannian metric is a Fisher metric. Theorem 5.1 is actually a simple consequence of the Nash embedding theorems [N1, N2].

5.2. Theorem. *Suppose that* (M^m, g, T) *is a* C^k*-statistical manifold,* $3 \le k \le \infty$*. There exists a number* $u(m) = 4(m+1)[((m(m+1)-1)m(m+1) + 2 + m) + (m+2)(m+3)]$ *such that we can embed* (M^m, g, T) *into* $Cap^{N(m)}$*. If* $M^m, g, T)$ *is* C^3*-bounded, i.e.* $|T|_{C^3} < \infty$ *then we can lower* $u(m)$ *by dividing* $(m+1)$*.*

5.3. Corollary. *Any finite dimensional* C^3*-statistical manifold is a statistical model. Hence we get a new proof of the Matumoto theorem on the existence of constrast function for these manifolds [M] (see also section 2).*

From Theorem 5.2 follows that any monotone invariant of Cap^N goes to infinity as N goes to infinity. This allows Cap^N to be universal spaces for statistical manifolds.

The rest of this section is devoted to the proof of Proposition 5.1 and Theorem 5.2. We also discuss at the end of the section C^1- embeddings of one statistical manifold into another one.

Proof of Proposition 5.1. As we have observed in 2.3.2 the existence of probability potential functions $\{f_i\}$ for a Riemannian manifold (M, g) is equivalent to the existence of an isometric embedding $(M, g) \to (S^n_+(2), g^0)$. Here we denote by g_0 the canonical metric on positive quadrant $S^n_+(2)$ of the sphere of radius 2 as in 2.3.2. This existence can be obtained by a general theory of isometric embeddings of Riemannian manifolds devoloped by Gromov [G] based on the Nash embedding theorem [N]. Here we give a simple explanation, how to get an isometric embedding $(M^m, g) \to (S^n_+(2), g^0)$, once we have Nash's isometric embedding (M^m, g) into the standard Euclidean space $(\mathbf{R}^{N(m)}, g^0)$. (The number $N(m)$ can be chosen as $(n/2)(3n+11)$ for a compact Riemannian manifold (M^m, g) and $(n/2)(n+1)(3n+11)$ for non-compact (M^m, g) [N1, N2]) (later Gormov [G] showed that we can chose $N(m) = (n+2)(n+3)/2$ as the best value for the smooth case, for C^1-embedding Nash has shown that wew can take $N(m) = 2m = 1$ [N1]). First we note that the image of (M^m, g) can be seen as in lying a ball $(B^{M(n)}, r)$ of radius r for a compact M and also for non-compact M, but in this case we have to add 1 to the dimension of $B^{M(m)}$. (This follows immediately from the Nash proof for the non-compact M^m, based on the dividing M^m to compact subsmanifolds). Next we can decrease r to $1/16$ by double the dimension of the ball, i.e. we embed the ball $(B^{M(n)}, r)$ to $(B^{2(M(n)}, 1/16)$. Now we have an isometric imbedding of (M^m, g) to a small region $R(1/16)$ in the flat torus $S^1 \times_{2M(n)} S^1$. This flat torus can be embedded as a Clifford torus in the standard sphere $S^{4M(n)-1}(2)$ such that the small region $R(1/16)$ lies in the positive quadrant $S_+^{4M(n)-1}$. Thus we get the required isometric embedding of (M^m, g). $\square$

Proof of Theorem 5.2. We first consider C^3-bounded statistical manifold (M, g, T), i.e. those such that $|T|_{C^3} < infty$. Our proof consists of 4 steps. In the first step we show that for a generic map f we can solve the linearized problem for the perturbed equations (5.4) and (5.5) at f. In the second step we apply the Nash-Gromov implicit function theorem to the first step and prove that our perturbation problem is actually can be solved for small value $\tilde{g}$ and $\tilde{T}$ and a generic f. In the third step we combine the Nash isometric embedding, the Nash-Gromov approximation theorem and our linear embedding theorem (Theorem 3.13.b) with our perturbation device in the second step to get M^m embedded in a linear statistical space $(\mathbf{R}^N(m), g^0, A\sum((dx^i)^3)$. Here A is a positive constant, possibly very big. In the last step we embed any given bounded domain of the last linear statistical space into a finite dimensional space Cap^N. This step is similar to the proof of Proposition 4.8.

Step 1. Linearized problem. Let f be a map $M^m \to (\mathbf{R}^q, g^0, T^0)$. Here g^0 is the standard Euclidean metric and $T^0 = \sum_{i=1}^q (dx^i)^3)$. Suppose that $\tilde{g} \in S^2(T^*M)$ is a small symmetric 2-form on M and $\tilde{T} \in S^3(T^*M)$ is a small 3-symmetric tensor on M. We want to find a small perturbation v at f such that

$$d(f+v) \cdot d(f+v) = df \cdot df + \tilde{g}, \tag{5.4}$$

$$d(f+v) \cdot d(f+v) \cdot d(f+v) = df \cdot df \cdot df + \tilde{T}. \tag{5.5}$$

Before solving (5.4) and (5.5) in the second step, we shall solve the corresponding linearized equations for v at a given map f. The linearized equation of the system (5.4) +(5.5) is the following system of (5.6) + (5.7)

$$\partial_i f \cdot \partial_j v + \partial_j f \cdot \partial_i v = \tilde{g}_{ij}, \text{ for all } i, j, k \tag{5.6}$$

$$S_{ijk} \circ T^0(\partial_i f, \partial_j f, \partial_k v) = \tilde{T}_{ijk} \text{ for all } i, j, k. \tag{5.7}$$

Here we denote by S_{ijk} the symmetrization over (ijk), i.e. we permute the operators $(\partial_i, \partial_J, \partial_k)$ in the LHS of (5.7). We shall simplify the form of (5.7) by introducing the bilinear symmetric map $A : \mathbf{R}^q \times \mathbf{R}^q \to \mathbf{R}^q$ as follows

$$g^0(A(v, w), e_i) = T^0(e_i, v, w).$$

Here $\{e_i\}$ is the dual basis to x^i on $\mathbf{R}^q$. Then we can rewrite (5.7) as

$$S_{ijk} < \partial_k v, A(\partial_i f, \partial_j f) >= \tilde{T}_{ijk}. \tag{5.7'}$$

Following Nash [N2] and Gromov [G] we shall consider only deformations v which are orthogonal to f and T^0-orthogonal to f i.e. we add the following equations for v

$$< v, \partial_i f >= 0, \text{ for all } i, \tag{5.8}$$

$$< v, A(\partial_i f, \partial_j f) >= 0, \text{ for all } i, j. \tag{5.9}$$

The Nash -Gromov trick makes it possible that under the conditions (5.8) and (5.9) (namely we apply the differentiation ∂_k to (5.8) and (5.9)) the system of linearized equations (5.6) + (5.7) is equivalent to the following linear algrbraic system for v

$$< v, \partial_i \partial_j f >= -\frac{1}{2}\tilde{g}_{ij}, \tag{5.10}$$

$$< v, A(\partial_i \partial_j f, f_k) >= -\tilde{T}_{ijk}. \tag{5.11}$$

Clearly the system (5.8)+(5.9)+(5.10)+(5.11) has a solution v for any given $\tilde{g}_{ij}$ and $\tilde{T}_{ijk}$, if the $(m + m^2 + m^2 + m^3)$ vectors

$$\{(\partial_i f, \partial_{ij} f, A(\partial_i f, \partial_j f), A(\partial_i \partial_j f, \partial_k f)\} \text{ for all } i, j = \overline{1, q}$$

are linearly independent at all $p \in M$. Following Gromov we shall call such an embedding (g^0, T^0) **-free embedding**. Thus if f is a (g^0, T^0)-free C^k-embedding, then the linearized equation (5.6)+(5.7) of our perturbation problem (5.4)+(5.5) at f has a solution v, because it can be chosen as a solution of the linear (w.r.t the second derivative of f) system (5.8)+(5.9)+(5.10)+(5.11).

We also observe that if f is a (g^0, T^0)-free embedding, then the composition $f \circ h$, where h is an embedding of linear statistical spaces $(\mathbf{R}^q, g^0, T^0) \to (\mathbf{R}^p, g^0, T^0)$ is also a (g^0, T^0) -free embedding. (Here we abuse the notion g^0, T^0 for the standard statistical structure on $\mathbf{R}^q$ with possible different dimension q.)

Now we shall show that the set of all (g^0, T^0)-free, C^k- embeddings is open, dense in the C^k-topology if $q > m(m+1)^2$ and $k \geq 2$. First we shall prove the following Lemma.

5.12. Lemma. *Suppose that $q \leq m(m-1)+2$. Then there is an open, dense set S^q_m of vectors $(v_1, \cdots, v_m) \in \mathbf{R}^q \oplus \cdots_{m\, times} \oplus \mathbf{R}^q$ with the following property (PA^q_m)*

all vectors $v_1, \cdots, v_m, A(v_i, v_j), i, j = 1, m,$ are linear independent in $\mathbf{R}^q$.

Proof of Lemma 5.12. We prove by induction on m. The condition P^q_1 for a vector v_1 can be expressed by the following $(q-1)$ equations on coordinates of v_1:

$$v^1_1 = v^2_1 = \cdots = v^q_1. \tag{5.12.1}$$

Thus for $m = 1$ the set S^q_1 of vectors v_1 with (PA^q_1) is a complement to a 1-dimensional subset of in $\mathbf{R}^q$.

Next we assume that the Lemma 5.12 is valid for the set S^q_{m-1} for all $q \geq 2$. We shall prove that the Lemma 5.12 is also valid for the set S^q_m. It suffices to show that S^q_m is $S^q_{m-1} \times D^q_m$, where D^q_m is the complement of a closed set C^q_m of codimension $(q - m(m+1) - 1)$ in $\mathbf{R}^q$.

The condition PA^q_m is the union of the following three conditions. The first one is that $(v_1, \cdots, v_{m-1}) \in S^q_{m-1}$ (the condition P^q_{m-1} for the first $(m-1)$ vectors.) The second condition is that the two vectors $v_m, A(v_m, v_m)$ are linearly independent (the condition PA^q_1 for v_m). The last condition is that these two vectors are together linearly indepedendent to the $(m-1)+(m-1)^2$ vectors $\{v_i, A(v_i, v_j),\ i, j \leq m-1\}$.

We define C^q_m to be the union of the following 2 sets corresponding to the last two conditions on v_m. The first set consists of vectors $v_m \in \mathbf{R}^q$ whose coordinates v^i_m satisfies the equation (5.12.1). The second set consists of those v_m such that the space generated by v_m and $A(v_m, v_m)$ has a non-trivial intersection with the linear subspace in $\mathbf{R}^q$ generated by $(v_i, A(v_i, v_j), i, j = 1, (m-1))$. Clearly the set C^q_m is a closed subset of dimension m(m-1)+1 in $\mathbf{R}^q$. Since $q \leq (m-1)m + 2$ the complement D^q_m von C^q_m in $\mathbf{R}^q$ is an open dense set. It is straightforward to verify that the set $S^q_m = S^q_{m-1} \oplus D^q_m$ satisfies our condition. This completes the induction step and completes the proof of Lemma 5.12. □

Continuation of step 1. Clearly the condition that f is a (g^0, T^0) -free mapping is equivalent to the following condition. For each $p \in M$ the set of $m(m+1)$ vectors $(\partial_i f(p), \partial_i\partial_j f(p))$ satisfies the condition $PA^q_{m(m+1)}$. The following Proposition summarizes the main result of step 1.

5.13. Proposition. *The set of all C^k, $k \geq 2$, and (g^0, T^0) -free embeddings from M to $\mathbf{R}^q$ is an open dense set in C^k-topology, if $q \geq (m(m+1)-1)m(m+1)+2+m$. For each C^k- and (g^0, T^0)-free embedding f the linearized equation (5.5)+(5.6) at f has a C^{k-2} solution v.*

Proof. If f is a free map, then the existence of a solution v follows from the existence of a solution of the system (5.8) + (5.9)+ (5.10) +(5.11) as we have shown above. To show that v is of class C^{k-2} we need to show a canonical way to get a unique solution v. We note that for a given f and at each p the set of all v which satisfies (5.8)+(5.9)+(5.10)+(5.11) is an affine subspace. Thus we can chose a unique solution $v(p)$ as the minimizer of the norm $|v(p)|$ (as Nash did for isometric embeddings).

It remains to prove the first statement on the open dense property. We denote by $\Gamma^{(2)}(M, \mathbf{R}^q)$ the space of 2-jets of C^k-mappings from M^m to $\mathbf{R}^q$ with $2 \leq k \leq \infty$. The condition of that f is (g^0, T^0) -free is equivalent to the following. For each $p \in M$ the basis vectors of $T_f(M,p) =< \partial_i f(p), \partial_i\partial_j f(v), i, j = 1, m >$ satisfy the condition $PA^q_{m(m+1)}$. (It is easy to check that this condition does not depend on the choice of coordinates at p). Lemma 5.12 shows that this condition is an open differential relation on the jet-space $\Gamma^{(2)}(M^m, \mathbf{R}^q)$ if $q \geq (m(m+1)-1)m(m+1)+2$. Thus the space (g^0, T^0) -free mappings is an open map. To show that this space is dense we use the following corollary of the Thom transversality theorem whose proof is given in [Gr].

5.13.1. Lemma. ([Gr, Corollary 1.3.2.D']). *If $S \subset \Gamma^{(r)}(M^m, \mathbf{R}^q)$ has codimension at least $m+1$, then for a generic map $f \in C^k(M^m, \mathbf{R}^q)$ (i.e. for f in a countable interstion of open dense sets) the jet $f^r \in \Gamma^{(r)}(M^m, \mathbf{R}^q)$ has no intersection with S.*

We apply this Lemma for the complement S of the open set defined by properties $PA^q_{m(m+1)}$ in $\Gamma^{(2)}(M^m, \mathbf{R}^q)$. This proves the dense property of (g^0, T^0)-free embeddings. □

Step 2. Implicit function theorem.

Now we can apply the Nash-Gromov implicit function theorem [Gr, 2.3.2] to show that

5.14. Proposition. (Perturbation device) *Suppose that $k \geq 3$. For any smooth (g^0, T^0)-free C^k embedding $f : M^m \to \mathbf{R}^q$ there exists a number $E(f)$ such that if (g, T) is a C^k-statistical structure on M^m with*

$$|g - f^*(g^0)|_{C^3} \leq E(f) \text{ and } |T - u^*(T^0)|_{C^3} \leq E(f),$$

then there exists a perturbed map $\tilde{f}$ such that $\tilde{u}^*(g^0) = g$ and $\tilde{u}^*(T^0) = T$.

Proof. It suffices to show that the Nash-Gromov implicit function theorem [Gr, 2.3.2] can be applied in our case. First we recall some neccesary definitions and notations in [Gr].

Let $X \to M^m$ and $G \to M^m$ smooth vector bundles over a smooth manifold M^m. We denote by $X^{(r)}$ the space of (r)-jets of sections of X and by $\mathcal{X}^\alpha$ (resp. $\mathcal{G}^\alpha$) the space of C^α sections of X (and of G respectively). Suppose that $\mathcal{D} : \mathcal{X}^r \to \mathcal{G}^0$ is a smooth differential operator of oder r, in other words, there exists a smooth map $\triangle : X^{(r)} \to G$ and $\mathcal{D}(x) = \triangle \circ J^r_x$. Here $J^{(r)}(x)$ is the extension of the section $x : M^m \to X$ to a section $M^m \to X^{(r)}$.

For any set $\mathcal{A}(= \mathcal{A}^d) \subset \mathcal{X}^d$ we denote by $\mathcal{A}^{d+\alpha}$ the intersection $\mathcal{A}(= \mathcal{A}^d) \cap \mathcal{X}^{\alpha+d}$.

A differential operator $\mathcal{D}$ of oder r is called **infinitesimal invertible** over a subset $\mathcal{A} = \mathcal{A}^d \subset \mathcal{X}^d$, if there exists a family of linear differential operators (of order s) $M_x : \mathcal{G}^s \to Yy$ for $x \in A$ with the following 3 properties.

(1) $\mathcal{A} = \mathcal{A}^d \subset \mathcal{X}^d$ consists of C^d-solutions of an open differential relation $A \subset X^{(d)}$. This number d is called **the defect** of the infinitesimal inverse M. We require that $d \geq r$.

(2) The operator $M_x(g) = M(x, g)$ is a differential operator of oder d in x. Moreover the global operator

$$M : \mathcal{A}^d \times \mathcal{G}^s \to \mathcal{T}^0 = T(\mathcal{X}^0)$$

is also a differential operator , i.e. it is defined via a smooth map from $A \oplus G^{(s)} \to T_{vert}(X)$.

(3) M_x is a right inverse of the linearization $L\mathcal{D}_x$ of $\mathcal{D}$ at x. In other words $L\mathcal{D}(x, M(x, g)) = g$ for all $x \in \mathcal{A}^d$ and $g \in \mathcal{G}^s$.

Nash-Gromov implicit funtion theorem. [Gr, 2.3.2] *Suppose that $\mathcal{D} : \mathcal{X}^r \to \mathcal{G}^0$ is a smooth differential operator of oder r which admits an infinitesimal inverse M of oder s and of defect d over an open set $\mathcal{A} = \mathcal{A}^d \subset \mathcal{X}^d$. Let us fix a number σ such that*

$$\sigma > \bar{s} + 1 = \max(d, 2r + s) + 1.$$

Then for any $x_0 \in \mathcal{A}^\infty$ there exists a fine $C^{\bar{s}+s+1}$-neighborhood $\mathcal{B}_0$ of zero in $\mathcal{G}^{\bar{s}+s+1}$ such that for each $C^{\sigma+s}$-section $g \in \mathcal{B}_0$ the equation $\mathcal{D}(x) = \mathcal{D}(x_0) + g$ has a C^σ-solution.

Now let us verify the condition in the Nash-Gromov implicit function theorem for our case. In our case $\mathcal{X}$ is the trivial fibration $\mathbf{R}^q \times M^m$ and $\mathcal{G} = S^2(T^*M^m) \oplus S^3(T^*M^m)$. Operator $\mathcal{D} : \mathcal{X}^1 \to \mathcal{G}^0$ with $\mathcal{D}(f) = f^*(g^0, T^0)$ is a smooth differential operator of first oder. Proposition 5.13 states that operator $\mathcal{D}$ is infinitesimal invertible over the set $\mathcal{A} = \mathcal{A}^2 \subset \mathcal{X}^2$ consisting of (g^0, T^0)-free embeddings. Indeed this set $\mathcal{A} = \mathcal{A}^2$ consists of C^2-solutions of an open differential relation A on X^2. Thus the condition (1) is satisfied.

For $x(= f) \in \mathcal{A} = \mathcal{A}^2$ and $(g) \in \mathcal{G}^0$ we denote by $M(x, g)$ the unique solution of (5.8)+(5.9)+(5.10)+(5.11) which satisfies the minimizing property (in Proposition 5.13 such a map x is denoted by f.) It is straightforward to check that the global operator

$$M : \mathcal{A}^2 \times \mathcal{G}^0 \to \mathcal{T}^0 = T(\mathcal{X}^0)$$

is a smooth differential operator. Thus the condition (2) is satisfied.

The condition (3) also follows from Proposition 5.13.

This completes also our proof of Proposition 5.14. □

<u>Step 3. C^k-embedding of (M^m, g, T) into a linear statistical space.</u> We fix now our free C^k-embedding $f : M \to \mathbf{R}^{N(m)}$ together with the constant $E := E(f)$ obtained by applying Proposition 5.14 to f. Here we set $N(m) = (m(m+1)-1)m(m+1)+2-m$. Thank to the Nash isometric embedding [N2] and the dense property of (g^0, T^0)-free embeddings, we can assume that f is sufficient close to an isometric embedding $f_i : M^m \to \mathbf{R}^{N(m)}$ with respect to the metric $(1/2)g$. Hence $g^1 = g - f^*(g^0) \sim g - f_i(g^0) = (1/2)g$ is a positive definite form. Such an embedding f is called short w.r.t. g. We compose f with an embedding $h : \mathbf{R}^{N(m)} \to \mathbf{R}^{2N(m)}$ in Theorem 3.13.b to get a new map $u = h \circ f$, which still free, short w.r.t g and more over $u(T^0) = 0$. Now we want to use the Nash approximation trick by adding an extra dimension $q(m)$ in the target space $\mathbf{R}^{2N(m)}$.

5.15. Lemma. *We set $q(m) = (m+2)(m+3)$. There exists a C^k embedding $u_1 : M^m \to \mathbf{R}^{q(m)}$ such that*

$$|g - u^*(g^0) - u^1(g^0)|_{C^3} \leq E(u), \qquad (5.15.1)$$

(5.15.2) $$u^1(T^0) = 0.$$

Proof. We can get the existence of u^1 satisifying (5.15.1) by applying the Nash approximation method [N2, Gr]. Namely we note that the form $g^1 = g - u^*(g^0)$ is positiv definite, therefore it can be approximated in C^k-topology by metrics g^1_j which is pull back of g^0 via some C^α-embedding $f_j : M^m \to V^{2l(m)}$. Here we always assume that $k \geq 3$. This number $2l$ first was obtained by Nash and then it was estimated from above via $(m+2)(m+3)/2$ by Gromov [Gr]. We denote by $h : \mathbf{R}^{q(m)} \to \mathbf{R}^{2q(m)}$ the isotropic embedding in Proposition 3.13.b. Now we let $u_1 = h \circ f_1$ such that

$$|f_1(g^0) - g^1|_{C^3} < E(u).$$

Clearly u_1 satisfies (5.15.1) and (5.15.2). □

We shall combine this approximating Lemma 5.15 with our perturbation device (Proposition 5.14) to get the following C^k-embedding into a linear statistical space. This embedding is the main result of step 3.

5.16. Proposition. *For any statistical manifold (M^m, g, T) there exists a positive number A and such that (M^m, g, T) can be embedded into the linear statistical space $(\mathbf{R}^{N(m)}, g^0, A \cdot T^0)$.*

Proof. We fix a (g^0, T^0)-free C^k-embedding u and an extra C^k-embedding u_1 as in Lemma 5.15. According to our perturbation device (Proposition 5.14) there exists an C^k-immersion $u_2 : M^m \to \mathbf{R}^{2N(m)}$ such that

$$u_2^*(g^0) - u^*(g^0) = g - u^*(g^0) - (u_1)^*(g^0),$$

$$u_2 * (T^0) - u^*(T^0) = T \cdot \frac{E(u)}{|T|_{C^3}}.$$

(This is the only place, where we need the C^3-boundedness of T in our proof of Theorem 5.2).

Now we consider the following C^k-embbedding $(u_2, u_1) : M^m \to \mathbf{R}^{2N(m)+q(m)}$. First we compute that

$$(u_2, u_1)(g^0) = (u_2)^* g^0 + (u_1)^* g^0 = g,$$

$$(u_2, u_1)(T^0) = (u_2)^*(T^0) = T \cdot \frac{E(u)}{|T|_{C^3}}.$$

We let $A = \frac{|T|_{C3}}{E(u)}$. Then our embedding (u_2, u_1) satistifies the condition of Proposition 5.16. □

Step 4. Embedding into Cap^N

From Proposition 5.16 we get that for any compact statistical manifold (M^m, g, T) there exists an embedding (M^m, g, T) to $(\mathbf{R}^{2N(m)+q(m)}, g^0, [|T|_{C^3}/E(u)]\cdot T^0)$. Thus the existence of embedding of C^3-bounded statistical manifolds into Cap^N for $N = 4(2N(m)+q(m))$ is a consequence of the following Lemma which can be considered as a generalization of Proposition 4.8.

5.17. Lemma. *Any bounded domain in a linear statistical manifold $(\mathbf{R}^n, g^0, A \cdot T^0)$ can be realized as a submanifold of Cap^{4n}.*

Proof of Lemma 5.17. The same proof as in proof of Proposition 4.8 yields that the diameter with weight ρ of the positive quadrant part $S^{N-1}(\lambda)_+$ of the sphere of radius λ (in $\mathbf{R}^N$ with T defined by (2.6.1))is equal infinity, if $\lambda < 1$ and $N \geq 4$. Now we embed $(R^n, g^0, A \cdots T^0)$ into the direct product of n copies of $S^3(\lambda)_+$ such that $n \cdot \lambda^2 = 4$. Clearly this product can be embedded into Cap^{4n}. □

Proof of Theorem 5.2 for the non bounded case. We can deal with this case by using the compact decomposition of M^m as Nash did for the isometric embedding in [N2]. Namely we cover $\mathcal{M}^m$ by disk neighborhoods N_i in such a way that we can divide N_i among $(m+1)$ classes where: No two N_i of the same class overlap, each N_i overlaps only a finite number of other N_i. Now we "compactify" N_i via an surjective mapping : $\phi_i : N_i \to S_i$, where S_i is a sphere of the same dimension m. The map ϕ_i can be extended to the whole M since it maps the boundary of N_i into the north point of the sphere. On the other hand, this map ϕ_i is injective in a large (enough) subdomain $\bar{N}_i \subset N_i$. We can furthermore use the unity partition function to define statistical structure on each S_i such that the (sum of) pill back via ϕ_i is the given statistical structure on M. In other words we can consider the statistical structure on M as induced from (infinitely many) spheres S_i via mapping ϕ_i.

Now for each class C of coverings N_i we shall define an embedding

$$\psi_C : M^m \to S^{4(2N(m)+q(m))}(\frac{2}{\sqrt{n+1}}).$$

via the embedding of compact S_i into these spheres, considered as submanifolds of $\mathbf{R}^m$ with T defined by (2.6.1). The only point we have to modify the

Nash argument is that, we cannot construct $\psi_C : S_i \to S^{4(2N(m)+q(m))}(\frac{2}{\sqrt{n+1}})$ at the same time (for all i), but we have to construct them inductively to get them not being overlapped. Now the product mapping $\psi = \psi_1 \times \cdots \times \psi_{m+1}$ is the desired imebedding. □

5.18. Remark-Problem. We conjecture that our embedding theorem is also valid for C^1 structures, moreover we can lower dimension of the ambient space Cap^N in this case.

5.19. Embeddings into quantum statistical manifolds.

Quantum statistical manifolds are an important class of statistical models, so we devote a section for consideration of these manifolds.

As we have seen in section 2, we can define a statistical structure (g, T) on a given manifold M once a divergence funtion ρ is given on $M \times M$. For classical statistical models, the divergence function ρ can be taken as the relative entropy funtion (also the Kullback divergence, the Kulback-Leibler information) or as the Jensen function of the entropy function. In the quantum world the analog of the entropy funtion is the Neumann entropy and we also have have the quantum relative entropy.

Let $\mathcal{H}^n$ be a Hilbert space. A **density operator** ρ is a non-negative defined Hermittian operator on $\mathcal{H}$ whose trace is equal to 1. We denote by $\mathcal{S}$ the set of all density operators on $\mathcal{H}$. We define by $\mathcal{S}_k$ the subset of operators ρ in $\mathcal{S}$ with $rk\,\rho = k$. It is well known that the space $\mathcal{S}_1$ can be identified with the space $\mathbf{C}P^{n-1}$.

A **quantum statistical model** is a submanifold M^m in a $\mathcal{S}$ with a statistical structure being derived from the **Neuman entropy**

$$H(\rho) := tr(\rho \ln \rho)$$

or the quantum relative entropy

$$D^1(\sigma, \rho) = tr[\rho(\ln \rho - \ln \sigma)].$$

It is known that a metric torsion of a Fisher metric derived from this quantum relative entropy can be non-zero. In otherword the tensor T may not be 3-symmetric. We call such a statistical structure **true quantum statistical structure.** Clearly no true quantum statsitical manifold can be realized as a submanifold of a classical statistical manifold (because the tensor T of a classical statistical manifold is a 3-symmetric tensor). It is also interesting question to consider the embedding of Cap^N into a "universal" quantum statistical manifolds.

References

[Ay] N. Ay, *An information-geometric approach to a theory of pragmatic structuring*, Ann. of Prob., v.30, N. 1, 416-436 (2002).

[A-N] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Trans. of Math. Monograph (2000).

[Am] S. Amari, *Information geometry*, Contemporary Math. vol 203, 81-95, (1997).

[C] N.N. Chentsov, *Statistical decision rules and optimal inference*, AMS., (1982), (originally published in Russian, Nauka, Moscow , 1972).

[E] S. Eguchi, *Geometry of minimum contrast*, Hiroshima Math. J. 22 (1992), 631-647.

[G] M. Gromov, *Partial Differential Relations,* Springer-Verlag, (1986).

[J] J. Jost, *Information geometry*, preprint MPI 2002.

[K-N] S. Kobayashi and K. Nomizu, *Foundation of differential geometry, v. I*, Interscience Publisher, 1963.

[L] S. Lauritzen, *Statistical manifolds*, In : Differential geometry in Statistical Inference, IMS Lecture Notes, Mongraph Serie 10., Inst. of Math. Stat. Hayward, California, (1987).

[Ma] T. Matumoto, *Any statistical manifold has a contrast function - On the C^3-functions taking the minimum at the diagonal of the product manifold*, Hiroshima Math. J., 23 , 327-332, (1993).

[M-C] E.A. Morozova and N.N. Chentsov, *Markov invariant geometry on manifolds of states* (in Russian), Itogi Nauki i Techniki 36, 69-102, (1990).

[N1] J. Nash, *C^1-isometric imbeddings*, Ann. Math. 60, 383-396, (1954).

[N2] J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. of Math., 63, 20-64, (1956).

[O-V] E.B.Vinberg, A.L. Onishchik, *Seminars on Lie groups and algebraic groups*, (in Russian) Nauka, Moscow, (1988).

[R1] C. R. Rao, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. of the Calcutta Math. Soc. 37, 89-91 (1945).

[R2] C. R Rao, *Differential metrics in probability spaces*, In : Differential geometry in Statistical Inference, IMS Lecture Notes, Mongraph Serie 10., Inst. of Math. Stat. Hayward, California, (1987).