## Søren Tolver Jensen

## Covariance Hypotheses which are Linear in both the Covariance and the Inverse Covariance



# S申ren Tolver Jensen <br> COVARIANCE HYPOTHESES WHICH ARE LINEAR <br> IN BOTH THE COVARIANCE AND <br> THE INVERSE COVARIANCE 

Preprint 1975 No. 1

## 1. INTRODUCTION AND SUMMARY

The purpose of the present paper is to clarify the following question: For which covariance models do the problems of statistical inference have an explicit solution ?

The question is relevant if one wishes to develop a general theory for covariance models because then one has to make a compromise between, on the one side, to make it as general as possible in order to cover most of the relevant examples and, on the other side, to provide that it can produce satisfactory solutions of concrete models and to make a right compromise it is necessary to know the nice models.

Since there has been a tendency to develop very general theories we shall here be rather concrete and only be interested in covariance models for which the maximum likelihood estimator and its distribution, in the case where the observation is normally distributed, can be given explicitiy.

The most general theory for statistical models which deals with exact problems is the theory of canonical exponential families, see Barndorff-Nielsen (1970). For the multidimensional normal distribution with mean zero the canonical hypotheses are those which are affine in the inverse covariance; for these models we have, for any choice of a basis in the sample space, that the orthogonal projection, with respect to the trace form, of the empirical covariance onto the space of inverse covariances is a minimal sufficient statistic, and the result of the general theory is that the maximum likelihood estimator, if it exists, is a unique solution to the likelihood equation and that this equation is given by setting the projection of the empirical covariance equal to the projection of the variance, see also Anderson (1969), Section 5. However, this equation cannot be solved explicitly in general and we have to make some restrictions. Since what one understands by an explicit solution is not objective, we shall not try to show that some restrictions are necessary, but only make one assumption which seems natural, namely that the hypothesis are linear in the
covariance too; in fact, the hypotheses which one can understand are those which are linear in the covariance while the condition that they are linear in the inverse covariance only shall provide that they are mathematically tractable.

In this paper we shall prove a structure theorem for covariance models which are linear in both the covariance and the inverse covariance and from this structure theorem show that these models can be solved explicitly.

In Section 2 we prove that the models can be characterized by means of Jordan algebras. In Section 3 and 4 we have stated the results from the theory of Jordan algebras which we shall use. In Section 6 we prove the structure theorem for the models. The theorem says that the models are products of models each of which being either I. independently identically distributed (multidimensional) real, complex, or quaternion variables with a completely unknown covariance or II. independently identically distributed (multidimensional) variables with a parametrization of the covariance which is given by means of the Clifford algebra. The models $I$ are known and it has been shown that they have explicit solutions, see Anderson (1958), Goodman (1963), Khatri (1965), and Andersson(1972). The models $I I$ are not known before and in Sections 7 and 8 we show that the problems of estimation and testing have rather explicit solutions; it turns out that the distribution problems are not more complicated than those for the twodimensional normal distribution. In Section 9 we have given one of the possible representations of the models which have a dimension which is less than sixteen. Finally in Section 10 we briefly discuss how the general theory of exponential families works on the models we treat.

Anderson (1969), (1973), (1974) has treated covariance models with linear structure but, as he states, the exact distribution of the maximum likelihood estimator can not be obtained in closed form in general. Seely (1971), (1972) has treated covariance models for which the smallest affine subspace containing the family of unknown covariances is a linear space $L$ and
(assuming $I \in L$ ) he prove that such a model admits a complete sufficient statistic if and only if L satisfy condition (2.3) in this paper (he called such a subspace a quadratic subspace but the right name must be a Jordan algebra). If one combines Seely's and our results one get that the linear covariance models which admit a complete sufficient statistic precisely are those which are characterized in Section 6 .
2. COVARIANCE MODELS GIVEN BY JORDAN ALGEBRAS.

Let $E$ be a real vector space and $\theta$ a non empty family of regular covariances on E. In the following we shall always assume that $\theta$ is a linear space; i.e. there exists a subspace L of the vector space $B_{S}\left(E^{*}\right)$ of symmetric bilinear forms on $E^{\prime} s$ dual space $E^{*}$ such that

$$
\begin{equation*}
\theta=\{\Sigma \in L \mid \Sigma \text { is positive definite }\} \text {. } \tag{2.1}
\end{equation*}
$$

The two lemmas below give necessary and sufficient conditions in order that the family of inverse covariances

$$
\theta^{-1}=\left\{\delta \in B_{S}(E) \mid \delta^{-1} \in \theta\right\}
$$

is a linear space too; i.e. that there exists a subspace M of the vector space $B_{S}(E)$ of symmetric bilinear forms on $E$ such that

$$
\begin{equation*}
\theta^{-1}=\{\delta \in M \mid \delta \text { is positive definite }\} \tag{2.2}
\end{equation*}
$$

We choose a fixed element $\delta_{0} \in \theta^{-1}$. With $\delta_{0}$ E is a vector space with an inner product and we can identify $E^{*}$ with $E$. Hereby both $B_{S}(E)$ and $B_{S}\left(E^{*}\right)$ are identified with the symmetric linear mappings of $E$ into itself and we can consider $\theta$ and $\theta^{-1}$ as subsets of the same vector space. Moreover, since $\delta_{0}$ and $\delta_{o}^{-1}$ correspond to the identical mapping of $E$ into itself, we have $1 \in \theta$ and $1 \in \theta^{-1}$.

Lemma 1. $\theta^{-1}$ is a linear space if and only if $\theta^{-1}=\theta$.

Proof. If is obvious. Suppose $\theta^{-1}$ has the form (2.2). Let a L. For $\lambda$ sufficiently smal1 $1-\lambda a \in \theta$ and $(1-\lambda a)^{-1}=$ $1+\lambda a+\lambda^{2} a^{2}+\ldots \epsilon \theta^{-1}$. Since $1 \in \theta^{-1},\left((1-\lambda a)^{-1}-1\right) / \lambda=$ $a+\lambda a^{2}+\ldots \in M$ and, letting $\lambda$ tend to zero, we have $a \in M$. Thus $L \subseteq M$ and by symetry it follows that $M \subseteq L$. Since $L=M$ we have $\theta=\theta^{-1}$.

Lemma 2. $\theta^{-1}$ is a linear space if and only if

$$
\begin{equation*}
\forall a, b \in L: a b+b a \in L \tag{2.3}
\end{equation*}
$$

Proof. Only if: It follows from lemma 1 that $L=M$. Let $a \in L$. For $\lambda$ sufficiently smal1 $1-\lambda a \in \theta$ and $(1-\lambda a)^{-1}=$ $1+\lambda a+\lambda^{2} a^{2}+\ldots \epsilon \theta^{-1}$. Since $1, a \in L,\left((1-\lambda a)^{-1}-1-\lambda a\right) / \lambda^{2}=$ $a^{2}+\lambda a^{3}+\ldots \in L$ and, letting $\lambda$ tend to zero, we have $a^{2} \in L:$ Hence $a b+b a=(a+b)^{2}-a^{2}-b^{2} \in L$ for $a, b \in L$. If: Let $a \in \theta$. By induction it follows from (2.3) that $a^{n} \in L$ for $n \geq 0$. For $\lambda$ sufficiently small and positive $a=(1-(1-\lambda a)) / \lambda$ and, since $1-\lambda a \in L, a^{-1}=\lambda\left((1-\lambda a)+(1-\lambda a)^{2}+\ldots\right) \in L$. Thus $\theta \subseteq \theta^{-1}$ and $\theta^{-1} \subseteq\left(\theta^{-1}\right)^{-1}=\theta$ and it follows from lemma 1 that $\theta^{-1}$ is a linear space.

The condition (2.3) says that L is a Jordan algebra of symmetric linear mappings of $E$ into itself. Now the theory of Jordan algebras is extensively treated in the literature, see Jacobson (1968), or Braun and Koecher (1966), and in the forlowing two sections we shall merely collect the results which we shall need later on.
3. THE STRUCTURE OF JORDAN ALGEBRAS.

A Jordan algebra over the real numbers $R$ is a real vector space J with a composition o satisfying

$$
\begin{align*}
& \forall a, b \in J: a \circ b=b o a,  \tag{3.1}\\
& \forall \lambda \in R \forall a, b \in J:(\lambda a) \circ b=\lambda(a \circ b),  \tag{3.2}\\
& \forall a, b_{1}, b_{2}, \in J: a \circ\left(b_{1}+b_{2}\right)=a \circ b_{1}+a \circ b_{2}, \tag{3.3}
\end{align*}
$$

$$
\begin{equation*}
\forall a, b \in J: \quad\left(a^{2} o b\right) \quad a \quad a=a^{2} o(b o c a) . \tag{3.4}
\end{equation*}
$$

Thus the composition o is commutative and distributive with repect to + , but the associative law is replaced by (3.4).

For an associative algebra $A$ we can define a new composition by setting

$$
\begin{equation*}
a \circ b=\frac{1}{2}(a b+b a) \tag{3.5}
\end{equation*}
$$

for $a, b \in A$, and it is easy to see that $A$ with the composition o is a Jordan algebra; it is denoted $A^{+}$.

A Jordan algebra J is called special if there exists an associative algebra $A$ such that $J$ is a Jordan subalgebra of $A^{+}$.

A Jordan algebra $J$ is called formally real if $a^{2}+b^{2}=0$ implies $a=0$ and $b=0$.

We can now formulate condition (2.3) in lemma 2 by saying that $L$ is a Jordan subalgebra of (End $\mathbb{R}^{+}{ }^{+}$, where End $R_{R}$ is the linear mappings of $E$ into itself. Hence $L$ is special and, since $L$ consists of symmetric mappings, it is clear that $L$ is formally real.

The structure of a finite-dimensional special formally real Jordan algebra is completely known according to the following theorem of Jordan, von Neumann, and Wigner:

Theorem 1. Let J be a finite-dimensional, special, and formally real Jordan algebra. Then $J$ is somorphic to a product of Jordan algebras, $J=J_{1} \times \ldots x J_{k}$, where each $J_{i}, i=1, \ldots, k$, is isomorphic to one of the following: (i)R, (ii) $R \times V$, where $V$ is an m-dimensional real vector space, m $\geq 2$, with a positive definite form $\phi$ and where the composition in $R \times V$ is given by

$$
\begin{equation*}
(\alpha, x) \circ(\beta, y)=(\alpha \beta+\phi(x, y), \beta x+\alpha y), \tag{3.6}
\end{equation*}
$$

(iii) $H_{r}(D), r \geq 3$, where $D$ is one of the classical fields, the real numbers R, the complex numbers $C$, or the quaternions $H$, and $H_{r}(D)$ the Hermitian $r \times r-m a t r i c e s$ with elements in $D$.

Proof. See Jacobson (1968), p 205, Braun and Koecher (1966), p. 331, or Jordan, von Neumann, and Wigner (1934), p. 63.

Two Jordan algebras $R \times V_{1}$ and $R \times V_{2}$ of type (ii) are isomorphic if dim $V_{1}=\operatorname{dim} V_{2}$. Apart from this case none of the Jordan algebras mentioned in theorem 1 are isomorphic. We shall see in Section 9 that $R \times V$ for $m=2,3$, and 5 is isomorphic to $H_{2}(D)$ for $D=R, C$, and $H$.

## 4. REPRESENTATIONS OF JORDAN ALGEBRAS.

Theorem 1 is not quite sufficient for our purpose; wish to know, not only the structure of a Jordan algebra, but also in which way it can be represented as symmetric linear mappings of $E$ into itself. We shall need the notion of a unital special universal envelope algebra, in the following abbreviated usuea.

Let $J$ be a Jordan algebra with a unit; an associative algebra $A$ and a Jordan algebra homomorphism $\sigma: J \rightarrow A^{+}$sending 1 into 1 is called a usuea for $J$ if, for any associative algebra B and any Jordan algebra homomorphism $f: J \rightarrow B^{+}$sending 1 into 1 , there exists a unique algebra homomorphism $\eta: A \rightarrow B$ sending 1 into 1 such that $f=\eta \circ \sigma$ :


Every Jordan algebra has one and, upto isomorphism, only one usuea and it is generated by the image of J, see Jacobson (1968), p. 73. If J is a special Jordan algebra there exists a B and an injective $f$ and it follows that $\sigma$ must be injective.

The usuea for a product of Jordan algebras is a product of the usuea's, see Jacobson (1969), p. 74. Thus we only need to know the usuea's for the Jordan algebras mentioned in
theorem 1. They are: (i) for $R: R$ with $\sigma: R \rightarrow R$ the identical mapping, see Jacobson (1968), p. 74, (ii) for $R \times V: T h e C 1 i f-$ ford algebra $C(V, \phi)$ with the canonical mapping $\sigma: R \times V \rightarrow$ $C(V, \phi)$, see Jacobson (1968), p. 74, (iii) for $H_{r}(D), r \geq 3: T h e$ algebra $M_{r}(D)$ of $r \times m$-matrices with elements in $D$ and $\sigma:$ $H_{r}(D) \rightarrow M_{r}(D)$ the inclusion mapping, see Jacobson (1968), p. 143.

For definition and properties of the Clifford algebra see Chevalley (1954) -or Bourbaki (1959); see also below.

According to Wedderburn's structure theorem the finite-dimensional simple associative algebras over R have the form End ${ }_{D} S$ where $D$ is $R, C$, of $H, S$ a finite-dimensional vector space over $D$, and End $S$ the $D$-linear mappings of $S$ into itself, see Bourbaki (1958), p. 49. We shall always assume that $S$ is a
 is isomorphic to $M_{r}(D)$ where $r=d i m_{D} S$.

We can say, therefore, that the usuea for a Jordan algebra of type (i) or (iii) is a simple algebra End $S$ and that there exists a D-basis in $S$ such that the elements in the Jordan algebra by $\sigma$ correspond to Hermitian matrices.

The structure of the Clifford algebra $C(V, \phi)$, where $\phi$ is a positive definite form on $V$ and $m=d i m V, i s$ given by the following table:

| m | $C(\mathrm{v}, \phi)$ | D | $\beta(\mathrm{m})=\mathrm{dim}_{\mathrm{D}} \mathrm{S}$ | $\alpha(m)=\mathrm{dim}_{\mathrm{R}} \mathrm{S}$ |
| :---: | :---: | :---: | :---: | :---: |
| $8 x+1$ | $E \mathrm{nd}_{R} \mathrm{~S} \times \mathrm{E} \mathrm{nd}_{\mathrm{R}} \mathrm{S}$ | R | $2^{4 x}$ | $2^{4 x}$ |
| $8 x+2$ | $E n d_{R} \mathrm{~S}$ | R | $2^{4 x+1}$ | $2^{4 x+1}$ |
| $8 x+3$ | End ${ }_{C} \mathrm{~S}$ | C | $2^{4 x+1}$ | $2^{4 x+2}$ |
| $8 x+4$ | End ${ }_{H} \mathrm{~S}$ | H | $2^{4 x+1}$ | $2^{4 x+3}$ |
| $8 x+5$ | $\operatorname{End}_{H} \mathrm{~S} \times \mathrm{End}_{H} \mathrm{~S}$. | H | $2^{4 x+1}$ | $: 2^{4 x+3}$ |
| $8 x+6$ | End ${ }_{H} \mathrm{~S}$ | H | $2^{4 x+2}$ | $2^{4 x+4}$ |
| $8 x+7$ | End ${ }_{C} \mathrm{~S}$ | C | $2^{4 x+3}$ | $2^{4 x+4}$ |
| $8 x+8$ | End $\mathrm{R}^{\text {S }}$ | R | $2^{4 x+4}$ | $2^{4 x+4}$ |

For a proof see Chevalley (1954), p. 66.

The usuea for a Jordan algebra of type (ii) is thus either a simple algebra or a product of two identical simple algebras. We shall say we are in case (iia) if $C(V, \phi)$ is simple and in case (iib) otherwise. Thus we are in case (iib) if m-1 is divisible by four.

We choose a basis in $S$ in case (iia) and a basis in each of the two $S^{\prime} s$ in case (iib); then the mapping $\sigma: R \times V \rightarrow \mathbb{C}(V, \phi)$ in case (iia) corresponds to a mapping

$$
\begin{equation*}
\sigma_{m}: R \times V \rightarrow M_{\beta(m)}(D) \tag{4.1a}
\end{equation*}
$$

and in case (iib) to a mapping

$$
\begin{equation*}
\sigma_{m}=\left(\sigma_{m}^{1}, \sigma_{m}^{2}\right): R \times V \rightarrow M_{\beta(m)}(D) x M_{\beta(m)}(D) \tag{4.1b}
\end{equation*}
$$

where $D$ and $\beta(m)=\operatorname{dim}_{D} S$ are given in the table above.

Lemma 3. The basis or the bases in $S$ can be chosen in such a way that $\sigma_{m}$ maps $R \times V$ into Hermitian matrices.

We shall need the following property of the Clifford algebra: The Clifford algebra $C(V, \phi)$ is a real algebra of dimension $2^{m}$ for which there exists an injective linear mapping

$$
i: V \rightarrow C(V, \phi)
$$

such that $i(V)$ generates $\mathbb{C}(V, \phi)$ and such that

$$
\begin{equation*}
\forall v \in V:(i(v))^{2}=\phi(v, v) \cdot 1 \tag{4.2}
\end{equation*}
$$

see Chevalley (1954), p. 40 .

From (4.2) it follows that

$$
\begin{equation*}
\forall v_{1}, v_{2} \in V: i\left(v_{1}\right) i\left(v_{2}\right)+i\left(v_{2}\right) i\left(v_{1}\right)=2 \phi\left(v_{1}, v_{2}\right) 1 \tag{4.3}
\end{equation*}
$$

and from (4.3) it can be seen that $1 \notin i(V)$; the mapping $\sigma: R \times V \rightarrow C(V, \phi)$ is given by $\sigma(\lambda, v)=\lambda \cdot 1+i(v)$.

Proof of lemma 3. We shall only give the proof in case (iib); in case (iia) the proof is simpler. Choose an orthonormal basis $e_{1}, \ldots, e_{m}$ in $V$. Set $x_{j}=i\left(e_{j}\right), j=1, \ldots, m$ and let $G$ be the subset of $C(V, \phi)$ which consist of the elements $\pm 1$ and $\pm \mathrm{x}_{\mathrm{i}_{1}} \ldots \mathrm{x}_{\mathrm{i}_{\mathrm{k}}}, 1 \leq \mathrm{i}_{1} \leq \cdots \leq \mathrm{i}_{\mathrm{k}} \leq \mathrm{m}, \mathrm{k}=1, \ldots, \mathrm{~m}$. It follows from (4.2) and (4.3) that $G$ is a finite group. Choose a positive definite Hermitian form $\psi_{0}$ on $S \times S$ such that the two $S^{\prime} s^{\prime}$ are orthogonal and define a new form $\psi$ on $S \times S$ by

$$
\psi(x, y)=\sum_{g \in G} \psi_{o}(g x, g y)
$$

It is clear that $\psi$ is a positive definite Hermitian form on $S \times S$ and, since the elements of the Clifford algebra map the two $S^{\prime} s$ into themselves, that the two $S^{\prime} s$ are orthogonal with respect to $\psi$. It follows from (4.2) that $\mathrm{x}_{\mathrm{j}}^{-1}=\mathrm{x}_{\mathrm{j}}, \mathrm{j}=1, \ldots \mathrm{~m}$, and we have

$$
\begin{aligned}
\psi\left(x_{j} x, y\right) & =\sum_{g \in G} \psi_{o}\left(g x_{j} x, g y\right)=\sum_{g \in G} \psi_{o}\left(g x, g x_{j}^{-1} y\right) \\
& =\sum_{g \in G} \psi_{o}\left(g x, g x_{j} y\right)=\psi\left(x, x_{j} y\right)
\end{aligned}
$$

Hence the elements $x_{1}, \ldots, x_{m}$ are Hermitian with respect to $\psi$, and the lemma follows if we choose bases in the two $\mathrm{S}^{\mathrm{r}}$ s which are orthonormal with respect to $\psi$.

## 5. HERMITIAN MATRICES.

In this section we shall briefly discuss the connection between complex or quaternion matrices and real matrices.

A complex matrix has the form

$$
\begin{equation*}
\Sigma+i F \tag{5.1}
\end{equation*}
$$

where $\Sigma$ and $F$ are real matrices. If we consider (5.1) as the matrix of a C-linear mapping of a complex vector space with basis ( $e_{1}, \ldots, e_{n}$ ) then the matrix of this mapping with respect to the real basis $\left(e_{1}, \ldots, e_{n}, e_{1} i, \ldots, e_{n} i\right)$ is

$$
\left\{\begin{array}{rr}
\Sigma & -F  \tag{5.2}\\
F & \Sigma
\end{array}\right\}
$$

(5.1) is Hermitian if and only if $\Sigma$ is symmetric and $F$ is antisymmetric; i.e. if and only if (5.2) is symmetric.

In a similar way a quaternion matrix has the form

$$
\begin{equation*}
\Sigma+i F_{1}+\mathrm{jF}_{2}+\mathrm{kF}_{3}, \tag{5.3}
\end{equation*}
$$

where $\Sigma, F_{1}, F_{2}$, and $F_{3}$ are real matrices, and it corresponds to the real matrix

$$
\left\{\begin{array}{ccc}
\Sigma & -F_{1}-F_{2}-F_{3}  \tag{5.4}\\
F_{1} & \Sigma-F_{3} & F_{2} \\
F_{2} & F_{3} & \Sigma-F_{1} \\
F_{3}-F_{2} & F_{1} & \Sigma
\end{array}\right\}
$$

(5.3) is Hermitian if and only if $\Sigma$ is symmetric and $F_{1}, F_{2}$, and $\mathrm{F}_{3}$ are antisymmetric; i.e. if and only if (5.4) is symmetric.

We shall say a real matrix has a complex or quaternion form if it has the form (5.2) or (5.4). Conversely (5.2) and (5.4) are called the real matrices corresponding to (5.1) and (5.3). Usually we use the same notation for a complex or quaternion matrix and its real form. Thus the matrices $\sigma_{m}(\lambda, v), \sigma_{m}^{1}(\lambda, v)$ and $\sigma_{m}^{2}(\lambda, v)$ given by (4.1a) and (4.1b) are, when they are considered as real matrices, $\alpha(\mathrm{m}) \times \alpha(\mathrm{m})$ - matrices where $\alpha(\mathrm{m})=$ $\operatorname{dim}_{R} S$ is given in the table in Section 4.

## 6. THE STRUCTURE THEOREM.

We are now in a position to prove the structure theorem for covariance models which are linear in both the covariance and the inverse covariance.

It follows from Sections 2 and 3 that the models precisely are
those which have the form (2.1), where $L$ is a Jordan algebra of symmetric linear mappings of $E$ into itself and $1 \in L$.

Before we state the theorem it shall be noticed that we by means of a fixed inner product on $E$ have identified the symmet tric bilinear forms on $E^{*}$, the symmetric bilinear forms on $E$, and the symmetric linear mappings of $E$ into itself. If we choose a basis in $E$ then anyone of these objects correspond to a matrix, but two objects which are identified have in general only the same matrix if the chosen basis is orthonormal with respect to the fixed inner product !

Let $L=J_{1} \times \ldots \times J_{k}$ be the decomposition of $L$ into a product of the Jordan algebras mentioned in theorem 1 . We have

Theorem 2. There exists an orthonormal basis in E such that $\theta$ is the family of matrices of the form

$$
\left\{\begin{array}{ccc}
A_{1} & \cdots & 0  \tag{6.1}\\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
0 & \cdots & A_{k}
\end{array}\right\}
$$

where $A_{i}$ is an $N_{i} \times N_{i}$-matrix, $i=1, \ldots, k, d_{i m} E=N_{1}+\ldots+N_{k}$, where the $A_{i}$ 's vary independently, and where the set of the $A_{i}$ 's are given by following parametrizations:
(i) If $\mathrm{J}_{\mathrm{i}}=\mathrm{R}$ :

$$
\begin{equation*}
A_{i}=a I_{N_{i}}, a>0 \tag{6.2}
\end{equation*}
$$

(iia) If $J_{i}=R \times V, m=d i m \quad V$, and $m-1$ is not divisible by four :

$$
\begin{equation*}
A_{i}=\sigma_{m}(\lambda, v) \not I_{n} \tag{6.3a}
\end{equation*}
$$

where ( $\lambda, v) \in R \times V$ such that $\sigma_{m}(\lambda, v)$ is positive definite and where $N_{i}=\alpha(m) n$.
(iib) If $J_{i}=R \times V, m=d i m V$, and $m-1$ is divisible by four:

$$
A_{i}=\left\{\begin{array}{ccccc}
\sigma_{m}^{1}(\lambda, v) & \rtimes I_{n_{1}} & 0 &  \tag{6.3b}\\
0 & & \sigma_{m}^{2}(\lambda, v) & \rtimes I_{n_{2}}
\end{array}\right\}
$$

where $(\lambda, v) \in R \times V$ such that $\sigma_{m}^{1}(\lambda, v)$ and $\sigma_{m}^{2}(\lambda, v)$ are positive definite and where $N_{i}=\alpha(m)\left(n_{1}+n_{2}\right)$.
(iii) $\operatorname{If} J_{i}=H_{r}(D), r \geq 3:$

$$
\begin{equation*}
A_{i}=\Sigma \rtimes I_{n}, \tag{6.4}
\end{equation*}
$$

where $\Sigma$ is a positive definite $q \times q-m a t r i x, ~ q=r \operatorname{dim}_{R} D$, which has a complex form if $D=C$ and a quaternion form if $D=H$ and where $N_{i}=q n$.

Remark. We shall see in Section 7 that $\sigma_{m}(\lambda, v), \sigma_{m}^{1}(\lambda, v)$, and $\sigma_{m}^{2}(\lambda, v)$ are positive definite precisely when $\lambda>0$ and $\lambda^{2}>$ $\phi(v, v)$.

Proof. It follows from Section 4 that the usuea for $L$ is a product of simple algebras $\underset{j}{ } \operatorname{End}_{\mathrm{D}}^{\mathrm{j}} \mathrm{S}_{\mathrm{j}}$. and that we in each $\mathrm{S}_{\mathrm{j}}$ can choose a $D_{j}$-basis such that the mapping $\sigma: L \rightarrow \underset{j}{\operatorname{tand}} \mathrm{D}_{\mathrm{j}} \mathrm{S}_{\mathrm{j}}$ maps the elements of $L$ into Hermitian matrices. As $L \subseteq E^{n} d_{R}$ and $1 \in L$ there exists a unique algebrahomomorphism $\eta$ : $\underset{j}{x} E_{j}{ }_{j} S_{j} \rightarrow E$ such that $a=\eta(\sigma(a))$ for every a $\in L$


By means of $\eta$ E is a modul over $\underset{j}{x} \operatorname{End}_{D_{j}} S_{j}$, and it follows from Bourbaki (1958), p. 42 that there exists left $D_{j}$ vector spaces $T_{j}$ and a bijective $R$-1inear mapping

$$
g: \underset{j}{x} S_{j}{ }^{夭_{D}}{ }_{j} T_{j} \rightarrow E
$$

such that

$$
\begin{equation*}
\eta\left(\left(a_{j}\right)\right)=g \circ\left(\underset{j}{\left(x a_{j} \rtimes 1_{T_{j}}\right)} \circ g^{-1}\right. \tag{6.6}
\end{equation*}
$$

for every $\left(a_{j}\right) \in \underset{j}{x} \operatorname{End}_{D}{ }_{j}{ }_{j}$..
We choose a basis ( $f_{\alpha}$ ) in each $T_{j}$. Together with the given basis ( $e_{\beta}$ ) in $S_{j}$ and the usual real basis ( $\varepsilon_{\gamma}$ ) in $D_{j}$ we obtain
 together we obtain a real basis in $x S_{j}{ }^{\gamma_{D}} \mathrm{~T}_{\mathrm{j}}$. In the follo-
 sponding to this basis.

Set $\left(a_{j}\right)=\sigma(a), a \in L$ From (6.5) and (6.6) we have

$$
\begin{equation*}
a=g \circ\left(\underset{j}{(x} a_{j} \nprec 1_{T_{j}}\right) \circ g^{-1} \tag{6.7}
\end{equation*}
$$

and, by transposition,

$$
\begin{equation*}
a=g^{-1^{*}} \circ\left(\underset{j}{x} a_{j} \nsucc 1_{T_{j}}\right) \circ g^{*} \tag{6.8}
\end{equation*}
$$

Hence

$$
\begin{equation*}
g^{*} \circ \mathrm{~g} \quad \mathrm{o}\left(\underset{j}{(x} a_{j} \not 夭 1_{T_{j}}\right)=\left(\underset{j}{x} a_{j} \rtimes 1_{T_{j}}\right) \circ g^{*} \circ g \tag{6.9}
\end{equation*}
$$

and, since $\sigma(L)$ generates $\underset{j}{ } \operatorname{End}_{j} \mathrm{~S}_{\mathrm{j}}$, it follows from Bourbaki (1958), p. 42 that

$$
\begin{equation*}
g^{*} o g=x_{j}^{x}\left(1_{S_{j}} \nsucc h_{j}\right), \tag{6.10}
\end{equation*}
$$

where $h_{j} \in \operatorname{End}_{D_{j}} T_{j}$ for each $j$.
From (6.10) it follows that the $h_{j}{ }^{\prime} \mathrm{s}$ are Hermitian and positive definite. Hence it follows from Bourbaki (1959), p. 123 that

$$
\begin{equation*}
h_{j}=r_{j} * o r_{j}, \tag{6.11}
\end{equation*}
$$

where $r_{j} \in E n d_{j} T_{j}$ for each $j$. Since $g$ is bijective $h_{j}$ and $r_{j}$ are bijective. From (6.10) and (6.11) we obtain

$$
\begin{aligned}
& g=g^{*-1} o\left(x_{j} 1_{S_{j}} \not{ }^{*} r_{j}^{*} \circ r_{j}\right)
\end{aligned}
$$

$$
\left.\left.=\left(\underset{j}{\left(x 1_{S_{j}}\right.} \nsucc r_{j}\right) \circ g^{-1}\right)^{*} \circ \underset{j}{\left(x 1_{S}\right.} \not r_{j}\right)
$$

and by substitution in (6.7) we have

Let

$$
k=g o\left(\underset{j}{x} 1_{S_{j}} \not \partial{\underset{j}{r}}_{-1}^{r^{-1}}\right)
$$

Then it follows from (6.12) that

$$
\begin{equation*}
k^{*} ð a \circ k=x_{j}\left(a_{j} \nprec 1_{T_{j}}\right) \tag{6.13}
\end{equation*}
$$

We now choose the image of the basis in $\underset{j}{ } S_{j} \gamma_{\mathrm{D}}{ }_{j} \mathrm{~T}_{\mathrm{j}}$ under the mapping $k$ as basis in E. Since $1 \in L$ and $\sigma(1)=\left(1_{S_{j}}\right){ }_{j}$ it follows from (6.13) that $k^{*} o k=\frac{x}{j} 1_{S_{j}} \quad \gamma 1_{T_{j}}$ and it is seen that the chosen basis is orthonormal. Moreover, it is seen from (6.13) that the matrix of a with respect to this basis is the same as the matrix of $\mathrm{x}_{\mathrm{j}}\left(\mathrm{a}_{\mathrm{j}}{ }^{\rtimes} 1_{\mathrm{T}_{j}}\right)$ with respect to the basis in $x S_{j} \gamma_{D} T_{j}$. The theorem then follows from the discussion of the usuea's in Section 4 .

If we assume that the distribution is normal then the structure theorem shows, since uncorrelated normally distributed variables are independent, that the model is a product of independent normal covariance models which have the form (6.2), ( 6.3 a ), (6.3b), or (6.4).
Models of the form (6.2) are of course trivial. Models of the form (6.4) for $D=R$ and $D=C$ are treated in detail in the literature. See e.q. T.W. Anderson (1958) for the real case and Khatri (1965) for the complex case.
S.A. Andersson (1972) has shown that covariance models which can be described by invariance under a group of linear mappings of the sample space are products of covariance models of the form (6.2) or (6.4) for $r \geq 2$. He also shows that the maximum
likelihood estimator and its distribution, for a model of the form (6.4), can be specified in a form which, at the same time, cover all the three cases $D=R, C$, and $H$.

The models of the form (6.3a) and (6.3b) do not seem to be known. In the following sections we shall show, therefore, that the problems of estimation and testing have complete so1utions.
7. MAXIMUM LIKELIHOOD ESTIMATION IN A MODEL OF TYPE (ii).

It is convenient to describe models of the form (6.3a) and (6.3b) in an invariant way.

The sample space is an $N$-dimensional real vector space E, and we assume that the observation $X$ has a normal distribution with mean value zero and that the family of unknown covariances has the form (2.1), where $L$ is a Jordan algebra of type (ii); i.e. we have an m-dimensional, $m \geq 2$, real vector space $V$ with an inner product $\phi$ and an injective Jordan algebra homomorphism $\tau$ of the Jordan algebra $R \times V$ into the Jordan algebra of symmetric linear mappings of $E$ into itself such that

$$
\theta=\{(\lambda, v) \in R \times V \mid \tau(\lambda, v) \text { is positive definite }\}
$$

is the parameter space and

$$
\tau: \theta \rightarrow E n d_{R} E
$$

the parametrization.

It follows from Section 4 that the homomorphism $\tau$ is composed of the mapping $\sigma_{m}: R \times V \rightarrow C(V, \phi)$ and an algebra homomorphism $\eta: C(V, \phi) \rightarrow E n d_{R} E$ Since $\sigma_{m}$ is injective we shall consider $R \times V$ as a subset of $C(V, \phi)$ and in the following omit $\sigma_{m}^{\prime}$; we shall also write $v$ instead of ( $0, v$ ) and $\lambda$ instead of ( $\lambda, 0$ ); thus $\lambda+v=(\lambda, v), \tau(\lambda)=\lambda 1_{E}$, and $\tau(\lambda)+\tau(v)=\tau(\lambda+v)=$ $\tau(\lambda, v)$.

For $v \in V$ have $v^{*}=v$ and $\eta(v)^{*}=\eta(v) ;$ thus $\eta\left(v^{*}\right)=\eta(v)^{*}$
and, since $R \times V$ generate $C(V, \phi)$, it follows that $\eta\left(s^{*}\right)=$ $\eta(s)^{*}$ for every $s \in C(v, \phi)$.
(6.3a) and (6.3b) show that $N=\alpha(m) \cdot n$ where $n \geq 1$ and $\alpha(m)$ is given in the table in Section 4 ; i.e. N has to be divisible by $\alpha(m)$.

It follows from (4.2) that ( $\lambda, v$ ) is invertible if and only if

$$
\begin{equation*}
\lambda^{2} \neq \phi(v, v) \tag{7.1}
\end{equation*}
$$

and in this case $(\lambda, v)^{-1}=\left(1 /\left(\lambda^{2}-\phi(v, v)\right)\right)(\lambda,-v)$. Since $\eta$ is an algebra homomorphism $\tau(\lambda, v)$ is also invertible if and only if (7.1) holds and then

$$
\begin{equation*}
\tau(\lambda, v)^{-1}=\frac{1}{\lambda^{2}-\phi(v, v)} \tau(\lambda,-v) . \tag{7.2}
\end{equation*}
$$

Now the set of positive definite $N \times N$-matrices is a convex connected component in the space of regular symmetric $N \times N$ matrices; hence it follows from (7.1), since $\tau(1,0)=1_{E}$, that $\tau(\lambda, v)$ is positive definite if and only if $\lambda>0$ and $\lambda^{2}>\phi(v, v) ;$ i.e. we have

$$
\theta=\left\{(\lambda, v) \in \mathrm{R} \times \mathrm{v} \mid \lambda>0 \wedge \lambda^{2}>\phi(\mathrm{v}, \mathrm{v})\right\}
$$

The density for the normal distribution with parameter ( $\lambda, v) \in \theta$ is
$\left.(2 \pi)^{-\frac{N}{2}} \right\rvert\,$ det $\left.\tau(\lambda, v)\right|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\left(\lambda^{2}-\phi(v, v)\right)}<\tau(\lambda,-\dot{v}), x x^{\prime}>\right)=$
$\left.(2 \pi)^{-\frac{N}{2}} \right\rvert\,$ det $\left.\tau(\lambda, v)\right|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\left(\lambda^{2}-\phi(v, v)\right)}<\tau(\lambda,-v), p\left(x x^{\prime}\right)>\right)$,
where <, > is the trace form on $E d_{R} E$ and $p$ the orthogonal projection, with respect to <, >, onto $\tau(R \times V)$.

The family (7.3) is a canonical exponential family; thus the likelihood equation is $p\left(x x^{\prime}\right)=E\left(p\left(X X^{\prime}\right)\right)$ or, since $E\left(p\left(X X^{\prime}\right)\right)=$ $p\left(E\left(X X^{\prime}\right)\right)=p(\tau(\lambda, v))=\tau(\lambda, v)$,

$$
\begin{equation*}
p\left(x x^{\prime}\right)=\tau(\lambda, v) . \tag{7.4}
\end{equation*}
$$

It follows from the general theory of exponential families, see Barndorff-Nielsen, theorem 6.8, that $p\left(x x^{\prime}\right) \in \overline{\tau(\theta)}=\tau(\bar{\theta})$ for every $x \in E$. Therefore, we can define random variables (Y, $Y_{1}$ ) with values in $\bar{\theta}=\left\{(\lambda, v) \mid \lambda \geq 0 \wedge \lambda^{2} \geqq \phi(v, v)\right\}$ by

$$
\begin{equation*}
p\left(X X^{\prime}\right)=\tau\left(Y, Y_{1}\right) \tag{7.5}
\end{equation*}
$$

since $\tau$ is injective on $R \times V$.

Then the condition for the maximum likelihood estimator to exist is that $\left(Y_{1}, Y_{1}\right) \in \theta$, or $Y>0$ and $Y^{2}>\phi\left(Y_{1}, Y_{1}\right)$.

In order to find the distribution of (Y, $Y_{1}$ ) we need some more results about the Clifford algebra.

Let $\Gamma=\left\{s \in C(V, \phi) \mid s^{*}=s^{-1} \wedge s v^{-1}=V\right\}$. It is obvious that $\Gamma$ is a group. For $s \in \Gamma$ the mapping $v \rightarrow \operatorname{svs}^{-1}$ of $V$ into itself is denoted $\chi$ (s). From (4.2) it follows that $\phi\left(\operatorname{svs}^{-1}, s^{-1}\right)=$ $\mathrm{svs}^{-1} \mathrm{svs}^{-1}=\mathrm{sv}^{2} \mathrm{~s}^{-1}=\mathrm{s} \phi(\mathrm{v}, \mathrm{v}) \mathrm{s}^{-1}=\phi(\mathrm{v}, \mathrm{v})$; thus $\chi(\mathrm{s})$ is an orthogonal transformation of $V$, and it is easy to see that the mapping

$$
\chi: \Gamma \rightarrow 0(\phi)
$$

of $\mathbb{R}$ into $\phi^{\prime} s$ orthogonal group is a group homomorphism.

Let $s \in V$ such that $\phi(s, s)=1$. It follows from (4.2) and lemma 3 that $s=s^{-1}$ and $s=s^{*}$; hence $s^{-1}=s^{*}$, and from (4.3) it follows that $\chi(s)(v)=s^{-1}=-(v-2 \phi(v, s) s)$ for $v \in V_{;}$thus $s \in \Gamma$ and $\chi(s)=-\rho(s)$, where $\rho(s)$ is the symmetry with respect to the hyperplane orthogonal to s. Now it is well known that any orthogonal transformation can be written as a product of symmetries. For $\pi \in O(\phi)$ there exist, therefore, unit vectors $s_{1}, \ldots, s_{k}$ in $V$ such that $\pi=\rho\left(s_{1}\right) \ldots \rho\left(s_{k}\right)=\xi^{k} X_{1}\left(s_{1}, \ldots, s_{k}\right)$, where $\xi(v)=-v$ for $v \in V$. If det $\pi=1$ it follows, since $\operatorname{det}(\bar{\rho}(s))=-1$, that $k$ is even and we have $\pi=\chi\left(s_{1} \ldots s_{k}\right)$. Thus

$$
\begin{equation*}
\chi(\Gamma) \supseteq 0_{+}(\phi), \tag{7.6}
\end{equation*}
$$

where $O_{+}(\phi)$ is the set of orthogonal transformation, of $V$, with determinant 1 .

Let $v \in V . S i n c e d i m V \geq 2$ there exists an orthogonal transformation $\pi$ with determinant 1 such that $\pi(v)=-v$ and it follows from (7.6) that there exists an $s \in \Gamma$ such that $\operatorname{svs}^{-1}=-\mathrm{v}$; since $\eta$. is an algebra homomorphism we have

$$
\begin{equation*}
\eta(s) \tau(v) \eta(s)^{-1}=\tau(-v) \tag{7.7}
\end{equation*}
$$

For $\lambda \in R$ it is obvious that $\eta(s) \tau(\lambda) \eta(s)^{-1}=\tau(\lambda)$; therefore, we have $\eta$ ís) $\tau(\lambda, v) \eta(s)^{-1}=\tau(\lambda,-v)$. If $(\lambda, v) \in \theta$ it follows from (7.2) that det $\tau(\lambda, v)$ e det $\tau(\lambda,-v)=\left(\lambda^{2}-\phi(v, v)^{N}\right.$, and we obtain

$$
\begin{equation*}
\operatorname{det} \tau(\lambda, v)=\left(\lambda^{2}-\phi(v, v)\right)^{N / 2} \tag{7.8}
\end{equation*}
$$

for $(\lambda, v) \in \theta$.

If we take the trace on both sides of (7.7) we have that $\operatorname{tr} \tau(v)=\operatorname{tr} \tau(-v)=\operatorname{tr}(-\tau(v))$ and it follows that

$$
\begin{equation*}
\left\langle 1_{E}, \quad \tau(v)\right\rangle=0 \tag{7.9}
\end{equation*}
$$

From (3.3) it follows that $\tau\left(\mathrm{v}_{1}\right) \tau\left(\mathrm{v}_{2}\right)+\tau\left(\mathrm{v}_{2}\right) \tau\left(\mathrm{v}_{1}\right)=2 \phi\left(\mathrm{v}_{\mathrm{p}}, \mathrm{v}_{2}\right) 1_{\mathrm{E}}$ for $v_{1}, v_{2} \in V$ and we obtain

$$
\begin{equation*}
\left\langle\tau\left(v_{1}\right), \tau\left(v_{2}\right)\right\rangle=N \phi\left(v_{1}, v_{2}\right) \tag{7.10}
\end{equation*}
$$

If we substitute (7.5) in (7.3) and use (7.8), (7.9), and (7.10), then the density (7.3) can be written
$(2 \pi)^{-\frac{N}{2}}\left(\lambda^{2}-\phi(v, v)\right)^{-\frac{N}{4}} \exp \left(-\frac{N}{2\left(\lambda^{2}-\phi(v, v)\right)}\left(\lambda y-\phi\left(v, y_{1}\right)\right)\right.$.
It follows from (7.5) and (7.9) that $\left\langle Y 1_{E}, 1_{E}\right\rangle=\left\langle\tau(Y), 1_{E}\right\rangle=$ $\left\langle\tau\left(Y, Y_{1}\right), 1_{E}\right\rangle=\left\langle X X^{\prime}, 1_{E}\right\rangle$ and we have

$$
\begin{equation*}
Y=\frac{1}{N} \operatorname{tr}\left(X X^{\prime}\right) \tag{7.12}
\end{equation*}
$$

Hence $Y$ is positive with probability one and we can define a random variable $Z \quad b y$

$$
\begin{equation*}
\mathrm{Z}=\mathrm{Y}_{1} / \mathrm{Y} \tag{7.13}
\end{equation*}
$$

Then $\left(Y, Y_{1}\right)=(Y, Y Z)$, and from (7.11) it is seen that (Y, $Z$ ) is a sufficient statistic for ( $\lambda, v$ ) ; hence it is sufficiently, at
first, to find the distribution of (Y, Z) for one value of the parameter. Until further notice we shall assume, therefore, that $v=0$.

From (7.12) we have that the distribution of $Y$ is a $\chi^{2}$-distribution with $N$ degrees of freedom and scaleparameter $\lambda / N$.

Let $c$ be an arbitrary constant; since $p\left((c X)(c X)^{\prime}\right)=$ $c^{2} p\left(X X^{\prime}\right)=c^{2} \tau(Y, Y Z)=\tau\left(X^{2} Y,\left(c^{2} Y\right) Z\right)$ we conclude that the distribution of $Z$ does not depend on $\lambda$ and since the density of $X$ only depends on $Y(v=0!)$, a simple calculation shows that $Y$ and $Z$ are independent random variables.

Let $s \in \Gamma$ and $a \in \tau(R \times V)$. From the definition of $\Gamma$ it follows that $\eta\left(s^{*}\right)^{-1} \operatorname{an}(s) \in \tau(R \times V)$ and $\eta(s)^{-1}=\eta\left(s^{-1}\right)=\eta\left(s^{*}\right)=$ $\eta(s)^{*}$; thus

$$
\begin{aligned}
& \left\langle p\left((\eta(s) X)(\eta(s) X)^{\prime}, a\right\rangle=\left\langle(\eta(s) X)(\eta(s) X)^{\prime}, a\right\rangle\right. \\
= & \left\langle\eta(s) X X^{\prime} \eta(s)^{-1}, a\right\rangle=\left\langle X X^{\prime}, \eta(s)^{-1} a \eta(s)\right\rangle \\
= & \left\langle p\left(X X^{\prime}\right), \eta(s)^{-1} a \eta(s)\right\rangle=\left\langle\eta(s) p\left(X X^{\prime}\right) \eta(s)^{-1}, a\right\rangle \\
= & \left\langle\eta(s) \tau(Y, Y Z) \eta(s)^{-1}, a\right\rangle=\langle\tau(Y, Y(X(s) Z), a\rangle,
\end{aligned}
$$

and we have

$$
\begin{equation*}
p\left((\eta(s) X)(\eta(s) X)^{\prime}\right)=\tau(Y, Y(X(s) Z)) \tag{7.14}
\end{equation*}
$$

Since the distribution of $X$ is invariant under the orthogonal transformation $\eta(s)$ it follows from (7.6) and (7.14) that the distribution of $Z$ is invariant under orthogonal transformation, of $V$, with determinant 1 .

Define a random variable $R$ by

$$
\begin{equation*}
R=\phi(Z, Z)^{1 / 2} \tag{7.15}
\end{equation*}
$$

From (7.5) and (7.13) it follows that $0 \leq R \leq 1$ and that the maximum likelihood estimator exists if and only if $R<1$.

We shall find the distribution of $R$. Let $u$ be an arbitrary unit vector in $V, \phi(u, u)=1$, and let $Z_{1}=\phi(Z, u)$. If $R=0$ then $Z_{1}=0$ and if $R>0$ then

$$
Z_{1}^{2} / R^{2}=\phi(Z, u)^{2} / R^{2}=\phi(Z / R, u)^{2}
$$

and it is seen that $Z_{1}^{2} / R^{2}$ is the square of the norm of the projection of $Z / R$ on the onedimensional space which contain $u$; thus, since $Z / R$ for given $R, R>0$, is uniformly distributed on the unit sphere, we have that the conditional distribution of $Z_{1}^{2} / R^{2}$ for given $R, R>0$, is a Betadistribution with ( $1, m-1$ ) degrees of freedom.

For $\alpha>0$ we have, therefore, that

$$
\begin{gather*}
E\left(Z_{1}^{2 \alpha}\right)=E\left(E\left(Z_{1}^{2 \alpha} \mid R\right)\right)=E\left(R^{2 \alpha B(1 / 2+\alpha,(m-1) / 2)}\right.  \tag{7.16}\\
B(1 / 2,(m-1) / 2) \\
\\
=E\left(R^{2 \alpha}\right) \frac{B(1 / 2+\alpha,(m-1) / 2)}{B(1 / 2,(m-1) / 2)},
\end{gather*}
$$

and to find the moments of $R$ we only have to find the distribution of $Z_{1}^{2}$.

From (7.12), (7.10), (7.9), and (7.5) it follows that

$$
\begin{aligned}
Z_{1} & =\phi(Z, u)=\phi\left(Y_{1} / Y, u\right) \\
& =\phi\left(Y_{1}, u\right) /\left(\frac{1}{N}\left\langle X X^{\prime}, 1_{E}\right\rangle\right)=\left\langle\tau\left(Y_{1}\right), \tau(u)\right\rangle /\left\langle X X^{\prime}, 1_{E}\right\rangle \\
& =\left\langle\tau\left(Y_{1}, Y_{1}\right), \tau(u)\right\rangle /\left\langle X X^{\prime}, 1_{E}\right\rangle=\left\langle X X^{\prime}, \tau(u)\right\rangle /\left\langle X^{\prime} X, 1_{E}\right\rangle
\end{aligned}
$$

and we have

$$
\begin{equation*}
\left(1-Z_{1}\right) / 2=\left\langle X^{\prime} X,\left(1_{E}-\tau(u)\right) / 2\right\rangle /\left\langle X X^{\prime}, 1_{E}\right\rangle \tag{7.17}
\end{equation*}
$$

Since $u$ is a unit vector it follows from (4.2) that $\tau(u)^{-1}=$ $\tau(u)=\tau(u)^{*} ;$ hence $(1-\tau(u)) / 2$ is an orthogonal projection, and from (7.9) we have that tr $\left(\left(1_{E}-\tau(u)\right) / 2\right)=N / 2$; thus the dimension of the range space is $N / 2$, and it follows from (7.17) that (1- $Z_{1}$ )/2 is Betadistributed with (N/2, N/2) degrees of freedom. By a simple transformation we have that the
distribution of $Z_{1}^{2}$ is a Betadistribution with (1,N/2) degrees of freedom and from (7.16) we obtain

$$
\begin{equation*}
E\left(R^{2 \alpha}\right)=\frac{\Gamma(N / 4+1 / 2) \Gamma(m / 2+\alpha)}{\Gamma(m / 2) \Gamma(N / 4+1 / 2+\alpha)} \tag{7.18}
\end{equation*}
$$

Since $N$ has to be divisible by $\alpha(m)$ it follows from the table in Section 4 that $N / 4+1 / 2=m / 2$ if $m=2,3,5$ or 9 and $N=\alpha(m)$ and that $N / 4+1 / 2>m / 2$ otherwise. In the first case it follows from (7.18) that the distribution of $R$ is degenerated in 1 and in the second case it is easy to see that the distribution of $R^{2}$ is a Betadistribution with (m,N/2-m+1) degrees of freedom.

Still we are assuming that the parameter $v$ is zero, but since the nullsets with respect to a normal distribution are the same for all parameters it follows that the maximum likelihood estimator exists with probability zero if $N / 4+1 / 2=m / 2$ and with probability one if $N / 4+1 / 2>m / 2$. In the following we shall therefore assume that $N / 2>m-1$.

Since $R$ is positive with probability one we can define a random variable $U$ by

$$
U=Z / R
$$

We have that $\phi(U, U)=1$, and, since the distribution of $Z$ is invariant under orthogonal transformation with determinant 1 , it follows that $R$ and $U$ are independent variables and that $U$ is uniformly distributed on the unit sphere.

The maximum likelihood estimator for $(\lambda, v)$ is $\left(Y, Y_{1}\right)=(Y, Y Z)=$ ( $\mathrm{Y}, \mathrm{YRU}$ ) and for $\mathrm{v}=0$ it follows from the results above that the distribution of (Y, R, Z) has density

$$
\begin{equation*}
\frac{N^{N / 2}}{\Gamma(N / 2)(2 \lambda)^{N / 2}} y^{N / 2-1} \exp \left(-\frac{N}{2 \lambda} y\right) \tag{7.20}
\end{equation*}
$$

$$
\times \frac{2}{B(m / 2, N / 4-m / 2+1 / 2)} r^{m-1}\left(1-r^{2}\right)^{N / 4-m / 2-1 / 2}
$$

From (7.11) we have that the density for the distribution of $X$ is

$$
(2 \pi)^{-N / 2}\left(\lambda^{2}-\phi(v, v)\right)^{-N / 4} \exp \left(-\frac{N}{2\left(\lambda^{2}-\phi(v, v)\right)} y(\lambda-r \phi(v, u))\right)
$$

and, since this density for $v=0$ is

$$
(2 \pi)^{-N / 2} \lambda^{-N / 2} \exp \left(-\frac{N}{2 \lambda} y\right),
$$

it follows from this and (7.20) that the distribution of $(Y, R, U)$ for an arbitrary parameter $(\lambda, V) \in \theta$ has density
$\frac{N^{N / 2}}{\Gamma(N / 2) 2^{N / 2}\left(\lambda^{2}-\phi(v, v)\right)^{N / 4}} y^{N / 2-1} \exp \left(-\frac{N}{2\left(\lambda^{2}-\phi(v, v)\right)} y(\lambda-r \phi(v, u))\right)$
$\times \frac{2}{B(m / 2, N / 4-m / 2+1 / 2)} r^{m-1}\left(1-r^{2}\right)^{N / 4-m / 2-1 / 2}$
with respect to the product of the Lebesgue measure on $[0, \infty[$ ( $y$ ), the Lebesgue measure on $[0,1](r)$, and the uniform distribution on the unit sphere (u).

From the structure theorem in Section 6 it is seen that the models of type (ii) are classified between the classical onedimentional models and the classical multidimensional models. Thus (7.21) also determine a distribution on a convex cone (the distribution of $\left(Y, Y_{1}\right)=(Y, Y R U)$ on $\theta$ ) which generalizes the $\chi^{2}$-distribution but which is simpler than the Wishart distribution. We shall see in Section 9 that (7.21) for special values of $N$ and $m$ describe the twodimensional real, complex, and quaternion Wishart distributions.

## 8. LIKELIHOOD RATIO TEST IN A MODEL OF TYPE (ii)

The natural hypotheses in model of type (ii) of course are those which again determine a model of type (ii). It follows from (3.6) that these hypotheses precisely are those which have the form
$\mathrm{H}:(\lambda, \mathrm{v}) \in \mathrm{R} \times \mathrm{V}_{1}$,
where $V_{1}$ is a $k$-dimensional, $1<k<m$, subspace of $V$.
From (7.9) and (7.10) it follows that the orthogonal projection, with respect to the trace form, of $p\left(X X^{\prime}\right)=\tau\left(Y, Y_{1}\right)$ on $\tau\left(R \times V_{1}\right)$ corresponds to the orthogonal projection, with respect to $\phi$, of $\left(Y, Y_{1}\right)=(Y, Y R U)$ on $R \times V_{1}$. Thus we have that the maximum likelihood estimator for ( $\lambda, v$ ) under the hypothesis $H$ is

$$
\left(Y, q\left(Y_{1}\right)\right)=(Y, Y R q(U))
$$

where $q$ is the orthogonal projection of $V$ onto $V_{1}$.
If we substitute (Y,YRU) for ( $\lambda, \mathrm{v}$ ) in (7.11) we get

$$
Y^{-N / 2}\left(1-R^{2}\right)^{-N / 4} e^{-N / 2}
$$

and if we substitute (Y,YRq(U)) we get

$$
Y^{-N / 2}\left(1-R^{2} \phi(q U, q U)\right)^{-N / 4} e^{-N / 2}
$$

and we have that the likelihood ratio statistic is

$$
Q=\left(\frac{1-R^{2}}{1-R^{2} \phi(q U, q U)}\right)^{N / 4}
$$

Let

$$
\begin{aligned}
& \mathrm{R}_{1}=\mathrm{R}(\phi(\mathrm{qU}, \mathrm{qU}))^{1 / 2}, \\
& \mathrm{U}_{1}=\mathrm{qU} /(\phi(\mathrm{qU}, \mathrm{qU}))^{1 / 2},
\end{aligned}
$$

and

$$
W=\left(1-R^{2}\right) /\left(1-R_{1}^{2}\right)
$$

Then the maximum likelihood estimator is (Y, $\mathrm{YR}_{1} \mathrm{U}_{1}$ ) and $\mathrm{Q}=$ $W^{N / 4}$ 。

We shall find the distribution of (Y, $\left.R_{1}, U_{1}, W\right)$ under the hypothesis H. Since we have again a model of type (ii) it follows from Section 7 that ( $Y, R_{1}, U_{1}$ ) is sufficient for ( $\left.\lambda, v\right)$. We
shall therefore first find the distribution under the assump tion that $v=0$. In this case we have from Section 7 that $Y$, $R$, and $U$ are independent variables, that $Y$ is $X^{2}$-distributed with $N / 2$ degrees of freedom and scaleparameter $\lambda / N$, that $R^{2}$ is Betadistributed with (m,N/2-m+1) degrees of freedom, and that $U$ is uniformly distributed on the unit sphere. Since U is uniformly distributed we have, by a well known result, that $\phi(q U, q U)$ and $U_{1}$ are independently distributed and that $\phi(q U, q U)$ is Betadistributed with (kgm-k) degrees of freedom and $U_{1}$ is uniformly distributed on the unit sphere in $V_{1}$. Since $R_{1}^{2}=R^{2}$. $\phi(q u, q u)$ we find by a simple transformation that $R_{1}^{2}$ and $W$ are independent variables which are Betadistributed with respectively $(k, N / 2-k+1)$ and ( $N / 2-m+1, m-k)$ degrees of freedom.

Now, let $v$ be an arbitrary vector in $V_{1}$. Then we have that $R \phi(v, U)=R \phi(v, q U)=R_{1} \phi\left(v, U_{1}\right)$, and from this and. (7.21) it follows, by an argument which is similar to that which led from (7.20) to (7.21), that $\left(Y, R_{1}, U_{1}\right)$ and $W$ are independently distributed, that the distribution of $\left(Y, R_{1}, U_{1}\right)$ has a density of the form (7.21) with $V$ replaced by $V_{1}$, and that $W$ is Betadistributed with (N/2-m+1,m-k) degrees of freedom.

It is important to notice that the maximum likelihood estimator and the likelihood ratio statistic are independently distributed under the hypothesis $H$, because this result ensures that one can allow oneself to test a sequence of hypotheses successively. Since then succesive test are independent.

If $\operatorname{dim} V_{1}=1$ it is not hard to see that the hypothesis $H$ gives a model of the form

$$
\left(\begin{array}{ll}
a & 0 \\
0 & b
\end{array}\right) \otimes I_{N / 2}
$$

where $\lambda=(a+b) / 2 \in R$ and $v=(a-b) / 2 \in V_{1}=R$, and the results above will still hold.

If $\operatorname{dim} V_{1}=0$ then $W=1-R^{2}$ and the maximum likelihood estimator $Y$ for $\lambda$ is $X^{2}$-distributed with $N / 2$ degrees of freedom, and scaleparameter $\lambda / N$ and independent of $W$ which is Betadistributed with (N/2-m+1,m) degrees of freedom.

## 9. EXAMPLES OF MODELS OF TYPE (ii).

We obtain a concrete representation of model of type (ii) by choosing a fixed $\sigma_{m}$. This is done by choosing a concrete representation of the Clifford algebra $C(V, \phi)$ and, by using the methods in the proof of theorem II. 2.5. in Chevalley (1954), we obtain, for an orthonormal basis in $V$, the following representations of $\sigma_{m}$ for $2 \leq m \leqq 9$ (i.e. $\alpha(m) \leq 16$ ):

$$
\sigma_{5}^{2}\left((a+b) / 2,((a-b) / 2, c, f, g, h)=\sigma_{5}^{1}((a+b) / 2,((a-b) / 2, c, f, g,-h)\right.
$$

$$
\sigma_{6}((a+b) / 2,((a-b) / 2, c, f, g, h, k))=\sigma_{7}((a+b) / 2,((a-b) / 2, c, f, 0, g, h, k))
$$

$$
\sigma_{7}\left((a+b) / 2,((a-b) / 2, c, f, i, g, h, k)=\left\{\begin{array}{rrrr}
A & B & C & D \\
B^{\prime} & A & D & -C \\
-C & -D & A & B \\
-D & C & B^{\prime} & A
\end{array}\right\}\right.
$$

where

$$
\begin{aligned}
& \sigma_{2}((a+b) / 2,((a-b) / 2, c))=\left\{\begin{array}{ll}
a & c \\
c & b
\end{array}\right\}, \\
& \sigma_{3}((a+b) / 2,((a-b) / 2, c, f))=\left\{\begin{array}{rrrr}
a & c & 0 & f \\
c & b & -f & 0 \\
0 & -f & a & c \\
f & 0 & c & b
\end{array}\right\}, \\
& \sigma_{4}((a+b) / 2,((a-b) / 2, c, f, g))=\sigma_{5}^{1}((a+b) 2,((a-b) / 2, c, f, g, 0)), \\
& \sigma_{5}^{1}((a+b) / 2,((a-b) / 2, c, f, g, h) \\
& =\left\{\begin{array}{rrrrrrrr}
a & c & 0 & f & 0 & g & 0 & h \\
c & b & -f & 0 & -g & 0 & -h & 0 \\
0 & -f & a & c & 0 & h & 0 & -g \\
f & 0 & c & b & -h & 0 & g & 0 \\
0 & -g & 0 & -h & a & c & 0 & f \\
g & 0 & h & 0 & c & b & -f & 0 \\
0 & -h & 0 & g & 0 & -f & a & c \\
h & 0 & -g & 0 & f & 0 & c & b
\end{array}\right\},
\end{aligned}
$$

$$
\begin{gathered}
A=\left\{\begin{array}{llll}
a & c & 0 & 0 \\
c & b & 0 & 0 \\
0 & 0 & a & c \\
0 & 0 & c & b
\end{array}\right\}, B=\left\{\begin{array}{rrrr}
0 & f & 0 & i \\
-f & 0 & -i & 0 \\
0 & -i & 0 & f \\
i & 0 & -f & 0
\end{array}\right\}, C=\left\{\begin{array}{rrrr}
0 & g & 0 & 0 \\
-g & 0 & 0 & 0 \\
0 & 0 & 0 & g \\
0 & 0 & -g & 0
\end{array}\right\}, \text { and } \\
D=\left\{\begin{array}{rrrr}
0 & -h & 0 & -k \\
h & 0 & k & 0 \\
0 & -k & 0 & h \\
k & 0 & -h & 0
\end{array}\right\}, \\
\sigma_{8}((a+b) / 2,((a-b) / 2, c, f, i, g, j, h, k)) \\
= \\
\sigma_{9}^{1}((a+b) / 2,((a-b) / 2, c, 0, f, i, j, h, k)), \\
\sigma_{9}^{1}\left((a+b) / 2,\left((a-b) / 2, c, d^{\prime}, f, i, g, j, h, k\right)\right) \\
\\
=\left\{\begin{array}{rlrl}
A_{1} & A_{2} & C & D \\
C^{\prime} & D & A_{2} & -B \\
-D & C^{\prime} & -B^{\prime} & A_{1}
\end{array}\right\}
\end{gathered}
$$

where

$$
\begin{aligned}
& A_{1}=\left\{\begin{array}{rrrr}
a & c & 0 & d \\
c & b & -d & 0 \\
0 & -d & a & c \\
d & 0 & c & b
\end{array}\right\}, A_{2}=\left\{\begin{array}{rrrr}
a & c & 0 & -d \\
c & b & d & 0 \\
0 & d & a & c \\
-d & 0 & c & b
\end{array}\right\}, B=\left\{\begin{array}{rrrr}
0 & -f & 0 & i \\
f & 0 & -i & 0 \\
0 & -i & 0 & -f \\
i & 0 & f & 0
\end{array}\right\} \\
& C=\left\{\begin{array}{rrrr}
0 & -g & 0 & -j \\
g & 0 & j & 0 \\
0 & j & 0 & -g \\
-j & 0 & g & 0
\end{array}\right\}, \text { and } D=\left\{\begin{array}{rrrr}
0 & h & 0 & k \\
-h & 0 & -k & 0 \\
0 & k & 0 & -h \\
-k & 0 & h & 0
\end{array}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
& \sigma_{9}^{2}((a+b) / 2,((a-b) / 2, c, d, f, i, g, j, h, k))= \\
& \left.\sigma_{9}^{1}((a+b) / 2,(a-b) / 2, c,-d, f ; i, g, j, h, k)\right)
\end{aligned}
$$

It is now seen that (6.3a) for $m=2$, (6.3a) for $m=3$ and (6.3b) for $m=5$ and $n_{1}=0$ or $n_{2}=0$ are the models of the form (6.4) for $r=2$ (two) and $D$ equal to respectively $R, C$, and H; therefore, these models can be described by invarians.

Apart from the three cases mentioned above $I$ have not succeeded
in finding an interpretation of modelof type (i). Here it shall, however, be emphasized that an arbitrary transformation of the covariance matrices above gives equivalent representations of the models of type (ii); thus, if we collect the observations with variance a in one component and the observations with variance $b$ in another component then it can be seen that the marginal distributions of both components give models of type (i) and that the conditional distribution of one component given the other gives an ordinary linear model with N/2 observation and with m - 1 parameters for the mean vector; these considerations make it possible to understand why the maximum likelihood estimator does not exists in the case where N/2 = m - 1 and why the test statistic $W$ is Betadistributed with (N/2-m+1,m-k) degrees of freedom.

It shall also be mentioned here that the Clifford algebra is of great importance in the mathematical theory of quantum mechanics in connection with the so-called spinrepresentation and that the covariance matrices above closely correspond to Dirac's Y matrices, see Varadarajan (1970), Chapter XII.4. The first time Jordan algebras were treated was in an attempt to formulate the foundation of quantum mechanics. Finally $I$ shall say that the relation (7.14) after my opinion is very important, since this relation shows that the orthogonal transformation $\Phi(s)$ of the observation $X$ corresponds to the orthogonal transformation $X(s)$ of the parameter $v$.

## 10. COVARIANCE MODELS AS EXPONENTIAL FAMILIES.

A linear covariance model can be given by saying that the unknown covariance matrix is a linear combination

$$
\begin{equation*}
\Sigma=\sum_{j=0}^{m} \sigma_{j} G_{j} \tag{10.1}
\end{equation*}
$$

of $m+1$ known, linear independent, symmetric matrices; here $\sigma_{0}, \ldots, \sigma_{m}$ are the parameters, see Anderson (1969).

It follows from Section 2 that it is not a restriction to assume that $G_{0}=I$. Then we have from (2.3) that the model is linear in the inverse covariance too if and only if $G_{i}{ }_{j}+$ $G_{j} G_{i}$ for any $i$ and $j$ is a linear combination of $G_{0}, \ldots, G_{m}$. In this case it follows, in the same way as we obtained (7.4), that the likelihood equations are

$$
\begin{equation*}
\sum_{j=0}^{m} \operatorname{tr}\left(G_{i} G_{j}\right) \sigma_{j}=\operatorname{tr}\left(G_{i} X X^{\prime}\right)=X^{\prime} G_{i} X, \tag{10.2}
\end{equation*}
$$

$i=0, \ldots, m ;$ see also Anderson (1969), Section 5.

Since $G_{0} ; . . ., G_{m}$ are linearly independent (tr $G_{i}{ }_{j}$ ) is a positive definite $(m+1) \times(m+1)$ matrix and (10.2) has a unique solution; if (10.1) is positive definite for this solution it is the maximum likelihood estimator; otherwise the maximum likelihood estimator does not exist.

Since $G_{0}=I$ we can always assume that $\operatorname{tr} G_{i}=0$ for $i=1, \ldots, m$. Then it follows from (7.10) that the model is a model of type (ii) if and only if

$$
\begin{equation*}
G_{i} G_{j}+G_{j} G_{i}=2 \operatorname{tr}\left(G_{i} G_{j}\right) I \tag{10.3}
\end{equation*}
$$

for $i, j=1, \ldots, m$. In this case we can, since the models of type (ii) are treated in an invariant way in Sections 7 and 8, at once give a complete solution to the problems of inference: $V=R^{m}, \lambda=\sigma_{0}, V=\left(\sigma_{1}, \ldots, \sigma_{m}\right),\left(1 / N \operatorname{tr}\left(G_{i} G_{j}\right)\right)$ is the matrix for the inner product $\phi$ on $V=R^{m}$, and the distribution of the maximum likelihood estimator for $\left(\sigma_{0},\left(\sigma_{1}, \ldots, \sigma_{m}\right)\right.$ ) is given by (7.21); a test for a linear hypothesis about ( $\sigma_{1}, \ldots, \sigma_{m}$ ) can be carried out on the basis of the Betadistributed statistic $W$ given in Section 8.

If one compares the considerations in the first part of this section with the structure theorem in Section 5 , which shows which highly structured models one in reality has to deal with, it seems quite clear that one cannot hope to solve the distribution problems for these models by means of the general theory of exponential families. The fact is that each class of nice models requires a special discussion.

## ACKNOWLEDGMENT

I have seen a short note of $T . W$. Anderson and lemma 1 is essentially due to him. My thanks are due to S申ren Johansen for critical reading and discussion.

## REFERENCES

[1] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.
[2] Anderson, T.W. (1969). Statistical inference for covariances with linear structure, in Proceedings of the Second International Symposium on Multivariate Analysis, ed. P.R. Krishnaiah, Academic Press, New York, 55-66.
[3] Anderson, T.W. (1973). Asymptotical efficient estimation of covariance matrices with linear structure. Ann.Statist.1135-141.
[4] Anderson T.W. (1974). Maximum likelihood estimation of parameters of an autoregressive process with moving everage residuals and other covariance matrices with linear structure. Technical Report No. 15, Stanford University, Stanford, California.
[5] Andersson, S. (1972). Invariant Hypotheses I and II. Preprints No. 6 and 7, Institute of Mathematical Statistics, University of Copenhagen, Copenhagen. (Will appear in Ann. Statist.).
[6] Barndorff-Nielsen, 0. (1970). Exponential Families. Exact Theory. Various Publications Series No. 19, Matematisk Institut, Aarhus Universitet.
[7] Bourbaki, N. (1958). E1éments de mathématique, Algèbre, Chapitre 8. Hermann, Paris.
[8] Bourbaki, N. (1959). Eléments de mathématique, Algèbre, Chapitre 9. Hermann, Peris.
[9] Braun, H. and Koecher, M. (1966). Jordan-algebren. Springer-Verlag, Berlin.
[10] Chevalley, C.C. (1954). The Algebraic Theory of Spinors. Columbia University Press, New York.
[11] Goodman, N.R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution Ann.Math.Statist. 34, 152-176.
[12] Jacobson, N. (1968). Structure and Representation of Jordan Algebras. American Mathematical Society, Providence, Rhode Island.
[13] Khatri, C.G. (1965). Classical statistical analysis based on a certain multivariate complex Gaussian distribution. Ann.Math.Statist. 36, 98114.
[14] Jordan, P., von Neumann, J., and Wigner, E., (1934). On an algebraic generalization of the quantum mechanical formalism. Ann. of Math. 36, 2964.
[15] Seely, J. (1971). Quadratic subspaces and completeness. Ann.Math.Statist. 42, 710-721.
[16] Seely, J. (1972). Completeness for a family of multivariate normal distributions. Ann.Math.Statist. 43, 1644-47.
[17] Varadarajan, V.S. (1970). Geometry of Quantum Theory, Volume II. Van Nostrand Reinhold Company, New York.

