The Annals of Statistics
2000, Vol. 28, No. 4, 1150–1178

INVARIANT NORMAL MODELS WITH RECURSIVE GRAPHICAL MARKOV STRUCTURE

BY JESPER MADSEN

University of Copenhagen

An extension of the class of GS-LCI normal models introduced by Andersson and Madsen is defined and studied. The models are defined in terms of symmetry restrictions given by a finite group and conditional independence restrictions given by an acyclic directed graph. Maximum likelihood estimation of the parameters in the models is discussed.

1. Introduction. Andersson and Madsen (1998) [referred to hereafter as AM (1998)] introduced a class of normal models combining group symmetry (GS) restrictions and conditional independence (CI) restrictions, the so-called *GS-LCI models*. In the present paper we define and study a larger class of normal models with GS and CI restrictions, the *GS-ADG models*, extending the GS-LCI models in the following two ways:

1. The CI restrictions are given by an acyclic directed graph (ADG) instead of a finite distributive lattice as in the case of the GS-LCI models. This extension is straightforward since the class of normal models with CI restrictions given by ADGs in a natural way extends the normal models with CI restrictions given by finite distributive lattices [see Andersson and Perlman (1998), Remark 4.1].
2. The condition on the interplay between the GS and the CI restrictions is relaxed such that GS restrictions *between* variables that appear in the CI restrictions are allowed. In the case of the GS-LCI models, GS restrictions are only allowed to operate *within* each of the multivariate variables that appear in the CI restrictions [see AM (1998), Section 2.4], and clearly, this condition is restrictive. A simple example of a GS-ADG model which is *not* a GS-LCI model due to this condition is where we consider a trivariate normal distributed variable (x_a, x_b, x_c) with the CI restriction that x_b and x_c are conditionally independent given x_a, and the GS restriction that (x_a, x_b, x_c) has the same distribution as (x_a, x_c, x_b). A possible application of this model could be where x_a, x_b, x_c are measurements of some variable on objects a, b, c, where b and c are symmetric ($\equiv$ interchangable in the sense that labeling of the two is arbitrary) and connected to each other only through a. The ML estimation properties of this example were discussed in detail in AM (1998), Example 6.1. A generalization to the multivariate case is presented in Examples 6.1 and 7.1 of this paper.

Received February 1999; revised April 2000.
AMS 1991 *subject classifications*. Primary 62H12, 62H15; secondary 62H10, 62H20, 62A05.
Key words and phrases. Group symmetry, invariance, orthogonal group representation, acyclic directed graph, conditional independence, graphical Markov structure, maximum likelihood estimator, multivariate normal distribution.

Hylleberg, Jensen and Ørnbøl (1993) combine GS restrictions given by a subgroup of permutations with CI restrictions given by an undirected graph. Andersen, Højbjerre, Sørensen and Eriksen (1995) combine the special symmetry given by the complex numbers; that is, the GS condition given by the group $\{\pm 1, \pm i\}$, with the CI restrictions given by an undirected graph. In both cases the models become special cases of the GS-ADG models when the graph considered is decomposable; however, these are not special cases of the GS-LCI models. When the graph is not decomposable, results on ML estimation similar to those presented in this paper (Theorem 7.1, Proposition 7.2) cannot be obtained in general.

The GS-ADG models seem to have statistical properties very similar to the GS-LCI models. In this paper we consider the problem of maximum likelihood (ML) estimation.

The general definition of the GS-ADG models is given in Section 6, and some fundamental properties of the models are listed. In particular we obtain an interpretation of the models in terms of certain symmetry restrictions on the conditional covariances and regression coefficients (Theorem 6.1).

In Section 7 a generalization of Theorem 3.1 in AM (1998) is presented. This means that we give a necessary and sufficient condition for the existence and uniqueness of the ML estimator for a fixed observation $x \in \mathbb{R}^I$ together with an almost explicit expression for the ML estimator (Theorem 7.1). Moreover, a generalization of Proposition 3.2 in AM (1998) is obtained; that is, it is shown that either the ML estimator exists and is unique with probability 1, or else, it does not exist or it is not unique (Proposition 7.2).

The basic notation used in this paper is similar to that of AM (1998) and is explained in Section 2. In Section 3, which is an extract of Section 2 in Andersson and Perlman (1998) [hereafter AP (1998)], the notation and concepts of acyclic directed graphs are described. A short introduction to normal models with GS restrictions (GS models) is given in Section 4. This section is a copy of Section 2.2 of AM (1998). In Section 3 we give an introduction to normal models with CI restrictions given by ADGs (ADG models). This section is based on definitions and results from Sections 4, 7 and 10 of AP (1998).

2. Notation. Let I be a finite index set and let $\mathbb{R}^I$ be the vector space of all families $x = (x_i \mid i \in I)$ of real numbers indexed by I. We define $\mathbb{R}^\varnothing = \{0\}$. For $K \subseteq I$, denote by x_K the canonical projection of x on $\mathbb{R}^K$; that is, $x_K = (x_i \mid i \in K) \in \mathbb{R}^K$.

Let J be (another) finite index set, and let $\mathrm{M}(I \times J) \equiv \mathbb{R}^{I\times J}$ denote the vector space of all $I \times J$ matrices. The algebra $\mathrm{M}(I \times I)$ is denoted by $\mathrm{M}(I)$. For $A \in \mathrm{M}(I \times J)$ let $A' \in \mathrm{M}(J \times I)$ denote the transposed matrix. The group of all nonsingular $I \times I$ matrices, the group of all orthogonal $I \times I$ matrices, the cone of all positive semidefinite $I \times I$ matrices, and the cone of all positive definite $I \times I$ matrices are denoted by $\mathrm{GL}(I)$, $\mathrm{O}(I)$, $\mathrm{PS}(I)$ and $\mathrm{P}(I)$, respectively. The $I \times I$ identity matrix is denoted by 1_I.

For $A = (a_{ii'} \mid (i, i') \in I \times I) \in \mathrm{M}(I)$ and $L, K \subseteq I$, let A_{LK} denote the $L \times K$ submatrix of A; that is, $A_{LK} = (a_{ii'} \mid (i, i') \in L \times K) \in \mathrm{M}(L \times K)$.

Let A'_{LK} denote $(A_{LK})'$. Note that $A'_{LK} = (A')_{KL}$. The submatrix A_{KK} is denoted by A_K. If A_K is nonsingular, then A_K^{-1} denotes the inverse matrix $(A_K)^{-1}$.

For a covariance matrix $\Sigma \in \mathrm{P}(I)$ and disjoint subsets L, K of I, we define the corresponding conditional covariance $\Sigma_{L\cdot K} = \Sigma_L - \Sigma_{LK}\Sigma_K^{-1}\Sigma_{KL} \in \mathrm{P}(L)$.

For $\xi \in \mathbb{R}^I$ and $\Sigma \in \mathrm{P}(I)$ let $N(\xi, \Sigma)$ denote the normal distribution on $\mathbb{R}^I$ with expectation ξ and covariance matrix Σ. Let $N(\Sigma)$ denote $N(0, \Sigma)$.

3. Acyclic digraphs (ADGs). A directed graph (digraph) D is a pair (V, E) where V is a finite set (the set of *vertices*) and $E \subseteq V \times V$ is a binary relation (the set of *directed edges*) such that

$$\forall u, v \in V\colon (u, v) \in E \Rightarrow (v, u) \notin E.$$

For $u, v \in V$ we write $u \to v$ whenever $(u, v) \in E$, and we write $u < v$ if $u \to v$ or if there exists $v_1, \dots, v_k \in V$, where $k \in \mathbb{N}$, such that $u \to v_1 \to \cdots \to v_k \to v$. The hereby defined relation $<$ on V is transitive.

An acyclic directed graph (ADG) is a directed graph $D = (V, E)$ with the property that $v \not< v$ for all $v \in V$. In this case the relation $\leq$ on V given by $u \leq v$ if and only if $u < v$ or $u = v$, $u, v \in V$, is a *partial ordering* on V; that is, it is reflexive, antisymmetric and transitive.

For an ADG $D = (V, E)$ and $v \in V$ we define $\mathrm{pa}(v) = \{u \in V \mid u \to v\}$ (the *parents* of v), $\mathrm{de}(v) = \{u \in V \mid v < u\}$ (the *descendants* of v) and $\mathrm{nd}(v) = \{u \in V \mid v \not\leq u\}$ (the *nondescendants* of v).

Let $D_1 = (V_1, E_1)$ and $D_2 = (V_2, E_2)$ be two ADGs. A mapping $f\colon V_1 \to V_2$ is called an *ADG homomorphism* if $u_1 \to v_1 \Rightarrow f(u_1) \to f(v_1)$ for all $u_1, v_1 \in V_1$. Usually we write $f\colon D_1 \to D_2$ for such an ADG homomorphism. For an ADG $D = (V, E)$, the set of all bijective ADG homomorphisms $D \to D$ is denoted by $\mathrm{Aut}(D)$.

4. The group symmetry model. Let I be a finite index set, G a finite group and $\rho\colon G \to \mathrm{O}(I)$ an *orthogonal group representation* of G on $\mathbb{R}^I$; that is, $\rho(1) = 1_I$ and $\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$ for all $g_1, g_2 \in G$. Let $\mathrm{M}_G(I)$ denote the subalgebra of all matrices $A \in \mathrm{M}(I)$ that commute with $\rho(G)$; that is, $A\rho(g) = \rho(g)A$ for all $g \in G$. The group of all nonsingular matrices, the cone of all positive semidefinite matrices and the cone of all positive definite matrices in $\mathrm{M}_G(I)$ are denoted by $\mathrm{GL}_G(I)$, $\mathrm{PS}_G(I)$, and $\mathrm{P}_G(I)$, respectively. Note that $\Sigma \in \mathrm{PS}_G(I)$ if and only if $\Sigma \in \mathrm{PS}(I)$ and Σ is G-invariant; that is, $\rho(g)\Sigma\rho(g)' = \Sigma$. The statistical model

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_G(I)) \tag{4.1}$$

with observation space $\mathbb{R}^I$ and parameter space $\mathrm{P}_G(I)$ is called the *group symmetry (GS) model given by G*. An algebraic theory containing a complete solution to the likelihood inference problem for these models was developed by Andersson, Brøns and Jensen in the years 1972–1985. A summary of the basic theory together with a complete list of references is given in AM (1998), Appendix A.

The smoothing ($\equiv$ averaging) mapping

$$\psi_I^G\colon \mathrm{PS}(I) \to \mathrm{PS}_G(I)$$
$$S \longmapsto \frac{1}{|G|}\sum(\rho(g)S_\rho(g)' \mid g \in G), \tag{4.2}$$

is fundamental for likelihood inference for group symmetry models. In fact, $\psi_I^G(xx')$ becomes, under a certain regularity condition, the unique maximum likelihood estimator of $\Sigma \in \mathrm{P}_G(I)$ based on the random observation $x \in \mathbb{R}^I$. When, I, G or both are subsumed we denote ψ_I^G by ψ^G, ψ_I and ψ respectively.

5. Normal ADG models. Let $D \equiv (V, E)$ denote an ADG (cf. Section 3) and let I_v, $v \in V$, denote nonempty finite index sets. We define the (entire) index set to be

$$I = \dot\cup(I_v \mid v \in V), \tag{5.1}$$

and for any subset A of V we define $I_A = \dot\cup(I_v \mid v \in A)$. Furthermore, for $v \in V$, define the following three subsets of I:

$$[v] = I_v, \qquad \langle v\rangle = I_{\mathrm{pa}(v)}, \qquad (v) = I_{\mathrm{nd}(v)\setminus\mathrm{pa}(v)}$$

(cf. Section 3). A covariance matrix $\Sigma \in \mathrm{P}(I)$ is said to have *conditional independence (CI) restrictions wrt D* if and only if for every $v \in V$, $x_{[v]}$ and $x_{(v)}$ are conditional independent given $x_{\langle v\rangle}$ [in short: $x_{[v]} \perp x_{(v)} \mid x_{\langle v\rangle}$] whenever $x \in \mathbb{R}^I$ follows $N(\Sigma)$. The set of all such covariance matrices is denoted by $\mathrm{P}_D(I)$. The statistical model

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_D(I)) \tag{5.2}$$

with observation space $\mathbb{R}^I$ and parameter space $\mathrm{P}_D(I)$ is called the *normal ADG model determined by D*; compare AP (1998), Section 6.

Let $A \in \mathrm{M}(I)$. For $u, v \in V$, let $A_{[u,v]}$ and $A_{\langle u,v\rangle}$ denote $A_{[u][v]}$ and $A_{\langle u\rangle\langle v\rangle}$, respectively. For $v \in V$, $A_{[v]}$ and $A_{\langle v\rangle}$ then denotes $A_{[v,v]}$ and $A_{\langle v,v\rangle}$, respectively (cf. Section 2). For $u, v \in V$ let $A_{[u,v\rangle}$ and $A_{\langle u,v]}$ denote $A_{[u]\langle v\rangle}$ and $A_{\langle u\rangle[v]}$, respectively, and similarly for convenience let $A_{[v\rangle}$ and $A_{\langle v]}$ denote $A_{[v]\langle v\rangle}$ and $A_{\langle v\rangle[v]}$, respectively. For $v \in V$, we thus have

$$A_{[v]\dot\cup\langle v\rangle} = \begin{pmatrix} A_{[v]} & A_{[v\rangle} \\ A_{\langle v]} & A_{\langle v\rangle}\end{pmatrix},$$

where $A_{[v]} \in \mathrm{M}([v])$, $A_{[v\rangle} \in \mathrm{M}([v]\times\langle v\rangle)$, $A_{\langle v]} \in \mathrm{M}(\langle v\rangle\times[v])$, and $A_{\langle v\rangle} \in \mathrm{M}(\langle v\rangle)$. Furthermore, define the vector space $\mathrm{M}_D(I) \subseteq \mathrm{M}(I)$ by $A \in \mathrm{M}_D(I)$ if and only if $A_{[u,v]} = 0$ for all $u, v \in V$ where $v \neq u$ and $v \notin \mathrm{pa}(u)$ [cf. AP (1998), Section 10].

For a covariance matrix $\Sigma \in \mathrm{P}(I)$ and $v \in V$ we use the short notation $\Sigma_{[v]\cdot}$ for the corresponding conditional covariance $\Sigma_{[v]\cdot\langle v\rangle} \in \mathrm{P}([v])$ (cf. Section 2). The family of matrices

$$\left(\left(\Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}, \Sigma_{[v]\cdot}\right) \mid v \in V\right) \in \times(\mathrm{M}([v]\times\langle v\rangle)\times \mathrm{P}([v]) \mid v \in V)$$

is called the family of *D-parameters* of $\Sigma \in \mathrm{P}(I)$ [cf. AP (1998), Definition 4.1].

In the solution to the likelihood inference problems for the model (5.2), the family of D-parameters plays an important role since both the likelihood function (LF) and the parameter space (PS) factorizes into products of LFs and PSs for simple MANOVA models indexed by V. The factor corresponding to $v \in V$ is equivalent to the MANOVA model on the sample space $\mathbb{R}^{[v]}$ given by the mean value subspace

$$\{R_{[v\rangle} x_{\langle v\rangle} \mid R_{[v\rangle} \in \mathrm{M}([v] \times \langle v\rangle)\},$$

where $x_{\langle v\rangle} \in \mathbb{R}^{\langle v\rangle}$, and the covariance matrix $\Lambda_{[v]} = \Sigma_{[v]\cdot} \in \mathrm{P}([v])$. [For a brief review of the MANOVA model, see AP (1998), Section 6.] This fact follows from the following three fundamental results for normal ADG models:

1. The mapping

$$\begin{aligned}\mathrm{P}_D(I) &\to \times(\mathrm{M}([v] \times \langle v\rangle) \times \mathrm{P}([v]) \mid v \in V)\\ \Sigma &\mapsto ((\Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}, \Sigma_{[v]\cdot}) \mid v \in V),\end{aligned}$$

is bijective [AP (1998), Proposition 4.1].

2. $\Sigma \in \mathrm{P}_D(I)$ if and only if

$$\mathrm{tr}(\Sigma^{-1}xx') = \sum(\mathrm{tr}(\Sigma_{[v]\cdot}^{-1}(x_{[v]} - \Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}x_{\langle v\rangle})(\cdots)') \mid v \in V), \tag{5.3}$$

for all $x \in \mathbb{R}^I$ [AP (1998), Proposition 4.2].

3. For $\Sigma \in \mathrm{P}_D(I)$,

$$\det(\Sigma) = \prod(\det(\Sigma_{[v]\cdot}) \mid v \in V) \tag{5.4}$$

[AP (1998), Proposition 4.2].

We notice, that (2) could equivalently be stated as

4. $\Sigma \in \mathrm{P}_D(I)$ if and only if

$$\begin{aligned}\mathrm{tr}(\Sigma^{-1}S) = \sum(\mathrm{tr}(\Sigma_{[v]\cdot}^{-1}(&S_{[v]} - S_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}\Sigma_{\langle v]}\\ &- \Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}S_{\langle v]} + \Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}S_{\langle v\rangle}\Sigma_{\langle v\rangle}^{-1}\Sigma_{\langle v]})) \mid v \in V),\end{aligned} \tag{5.5}$$

for all $S \in \mathrm{PS}(I)$.

Another important property of the normal ADG models which we shall use is the following:

5. The mapping

$$\begin{aligned}\mathrm{M}_D(I) &\to \times(\mathrm{M}([v] \times \langle v\rangle) \times \mathrm{M}([v]) \mid v \in V)\\ A &\mapsto ((A_{[v\rangle}, A_{[v]}) \mid v \in V)\end{aligned}$$

is bijective [AP (1998), (10.4)].

6. Normal GS-ADG models. Let I be a finite set, $D = (V, E)$ an ADG, and let $I = \dot\cup(I_v \mid v \in V)$ be a partitioning of I in nonempty finite index sets I_v, $v \in V$ (cf. Section 5).

Let G be a finite group and ρ: $G \to \mathrm{O}(I)$ an orthogonal group representation of G on $\mathbb{R}^I$. We denote by $P_{G,D}(I)$ the intersection $\mathrm{P}_G(I) \cap \mathrm{P}_D(I)$; that is, $\mathrm{P}_{G,D}(I)$ is the set of covariance matrices with both GS restrictions given by G and CI restrictions given by D.

The corresponding normal statistical model with observation space $\mathbb{R}^I$ is thus

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_{G,D}(I)). \tag{6.1}$$

Now let

$$\begin{aligned} &G \to \mathrm{Aut}(D), \\ &g \mapsto (v \mapsto gv), \end{aligned} \tag{6.2}$$

be a *representation of* G *on* D; that is, $(g_1g_2)v = g_1(g_2v)$ and $e_Gv = v$ for all $g_1, g_2 \in G$ and $v \in V$, where e_G denotes the one-element in G.

We have thus introduced two different representations of the group G, one on $\mathbb{R}^I$ and one on D. This is done in order to specify the restriction on the interplay between the GS restrictions and the CI restrictions used in the definition of the GS-ADG model (cf. Lemma 6.1 and Definition 6.1 below).

LEMMA 6.1. *For* $g \in G$ *the following conditions are equivalent*:

(i) $\forall\, x \in \mathbb{R}^I\ \forall\, v \in V$: $x_{[v]} = 0 \Rightarrow (\rho(g)x)_{[gv]} = 0$;
(ii) $\forall\, x \in \mathbb{R}^I\ \forall\, v \in V$: $(\rho(g)x)_{[gv]} = \rho(g)_{[gv,v]}x_{[v]}$;
(iii) $\forall\, u, v \in V$: $u \neq gv \Rightarrow \rho(g)_{[u,v]} = 0$.

PROOF. (ii) $\Rightarrow$ (i) is trivial. To show (iii) $\Rightarrow$ (ii), let $x \in \mathbb{R}^I$ and $v \in V$. Then

$$(\rho(g)x)_{[gv]} = \sum(\rho(g)_{[gv,u]}x_{[u]} \mid u \in V) = \rho(g)_{[gv,v]}x_{[v]}.$$

To show (i) $\Rightarrow$ (iii), consider $u, v \in V$ with $u \neq gv$ and choose $x \in \mathbb{R}^I$ such that $x_{[t]} = 0$ for $t \in V$ where $t \neq v$,

$$\rho(g)_{[u,v]}x_{[v]} = \sum(\rho(g)_{[u,t]}x_{[t]} \mid t \in V) = (\rho(g)x)_{[u]} = (\rho(g)x)_{[g(g^{-1}u)]} = 0,$$

where the last equation follows by assumption since $g^{-1}u \neq v$ and hence $x_{[g^{-1}u]} = 0$. The assertion now follows since the choice of $x_{[v]} \in \mathbb{R}^{[v]}$ was arbitrary. □

DEFINITION 6.1. Under the assumption that the three equivalent conditions of Lemma 6.1 hold for all $g \in G$, the model (6.1) is called the normal GS-ADG model determined by G and D.

For the remainder of the section we shall assume that the three equivalent conditions of Lemma 6.1 hold.

REMARK 6.1. It follows from (iii) of Lemma 6.1, that

$$\rho(g)_{[gv,\, v\rangle} = \rho(g)_{\langle gv,\, v]} = 0, \tag{6.3}$$

for all $g \in G$ and $v \in V$. From the fact that the CI-conditions given by a finite lattice (ring) of subsets of the index set I equivalently can be given by a *transitive* ADG [see Andersson, Madigan, Perlman and Triggs (1995a, b)], it then follows that the GS-LCI models [cf. AM (1998), Section 2.4] correspond to the special case of the GS-ADG models where D is transitive and (6.2) is trivial; that is, $gv = v$ for all $g \in G$ and $v \in V$. (The latter condition implies that the GS-restrictions only operate inside each of the multivariate nodes.)

REMARK 6.2. Another special case is where the GS-restrictions are given only by the permutations (6.2) of the nodes in D. This means that in (5.1) we assume $I_v = I_{gv}$ for all $g \in G$ and $v \in V$, in which case (6.2) induces a representation ρ of G on $\mathbb{R}^I$ simply by permutation of the (multivariate) variables according to the permutation of the nodes in the graph; that is,

$$\rho(g)((x_v \mid v \in V)) = (x_{g^{-1}(v)} \mid v \in V),$$

for $g \in G$ and $x = (x_v \mid v \in V) \in \mathbb{R}^I \equiv \times(\mathbb{R}^{I_v} \mid v \in v)$.

The formulas (6.4)–(6.17) below are fundamental and used several times in calculations further on in the paper.

Let $g \in G$, $u, v \in V$, and $A \in \mathrm{M}(I)$. We have by (ii) of Lemma 6.1 that

$$(\rho(g)A)_{[gu,\, v]} = \rho(g)_{[gu,\, u]} A_{[u,\, v]}, \tag{6.4}$$

and hence also

$$(A\rho(g)')_{[v,\, gu]} = A_{[v,\, u]} \rho(g)'_{[gu,\, u]}. \tag{6.5}$$

From the fact that $\rho(g^{-1}) = \rho(g)^{-1} = \rho(g)'$, we get from (6.4) replacing g by g^{-1} and u by gv, that

$$1_{[v]} = \rho(g)'_{[gv,\, v]} \rho(g)_{[gv,\, v]}. \tag{6.6}$$

Since $\langle u \rangle = \dot\cup([u'] \mid u' \in \mathrm{pa}(u))$ and $\langle v \rangle = \dot\cup([v'] \mid v' \in \mathrm{pa}(v))$, we similarly obtain the identities

$$(\rho(g)x)_{\langle gv \rangle} = \rho(g)_{\langle gv,\, v \rangle} x_{\langle v \rangle}, \tag{6.7}$$

$$(\rho(g)A)_{\langle gu,\, v \rangle} = \rho(g)_{\langle gu,\, u \rangle} A_{\langle u,\, v \rangle}, \tag{6.8}$$

$$(A\rho(g)')_{\langle v,\, gu \rangle} = A_{\langle v,\, u \rangle} \rho(g)'_{\langle gu,\, u \rangle}, \tag{6.9}$$

$$1_{\langle v \rangle} = \rho(g)'_{\langle gv,\, v \rangle} \rho(g)_{\langle gv,\, v \rangle}, \tag{6.10}$$

$$(\rho(g)A)_{[gu,\, v\rangle} = \rho(g)_{[gu,\, u\rangle} A_{\langle u,\, v \rangle} \tag{6.11}$$

and

$$(A\rho(g)')_{[v,\, gu\rangle} = A_{[v,\, u\rangle} \rho(g)'_{\langle gu,\, u \rangle}, \tag{6.12}$$

respectively.

Let $\Sigma \in \mathrm{P}(I)$, $v \in V$ and $g \in G$. From (6.8) and (6.9), it follows that

$$(\rho(g)\Sigma\rho(g)')_{\langle gv\rangle} = \rho(g)_{\langle gv,\, v\rangle}\Sigma_{\langle v\rangle}\rho(g)'_{\langle gv,\, v\rangle},$$

and then by (6.10),

$$(\rho(g)\Sigma\rho(g)')^{-1}_{\langle gv\rangle} = \rho(g)_{\langle gv,\, v\rangle}\Sigma^{-1}_{\langle v\rangle}\rho(g)'_{\langle gv,\, v\rangle}. \tag{6.13}$$

From (6.11) and (6.12) we get that

$$(\rho(g)\Sigma\rho(g)')_{[gv\rangle} = \rho(g)_{[gv,\, v]}\Sigma_{[v\rangle}\rho(g)'_{\langle gv,\, v\rangle}, \tag{6.14}$$

and then by (6.13) and (6.10),

$$(\rho(g)\Sigma\rho(g)')_{[gv\rangle}(\rho(g)\Sigma\rho(g)')^{-1}_{\langle gv\rangle} = \rho(g)_{[gv,\, v]}\Sigma_{[v\rangle}\Sigma^{-1}_{\langle v\rangle}\rho(g)'_{\langle gv,\, v\rangle}. \tag{6.15}$$

In a similar way we obtain the equations

$$(\rho(g)\Sigma\rho(g)')_{[gv].} = \rho(g)_{[gv,\, v]}\Sigma_{[v].}\rho(g)'_{[gv,\, v]}, \tag{6.16}$$

$$(\rho(g)\Sigma\rho(g)')^{-1}_{[gv].} = \rho(g)_{[gv,\, v]}\Sigma^{-1}_{[v].}\rho(g)'_{[gv,\, v]}. \tag{6.17}$$

We shall now give a characterization of the GS-ADG models in terms of certain invariance restrictions on the D-parameters (cf. Section 5). The characterization is given in Theorem 6.1 and uses Proposition 6.1 and Definition 6.2 below.

PROPOSITION 6.1. *If* $\Sigma \in \mathrm{P}_D(I)$ *and* $g \in G$, *then* $\rho(g)\Sigma\rho(g)' \in \mathrm{P}_D(I)$.

PROOF. Let $\Sigma \in \mathrm{P}_D(I)$, $g \in G$ and $x \in \mathbb{R}^I$. We show that (5.3) holds for $\rho(g)\Sigma\rho(g)'$. Thus

$$\begin{aligned}
&\mathrm{tr}((\rho(g)\Sigma\rho(g)')^{-1}xx')\\
&\quad = \mathrm{tr}(\Sigma^{-1}(\rho(g^{-1})x)(\rho(g^{-1})x)')\\
&\quad = \sum(\mathrm{tr}(\Sigma^{-1}_{[v].}((\rho(g^{-1})x)_{[v]} - \Sigma_{[v\rangle}\Sigma^{-1}_{\langle v\rangle}(\rho(g^{-1})x)_{\langle v\rangle})(\cdots)') \mid v \in V)\\
&\quad = \sum(\mathrm{tr}(\Sigma^{-1}_{[v].}(\rho(g^{-1})_{[v,\, gv]}x_{[gv]} - \Sigma_{[v\rangle}\Sigma^{-1}_{\langle v\rangle}\rho(g^{-1})_{\langle v,\, gv\rangle}x_{\langle gv\rangle})(\cdots)') \mid v \in V)\\
&\quad = \sum(\mathrm{tr}((\rho(g)\Sigma\rho(g)')^{-1}_{[gv].}(x_{[gv]}\\
&\qquad\qquad - (\rho(g)\Sigma\rho(g)')_{[gv\rangle}(\rho(g)\Sigma\rho(g)')^{-1}_{\langle gv\rangle}x_{\langle gv\rangle})(\cdots)') \mid v \in V)\\
&\quad = \sum(\mathrm{tr}((\rho(g)\Sigma\rho(g)')^{-1}_{[v].}(x_{[v]}\\
&\qquad\qquad - (\rho(g)\Sigma\rho(g)')_{[v\rangle}(\rho(g)\Sigma\rho(g)')^{-1}_{\langle v\rangle}x_{\langle v\rangle})(\cdots)') \mid v \in V),
\end{aligned}$$

where the second equality is just (5.3), the third follows from (ii) of Lemma 6.1 and (6.7), the fourth follows from (6.15), (6.17), (6.6) and the fact that $\rho(g^{-1}) = \rho(g)^{-1} = \rho(g)'$, and the fifth since the mapping $V \to V (v \mapsto gv)$ is one-to-one. □

DEFINITION 6.2. (i) The family

$$(R_{[v\rangle} \mid v \in V) \in \times(\mathrm{M}([v] \times \langle v\rangle) \mid v \in V)$$

is called *invariant wrt* G if and only if

$$R_{[gv\rangle} = \rho(g)_{[gv,\, v]} R_{[v\rangle} \rho(g)'_{\langle gv,\, v\rangle},$$

for all $v \in V$ and $g \in G$.

(ii) The family

$$(\Lambda_{[v]} \mid v \in V) \in \times(\mathrm{M}([v]) \mid v \in V)$$

is called *invariant wrt* G if and only if

$$\Lambda_{[gv]} = \rho(g)_{[gv,\, v]} \Lambda_{[v]} \rho(g)'_{[gv,\, v]},$$

for all $v \in V$ and $g \in G$.

(iii) If both $(R_{[v\rangle} \mid v \in V) \in \times(\mathrm{M}([v] \times \langle v\rangle) \mid v \in V)$ and $(\Lambda_{[v]} \mid v \in V) \in \times(\mathrm{M}([v]) \mid v \in V)$ are invariant wrt G, then the family $((R_{[v\rangle}, \Lambda_{[v]}) \mid v \in V)$ is called *invariant wrt* G.

THEOREM 6.1. *The mapping* $\Sigma \mapsto ((\Sigma_{[v\rangle}\Sigma^{-1}_{\langle v\rangle}, \Sigma_{[v]\cdot}) \mid v \in V)$ [*cf.* (1) *of Section* 5] *constitutes a one-to-one correspondence between* $\mathrm{P}_{G,D}(I)$ *and families of D-parameters which are invariant wrt G*.

PROOF. Let $\Sigma \in \mathrm{P}_D(I)$ and $g \in G$. By (1) of Section 5, Proposition 6.1, and the fact that the mapping $V \to V(v \mapsto gv)$ is one-to-one, it follows that $\Sigma = \rho(g)\Sigma\rho(g)'$ if and only if

$$\Sigma_{[gv\rangle}\Sigma^{-1}_{\langle gv\rangle} = (\rho(g)\Sigma\rho(g)')_{[gv\rangle}(\rho(g)\Sigma\rho(g)')^{-1}_{\langle gv\rangle}$$

and

$$\Sigma_{[gv]\cdot} = (\rho(g)\Sigma\rho(g)')_{[gv]\cdot},$$

for all $v \in V$. The theorem now follows since the right-hand side of these equations are given by (6.15) and (6.16), respectively. □

Theorem 6.1 states that the GS-ADG model can be interpreted in terms of invariance restrictions on the D-parameters. In words, these restrictions imply that vertices which are equivalent under the action (6.2) have D-parameters which are one-to-one functions of each other. Hence, the D-parameters are *not* variation independent in general, which is the case for the GS-LCI models [cf. AM, Theorem 2.1]. Furthermore, Theorem 6.1 implies that there may be certain invariance restrictions on each of the marginal D-parameters. In order to characterize these restrictions further, we need the following definition. (The characterization is important and used to state and prove our main result on maximum likelihood estimation, Theorem 7.1.)

DEFINITION 6.3. For $v \in V$, let G_v denote the isotropy group corresponding to the action (6.2), that is,

$$G_v = \{g \in G \mid gv = v\}.$$

Furthermore, denote by $\mathrm{M}_{G_v}([v] \times \langle v \rangle)$ the vector space of all $[v] \times \langle v \rangle$ matrices $R_{[v\rangle}$ satisfying that $\rho(g)_{[v]} R_{[v\rangle} \rho(g)'_{\langle v \rangle} = R_{[v\rangle}$ for all $g \in G_v$, and by $\mathrm{P}_{G_v}([v])$ the cone of all $[v] \times [v]$ covariance matrices $\Lambda_{[v]}$ satisfying that $\rho(g)_{[v]} \Lambda_{[v]} \rho(g)'_{[v]} = \Lambda_{[v]}$ for all $g \in G_v$.

For $v \in V$ and $\Sigma \in \mathrm{P}_{G,D}(I)$, it is now easily verified that the invariance restrictions of the marginal D-parameters $R_{[v\rangle} = \Sigma_{[v\rangle} \Sigma^{-1}_{\langle v \rangle}$ and $\Lambda_{[v]} = \Sigma_{[v]}$. (cf. Theorem 6.1) are given as $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v \rangle)$ and $\Lambda_{[v]} \in \mathrm{P}_{G_v}([v])$.

REMARK 6.3. In the case where (6.2) is trivial (cf. Remark 6.1), we have $G_v = G$ for all $v \in V$, and hence all invariance restrictions on the D-parameters (cf. Theorem 6.1) are given as $R_{[v\rangle} \in \mathrm{M}_G([v] \times \langle v \rangle)$, and $\Lambda_{[v]} \in \mathrm{P}_G([v])$, $v \in V$, that is, the restrictions are given only in terms of additional restrictions on each of the marginal D-parameters [compare AM (1998), Theorem 2.1].

REMARK 6.4. The opposite (dual) case of the one in Remark 6.3 is where the representation of G_v on $\mathbb{R}^{[v]}$ ($\equiv$ subrepresentation of ρ) is trivial for all $v \in V$. In this case, the invariance restrictions on the D-parameters (cf. Theorem 6.1) are given as

$$R_{[u\rangle} = \rho(g)_{[u,v]} R_{[v\rangle} \rho(g)'_{\langle u,v \rangle}, \qquad \Lambda_{[u]} = \rho(g)_{[u,v]} \Lambda_{[v]} \rho(g)'_{[u,v]},$$

for all $u, v \in V$ and $g \in G$ where $u \neq v$ and $u = gv$; that is, the restrictions are all given in terms of one-to-one relations *between* some of the D-parameters. In particular for $v \in V$, the marginal D-parameters $R_{[v\rangle}$ and $\Lambda_{[v]}$ are unrestricted.

Many, but not all, of the special cases mentioned in Remark 6.2 are of this type (cf. Examples 6.1–6.4 below).

A series of seven simple examples will illustrate the result obtained in Theorem 6.1. In Examples 6.1–6.4 the restrictions are given as in Remark 6.2; that is, the GS-restrictions are induced by permutations of the vertices in D. Additionally, Examples 6.1–6.3 are of the type where the representation of G_v on $\mathbb{R}^{[v]}$ is trivial for all $v \in V$ (cf. Remark 6.4). Example 6.5 is of the type where $G_v = G$ for all $v \in V$ (cf. Remark 6.3). In this example D is not transitive, hence this is not a GS-LCI model (cf. Remark 6.1). Examples 6.6 and 6.7 are not of any of the special types mentioned in Remarks 6.1–6.4. About the notation: for $u, v \in V$ we denote by (u, v) the transposition of u and v; that is, the simple permutation of V only interchanging u and v. For $u_1, v_1, u_2, v_2 \in V$, $(u_1, v_1)(u_2, v_2)$ denotes the composite of the two transpositions. The identity mapping of V [the one-element in $\mathscr{S}(V)$] is denoted by 1_V.

EXAMPLE 6.1. Let $V = \{1, 2, 3\}$, and let $D = (V, E)$ be the ADG below:

$$\begin{matrix} 2 & & 3 \\ & 1 & \end{matrix}.$$

Furthermore, let J and L be finite sets, and define $I_1 = J$, $I_2 = I_3 = L$ and

$$I = I_1 \dot{\cup} I_2 \dot{\cup} I_3.$$

Then $\mathbb{R}^I$ consists of families (x_1, x_2, x_3) of multivariate observations where $x_1 \in \mathbb{R}^J$ and $x_2, x_3 \in \mathbb{R}^L$.

Let G be the subgroup $\{1_V, (2, 3)\}$ of $\mathscr{S}(V)$. Consider the representations of G on D and $\mathbb{R}^I$, respectively, given by Remark 6.2.

The only nontrivial CI restriction given by D is that x_2 and x_3 are conditionally independent given x_1 (in short: $x_2 \perp x_3 \mid x_1$), and the only nontrivial GS restriction given by G is that the joint distribution of x_1, x_2, x_3 is invariant under the permutation of x_2 and x_3; that is, (x_1, x_2, x_3) has the same distribution as (x_1, x_3, x_2) [in short: $(x_1, x_2, x_3) =^{\mathscr{D}} (x_1, x_3, x_2)$].

From (1) of Section 5 it follows that $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters,

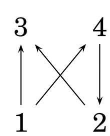
$$\begin{aligned} R_2 &= \Sigma_{[2\rangle} \Sigma_{\langle 2 \rangle}^{-1}, & R_3 &= \Sigma_{[3\rangle} \Sigma_{\langle 3 \rangle}^{-1}, \\ \Lambda_2 &= \Sigma_{[2]\cdot}, & \Lambda_3 &= \Sigma_{[3]\cdot}, \\ \Lambda_1 &= \Sigma_{[1]\cdot} = \Sigma_{[1]}, \end{aligned}$$

where $\langle 2 \rangle = \langle 3 \rangle = I_1$. The hypothesis $\Sigma \in P_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$R_2 = R_3 = R, \qquad \Lambda_2 = \Lambda_3 = \Lambda,$$

that is, $\Sigma \in \mathrm{P}_{G,D}(I)$ is uniquely determined from the parameters Λ_1, Λ, and R, respectively.

EXAMPLE 6.2. Let $V = \{1, 2, 3, 4\}$, and let $D = (V, E)$ be the ADG below:



Furthermore, let J and L be finite sets, and define $I_1 = I_2 = J$, $I_3 = I_4 = L$, and

$$I = I_1 \dot{\cup} I_2 \dot{\cup} I_3 \dot{\cup} I_4.$$

Then $\mathbb{R}^I$ consists of families (x_1, x_2, x_3, x_4) of multivariate observations where $x_1, x_2 \in \mathbb{R}^J$ and $x_3, x_4 \in \mathbb{R}^L$.

Let G be the subgroup $\{1_V, (3, 4)\}$ of $\mathscr{S}(V)$. Consider the representations of G on D and $\mathbb{R}^I$, respectively, given by Remark 6.2.

The CI restrictions given by D are

$$x_1 \perp x_2, \qquad x_3 \perp x_4 \mid (x_1, x_2),$$

and the only nontrivial GS restriction given by G is that $(x_1, x_2, x_3, x_4) =^{\mathscr{D}} (x_1, x_2, x_4, x_3)$.

From (1) of Section 5, it follows that $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters

$$\begin{aligned} R_3 &= \Sigma_{[3\rangle}\Sigma_{\langle 3\rangle}^{-1}, & R_4 &= \Sigma_{[4\rangle}\Sigma_{\langle 4\rangle}^{-1}, \\ \Lambda_3 &= \Sigma_{[3]\cdot}, & \Lambda_4 &= \Sigma_{[4]\cdot}, \\ \Lambda_1 &= \Sigma_{[1]}, & \Lambda_2 &= \Sigma_{[2]}, \end{aligned} \tag{6.18}$$

and we have $\langle 3\rangle = \langle 4\rangle = I_1 \cup I_2$. The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$R_3 = R_4 = R, \qquad \Lambda_3 = \Lambda_4 = \Lambda,$$

that is, $\Sigma \in \mathrm{P}_{G,D}(I)$ is uniquely determined from the parameters Λ_1, Λ_2, R and Λ, respectively.

EXAMPLE 6.3. Let D and I be as in Example 6.2. Instead consider the subgroup $G = \{1_V, (1,2)(3,4)\}$ of $\mathscr{S}(V)$, analogously with the representations of G on D and $\mathbb{R}^I$, respectively, given by Remark 6.2. The only nontrivial GS restriction given by G is that $(x_1, x_2, x_3, x_4) =^{\mathscr{D}} (x_2, x_1, x_4, x_3)$. As in Example 6.2, $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters (6.18). The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$\begin{aligned} R_3 &= (A, B), \\ R_4 &= (B, A), \\ \Lambda_3 &= \Lambda_4 = \Lambda_{34}, \\ \Lambda_1 &= \Lambda_2 = \Lambda_{12}, \end{aligned}$$

where $A, B \in \mathrm{M}(L \times J)$; that is, $\Sigma \in \mathrm{P}_{G,D}(I)$ is uniquely determined from the parameters A, B, Λ_{34} and Λ_{12}, respectively.

EXAMPLE 6.4. Let D and I be as in Example 6.2. Instead consider the subgroup $G = \{1_V, (1,2), (3,4), (1,2)(3,4)\}$ of $\mathscr{S}(V)$, analogously with the representations of G on D and $\mathbb{R}^I$, respectively, given by Remark 6.2. The GS restrictions given by G are

$$(x_1, x_2, x_3, x_4) \overset{\mathscr{D}}{=} (x_2, x_1, x_3, x_4) \overset{\mathscr{D}}{=} (x_1, x_2, x_4, x_3) \overset{\mathscr{D}}{=} (x_2, x_1, x_4, x_3).$$

As in Example 6.2, $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters (6.18). The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$R_3 = R_4 = (B, B),$$
$$\Lambda_3 = \Lambda_4 = \Lambda_{34},$$
$$\Lambda_1 = \Lambda_2 = \Lambda_{12},$$

where $B \in \mathrm{M}(L \times J)$; that is, $\Sigma \in \mathrm{P}_{G,D}(I)$ is uniquely determined from the parameters B, Λ_{34}, and Λ_{12}, respectively.

EXAMPLE 6.5. Let $V = \{1, 2, 3, 4\}$, and let $D = (V, E)$ be the ADG below:

$$\begin{array}{ccccc} & & 4 & & \\ & & \uparrow & & \\ 1 & \longrightarrow & 3 & \longleftarrow & 2\ . \end{array}$$

Furthermore, let J_l, $l = 1, \ldots, 4$, be finite sets; let $J_{lk} = J_l$, $l = 1, \ldots, 4$, $k = 1, 2$, let

(6.19) $$I_l = J_{l1} \dot{\cup} J_{l2},$$

$l = 1, \ldots, 4$, and define

$$I = I_1 \dot{\cup} I_2 \dot{\cup} I_3 \dot{\cup} I_4.$$

Then $\mathbb{R}^I$ consists of families $(x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32}, x_{41}, x_{42})$ of multivariate observations where $x_{l1}, x_{l2} \in \mathbb{R}^{J_l}$, $l = 1, \ldots, 4$.

Let σ denote the permutation of $\mathbb{R}^I$ where x_{l1} and x_{l2} are permuted simultaneously for $l = 1, \ldots, 4$. We then consider $G = \{1_I, \sigma\}$ with the trivial ($\equiv$ vacuous) representation on D.

The CI restrictions given by D are

$$(x_{11}, x_{12}) \perp (x_{21}, x_{22}),$$
$$(x_{11}, x_{12}, x_{21}, x_{22}) \perp (x_{41}, x_{42}) \mid (x_{31}, x_{32}),$$

and the only nontrivial GS restriction given by G is

$$(x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32}, x_{41}, x_{42})$$
$$\overset{\mathcal{D}}{=} (x_{12}, x_{11}, x_{22}, x_{21}, x_{32}, x_{31}, x_{42}, x_{41}).$$

From (1) of Section 5, if follows that $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters

$$R_3 = \Sigma_{[3\rangle}\Sigma^{-1}_{\langle 3\rangle}, \qquad R_4 = \Sigma_{[4\rangle}\Sigma^{-1}_{\langle 4\rangle},$$
$$\Lambda_3 = \Sigma_{[3]\cdot}, \qquad \Lambda_4 = \Sigma_{[4]\cdot},$$
$$\Lambda_1 = \Sigma_{[1]}, \qquad \Lambda_2 = \Sigma_{[2]},$$

where $\langle 3 \rangle = I_1 \cup I_2$ and $\langle 4 \rangle = I_3$. The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$
\begin{aligned}
&R_3 = \begin{pmatrix} A_{31} & B_{31} & A_{32} & B_{32} \\ B_{31} & A_{31} & B_{32} & A_{32} \end{pmatrix}, && R_4 = \begin{pmatrix} A_{43} & B_{43} \\ B_{43} & A_{43} \end{pmatrix},\\
&\Lambda_3 = \begin{pmatrix} A_3 & B_3 \\ B_3 & A_3 \end{pmatrix}, && \Lambda_4 = \begin{pmatrix} A_4 & B_4 \\ B_4 & A_4 \end{pmatrix},\\
&\Lambda_1 = \begin{pmatrix} A_1 & B_1 \\ B_1 & A_1 \end{pmatrix}, && \Lambda_2 = \begin{pmatrix} A_2 & B_2 \\ B_2 & A_2 \end{pmatrix},
\end{aligned} \tag{6.20}
$$

where $A_{31}, B_{31} \in \mathrm{M}(J_3 \times J_1)$, $A_{32}, B_{32} \in \mathrm{M}(J_3 \times J_2)$, $A_{43}, B_{43} \in \mathrm{M}(J_4 \times J_3)$ and A_l, B_l are $J_l \times J_l$ matrices, $l = 1, \dots, 4$; that is, $\Sigma \in \mathrm{P}_{G,D}(I)$ is uniquely determined from these parameters.

EXAMPLE 6.6. Let $V = \{2, 3, 4\}$, and let $D = (V, E)$ be the ADG below:

$$
\begin{array}{ccc} 3 & & 4 \\ & \nwarrow \ \nearrow & \\ & 2 & \end{array} .
$$

Furthermore, let J and L be finite sets and define

$$
I_2 = J_1 \dot{\cup} J_2, \tag{6.21}
$$

where $J_1 = J_2 = J$; let $I_3 = I_4 = L$, and let

$$
I = I_2 \dot{\cup} I_3 \dot{\cup} I_4.
$$

Then $\mathbb{R}^I$ consists of families (x_1, x_2, x_3, x_4) of multivariate observations where $x_1, x_2 \in \mathbb{R}^J$ and $x_3, x_4 \in \mathbb{R}^L$.

Let σ_{12} denote the permutation of $\mathbb{R}^I$ given by the permutation of x_1 and x_2, and similarly, let σ_{34} denote the permutation of x_3 and x_4. We then consider $G = \{1_I, \sigma_{12} \circ \sigma_{34}\}$ with the representation on D given such that $\sigma_{12} \circ \sigma_{34}$ corresponds to $(1, 2)\,(3, 4)$.

The only nontrivial CI restriction given by D is that

$$
x_3 \perp x_4 \mid (x_1, x_2),
$$

and the only nontrivial GS restriction given by G is that $(x_1, x_2, x_3, x_4) \stackrel{\mathscr{D}}{=} (x_2, x_1, x_4, x_3)$. From (1) of Section 5, it follows that $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters

$$
\begin{aligned}
&R_3 = \Sigma_{[3\rangle} \Sigma_{\langle 3 \rangle}^{-1}, && R_4 = \Sigma_{[4\rangle} \Sigma_{\langle 4 \rangle}^{-1},\\
&\Lambda_3 = \Sigma_{[3]\cdot}, && \Lambda_4 = \Sigma_{[4]\cdot},\\
&\Lambda_2 = \Sigma_{[2]},
\end{aligned} \tag{6.22}
$$

where $\langle 3 \rangle = \langle 4 \rangle = I_2$. The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$R_3 = (A, B),$$
$$R_4 = (B, A),$$
$$\Lambda_3 = \Lambda_4 = \Lambda,$$
$$\Lambda_2 = \begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix},$$

where $A, B \in \mathrm{M}(L \times J)$, and Φ, Δ are $J \times J$ matrices.

EXAMPLE 6.7. Let D and I be as in Example 6.6. Instead consider the larger group $G = \{1_I, \sigma_{12}, \sigma_{34}, \sigma_{12} \circ \sigma_{34}\}$, with the representation on D such that σ_{12} corresponds to $(1, 2)$, σ_{34} to $(3, 4)$, and (hence) $\sigma_{12} \circ \sigma_{34}$ to $(1, 2)\,(3, 4)$. The GS restrictions given by G are that

$$(x_1, x_2, x_3, x_4) \overset{\mathscr{D}}{=} (x_2, x_1, x_3, x_4) \overset{\mathscr{D}}{=} (x_1, x_2, x_4, x_3) \overset{\mathscr{D}}{=} (x_2, x_1, x_4, x_3).$$

As in Example 6.6, $\Sigma \in \mathrm{P}_D(I)$ is uniquely determined by the D-parameters (6.22). The hypothesis $\Sigma \in \mathrm{P}_{G,D}(I)$ could then by Theorem 6.1 be expressed in terms of the additional restrictions

$$R_3 = R_4 = (A, A),$$
$$\Lambda_3 = \Lambda_4 = \Lambda,$$
$$\Lambda_2 = \begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix},$$

where $A \in \mathrm{M}(L \times J)$, and Φ, Δ are $J \times J$ matrices.

We close the section with a characterization of the intersection $\mathrm{M}_{G,D}(I)$ of the two algebras $\mathrm{M}_G(I)$ and $\mathrm{M}_D(I)$ (cf. Sections 4 and 5, respectively). The characterization is used in the proof of our main result on ML estimation (Theorem 7.1).

PROPOSITION 6.2. *If* $A \in \mathrm{M}_D(I)$ *and* $g \in G$, *then* $\rho(g) A \rho(g)' \in \mathrm{M}_D(I)$.

PROOF. Consider $u, v \in V$ such that $u \neq v$ and $v \notin \mathrm{pa}(u)$. Since the mapping $V \to V (v \mapsto g^{-1}v)$ is a bijective ADG homomorphism, it follows that $g^{-1}u \neq g^{-1}v$ and $g^{-1}v \notin \mathrm{pa}(g^{-1}u)$, and hence by assumption, $A_{[g^{-1}u,\, g^{-1}v]} = 0$. Then by (6.4) and (6.5),

$$(\rho(g) A \rho(g)')_{[uv]} = \rho(g)_{[u,\, g^{-1}u]} A_{[g^{-1}u,\, g^{-1}v]} \rho(g)'_{[v,\, g^{-1}v]} = 0. \qquad \square$$

THEOREM 6.2. *The mapping* $A \mapsto ((A_{[v\rangle}, A_{[v]}) \mid v \in V)$ *constitutes a one-to-one correspondence between* $\mathrm{M}_{G,D}(I)$ *and families of matrices which are invariant wrt*. G [*cf. Definition* 6.2, (iii)].

PROOF. Let $g \in G$. By (5) of Section 5, Proposition 6.2, and the fact that the mapping $V \to V(v \mapsto gv)$ is one-to-one, it follows that $A = \rho(g)A\rho(g)'$ if and only if

$$A_{[gv\rangle} = (\rho(g)A\rho(g)')_{[gv\rangle}$$

and

$$A_{[gv]} = (\rho(g)A\rho(g)')_{[gv]},$$

for all $v \in V$. The theorem now follows since the right-hand side of these equations can be calculated using (6.4), (6.5), (6.11) and (6.12), respectively. □

7. ML estimation in GS-ADG models. In this section we shall consider ML estimation in the GS-ADG model

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_{G,D}(I)) \tag{7.1}$$

(cf. Definition 6.1). We start with some basic and probably well-known results.

Let I_1 and I_2 be finite sets and assume that

$$I = I_1 \dot{\cup} I_2. \tag{7.2}$$

LEMMA 7.1. *Let $S \in \mathrm{PS}(I)$ and let*

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

be the partitioning of S according to (7.2).

The equation

$$S_{21} = R_{21}S_{11}, \tag{7.3}$$

has a solution $\widehat{R}_{21}$ for $R_{21} \in \mathrm{M}(I_2 \times I_1)$. Furthermore, the matrix $S_{22} - \widehat{R}_{21}S_{12}$ does not depend on the choice of solution $\widehat{R}_{21}$ to (7.3).

PROOF. Choose $y \in \mathrm{M}(I)$ such that $S = yy'$, and let

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

be the partitioning of y according to (7.2); that is, $y_1 \in \mathrm{M}(I_1 \times I)$ and $y_2 \in \mathrm{M}(I_2 \times I)$, respectively. Let $P_2 \in \mathrm{M}((I_2 \times I) \times (I_2 \times I))$ denote the orthogonal projection matrix onto the subspace

$$L_2(y_1) = \{R_{21}y_1 \mid R_{21} \in \mathrm{M}(I_2 \times I_1)\}$$

of $\mathbb{R}^{I_2 \times I}$ (wrt the usual inner product on $\mathbb{R}^{I_2 \times I}$). We can choose $\widehat{R}_{21} \in \mathrm{M}(I_2 \times I_1)$ such that

$$\widehat{R}_{21}y_1 = P_2 y_2. \tag{7.4}$$

Then for every $R_{21} \in \mathrm{M}(I_2 \times I_1)$,

$$0 = \mathrm{tr}((R_{21}y_1)'(y_2 - \widehat{R}_{21}y_1)) = \mathrm{tr}(R_{21}'(S_{21} - \widehat{R}_{21}S_{11})), \tag{7.5}$$

which implies that $S_{21} - \widehat{R}_{21}S_{11} = 0$; that is, R_{21} is a solution to (7.3). On the other hand, if $\widehat{R}_{21}$ is a solution to (7.3), then by the last equation in (7.5),

$$\mathrm{tr}((R_{21}y_1)'(y_2 - \widehat{R}_{21}y_1)) = 0,$$

for all $R_{21} \in \mathrm{M}(I_2 \times I_1)$, and hence $\widehat{R}_{21}y_1 = P_2y_2$. In this case

$$\begin{aligned}(y_2 - P_2y_2)(y_2 - P_2y_2)' &= (y_2 - \widehat{R}_{21}y_1)(y_2 - \widehat{R}_{21}y_1)' \\ &= S_{22} - \widehat{R}_{21}S_{12};\end{aligned}$$

that is, $S_{22} - \widehat{R}_{21}S_{12}$ does not depend on the solution $\widehat{R}_{21}$ to (7.3). □

It follows from Lemma 7.1 that for $S \in \mathrm{PS}(I)$ with the partitioning according to (7.2) as above, the matrix $S_{22} - \widehat{R}_{21}S_{12}$, where $\widehat{R}_{21}$ is any solution to (7.3) is well defined; that is, it exists and is independent of the choice of solution $\widehat{R}_{21}$. This matrix is denoted by $S_{2\circ 1}$. Note that in the case where S_{11} is nonsingular, $S_{2\circ 1} = S_{2\cdot 1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$.

PROPOSITION 7.1. *Let $S \in \mathrm{PS}(I)$ and let*

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

be the partitioning of S according to (7.2). *Furthermore, let $\alpha > 0$ and consider the function*

$$\begin{aligned}\mathbf{L}(R_{21}, \Lambda_2) = \det(\Lambda_2)^{-\alpha/2} \exp\bigl(-\tfrac{1}{2}\mathrm{tr}(\Lambda_2^{-1}(S_{22} - S_{21}R_{21}' \\ - R_{21}S_{12} + R_{21}S_{11}R_{21}'))\bigr),\end{aligned}$$

where $(R_{21}, \Lambda_2) \in \mathrm{M}(I_2 \times I_1) \times \mathrm{P}(I_2)$.

Then $\mathbf{L}$ has a maximum if and only if $S_{2\circ 1}$ is nonsingular. In this case all maximas are given as pairs $(\widehat{R}_{21}, \widehat{\Lambda}_2)$ where $\widehat{R}_{21}$ is a solution to the equation

$$S_{21} = R_{21}S_{11}, \tag{7.6}$$

and $\widehat{\Lambda}_2 = (1/\alpha)S_{2\circ 1}$.

A maximum for $\mathbf{L}$ is unique if and only if the equation

$$R_{21}S_{11} = 0 \tag{7.7}$$

only has the solution $R_{21} = 0$ for $R_{21} \in \mathrm{M}(I_2 \times I_1)$.

PROOF. Let y, P_2 and $L_2(y_1)$ be as in the proof of Lemma 7.1. Then for arbitrary $(R_{21}, \Lambda_2) \in \mathrm{M}(I_2 \times I_1) \times \mathrm{P}(I_2)$,

$$\mathbf{L}(R_{21}, \Lambda_2) = \det(\Lambda_2)^{-\alpha/2} \exp\bigl(-\tfrac{1}{2}\mathrm{tr}\bigl(\Lambda_2^{-1}(y_2 - R_{21}y_1)(y_2 - R_{21}y_1)'\bigr)\bigr),$$

and it follows that for fixed value of $\Lambda_2 \in \mathrm{P}(I_2)$, $\mathbf{L}(R_{21}, \Lambda_2)$ is maximized for any solution $\widehat{R}_{21}$ to (7.4), or equivalently, to (7.6). In this case

$$\mathbf{L}(\widehat{R}_{21}, \Lambda_2) = \det(\Lambda_2)^{-\alpha/2} \exp\left(-\tfrac{1}{2}\mathrm{tr}(\Lambda_2^{-1} S_{2\circ 1})\right);$$

that is, $\mathbf{L}(\widehat{R}_{21}, \Lambda_2)$ does not depend on the choice of solution $\widehat{R}_{21}$ to (7.6). Moreover, $\mathbf{L}(\widehat{R}_{21}, \Lambda_2)$ has a maximum for $\Lambda_2 \in \mathrm{P}(I_2)$ if and only if $S_{2\circ 1}$ is nonsingular [cf. Andersson (1984), Lemma 3.2.2]. In this case the maximum is unique and given as $\widehat{\Lambda}_2 = (1/\alpha) S_{2\circ 1}$. A maximum $(\widehat{R}_{21}, \widehat{\Lambda}_2)$ for $\mathbf{L}$ is thus unique if and only if (7.6) has a unique solution, or equivalently, if and only if (7.7) only has the null-solution. □

COROLLARY 7.1. *There exists a unique maximum for* $\mathbf{L}$ *if and only if* S *is nonsingular*.

PROOF. It follows that (7.7) only has the null-solution if and only if S_{11} is nonsingular. In this case $S_{2\circ 1} = S_{2\cdot 1} = S_{22} - S_{21} S_{11}^{-1} S_{12}$, and S is nonsingular if and only if S_{11} and $S_{2\cdot 1}$ are nonsingular [cf. Rao (1973), Problem 1.2.4]. □

For the model (7.1) we now consider the problem of existence and uniqueness of the ML estimator based on an observation $x \in \mathbb{R}^I$.

THEOREM 7.1. *In the model* (7.1), *the maximum likelihood estimator* $\widehat{\Sigma} = \widehat{\Sigma}(x)$ *of* $\Sigma \in \mathrm{P}_{G,D}(I)$ *for the observation* $x \in \mathbb{R}^I$ *exists if and only if the matrices* $\psi(xx')_{[v]\circ}$, $v \in V$, *all are positive definite*.

In this case, *there is a one-to-one correspondence between all families of solutions* $(\widehat{R}_{[v\rangle} \mid v \in V)$ *to the equations*

$$\psi(xx')_{[v\rangle} = R_{[v\rangle} \psi(xx')_{\langle v\rangle}, \tag{7.8}$$

where $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$, $v \in V$, *and all ML estimators* $\widehat{\Sigma} = \widehat{\Sigma}(x)$ *given by the equations*

$$\widehat{\Sigma}_{[v\rangle} \widehat{\Sigma}_{\langle v\rangle}^{-1} = \widehat{R}_{[v\rangle}, \qquad \widehat{\Sigma}_{[v]\cdot} = \psi(xx')_{[v]\circ}, \tag{7.9}$$

$v \in V$. *The maximum likelihood estimator* $\widehat{\Sigma} = \widehat{\Sigma}(x)$ *is then unique if and only if the equations*

$$R_{[v\rangle} \psi(xx')_{\langle v\rangle} = 0,$$

where $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$, $v \in V$, *only have the solutions* $R_{[v\rangle} = 0$, $v \in V$.

PROOF. From (A.1) in AM (1998) and (5.5) it follows that the likelihood function $\mathbf{L}\colon \mathrm{P}_{G,D}(I) \to]0,\infty[$ can be calculated as follows:

$$\begin{aligned}\mathbf{L}(\Sigma) &= \det(\Sigma)^{-1/2}\exp(-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1}xx')) \\ &= \det(\Sigma)^{-1/2}\exp(-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1}\psi(xx'))) \\ &= \prod(\det(\Sigma_{[v]\cdot})^{-1/2} \\ &\quad \times \exp(-\tfrac{1}{2}\mathrm{tr}(\Sigma_{[v]\cdot}^{-1}((\psi(xx')_{[v]} - \psi(xx')_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}\Sigma_{\langle v]} \\ &\qquad - \Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1}\psi(xx')_{\langle v]} + \Sigma_{[v\rangle}\Sigma_{\langle v\rangle}^{-1} \\ &\qquad \times \psi(xx')_{\langle v\rangle}\Sigma_{\langle v\rangle}^{-1}\Sigma_{\langle v]}))) \mid v \in V).\end{aligned} \tag{7.10}$$

Thus from Theorem 6.1, it follows that it suffices to consider the problem of maximizing

$$\prod(\mathbf{L}_v(R_{[v\rangle}, \Lambda_{[v]}) \mid v \in V), \tag{7.11}$$

where

$$\begin{aligned}\mathbf{L}_v(R_{[v\rangle}, \Lambda_{[v]}) = \det(\Lambda_{[v]})^{-1/2}\exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_{[v]}^{-1}((\psi(xx')_{[v]} - \psi(xx')_{[v\rangle} \\ \times R'_{[v\rangle} - R_{[v\rangle}\psi(xx')_{\langle v]} + R_{[v\rangle}\psi(xx')_{\langle v\rangle}R'_{[v\rangle}))),\end{aligned} \tag{7.12}$$

for all families

$$((R_{[v\rangle}, \Lambda_{[v]}) \mid v \in V) \in \times(\mathrm{M}([v] \times \langle v\rangle) \times \mathrm{P}([v]) \mid v \in V),$$

which are invariant wrt G.

For each $v \in V$, there exists by Lemma 7.1 $\widetilde{R}_{[v\rangle} \in \mathrm{M}([v] \times \langle v\rangle)$ such that

$$\psi(xx')_{[v\rangle} = \widetilde{R}_{[v\rangle}\psi(xx')_{\langle v\rangle}. \tag{7.13}$$

By (5) of Section 5, there exists a matrix $\widetilde{R} \in \mathrm{M}_D(I)$ such that its $[v] \times \langle v\rangle$ submatrix equals $\widetilde{R}_{[v\rangle}$, $v \in V$. Now define

$$R = \frac{1}{|G|}\sum(\rho(g)\widetilde{R}\rho(g)' \mid g \in G). \tag{7.14}$$

Then for $v \in V$, it follows from (6.11) and (6.12) that

$$\begin{aligned}\widehat{R}_{[v\rangle} &= \frac{1}{|G|}\sum((\rho(g)\widetilde{R}\rho(g)')_{[v\rangle} \mid g \in G) \\ &= \frac{1}{|G|}\sum(\rho(g)_{[v,\, g^{-1}v]}\widetilde{R}_{[g^{-1}v\rangle}\rho(g)'_{\langle v,\, g^{-1}v\rangle} \mid g \in G),\end{aligned} \tag{7.15}$$

and hence

$$
\begin{aligned}
&\widehat{R}_{[v\rangle}\psi(xx')_{\langle v\rangle}\\
&\quad=\frac{1}{|G|}\sum\big(\rho(g)_{[v,\,g^{-1}v]}\widetilde{R}_{[g^{-1}v\rangle}\rho(g)'_{\langle v,\,g^{-1}v\rangle}\psi(xx')_{\langle v\rangle}\mid g\in G\big)\\
&\quad=\frac{1}{|G|}\sum\big(\rho(g)_{[v,\,g^{-1}v]}\widetilde{R}_{[g^{-1}v\rangle}\psi(xx')_{\langle g^{-1}v\rangle}\rho(g)'_{\langle v,\,g^{-1}v\rangle}\mid g\in G\big)\\
&\quad=\frac{1}{|G|}\sum\big(\rho(g)_{[v,\,g^{-1}v]}\psi(xx')_{[g^{-1}v\rangle}\rho(g)'_{\langle v,\,g^{-1}v\rangle}\mid g\in G\big)\\
&\quad=\frac{1}{|G|}\sum\big(\psi(xx')_{[v\rangle}\mid g\in G\big)=\psi(xx')_{[v\rangle},
\end{aligned}
\tag{7.16}
$$

where the second equation follows from (6.8), (6.9) and the fact that $\psi(xx') \in \mathrm{P}_G(I)$, the third follows from (7.13) and the fourth from (6.11) and (6.12) and the fact that $\psi(xx') \in \mathrm{P}_G(I)$. Obviously, $\widehat{R} \in \mathrm{M}_G(I)$, and from Proposition 6.2 and the fact that $\mathrm{M}_D(I)$ is a vector space, it follows that $\widehat{R} \in \mathrm{M}_D(I)$. From Theorem 6.2 it then follows that the family $(\widehat{R}_{[v\rangle} \mid v \in V)$ is invariant wrt G. Since $\psi(xx')_{[v]^\circ} = \psi(xx')_{[v]} - \widehat{R}_{[v\rangle}\psi(xx')_{\langle v]}$, $v \in V$, it then follows from (6.4), (6.5), (6.11), (6.12) and the fact that $\psi(xx') \in \mathrm{P}_G(I)$, that $(\psi(xx')_{[v]^\circ} \mid v \in V)$ is invariant wrt G. Thus, from Proposition 7.1 it follows that if the matrices $\psi(xx')_{[v]^\circ}$, $v \in V$, all are nonsingular, then $((\widehat{R}_{[v\rangle}, \psi(xx')_{[v]^\circ}) \mid v \in V)$ is a maximum for (7.11), which is invariant wrt G.

Conversely, assume that there exists $u \in V$ such that $\psi(xx')_{[u]^\circ}$ is singular. For $n \in \mathbb{N}$ and $v \in V$, let $R^{(n)}_{[v\rangle} = \widehat{R}_{[v\rangle}$, where $\widehat{R}$ is defined as in (7.14), and let $\Lambda^{(n)}_{[v]} = n^{-1}1_{[v]} + \psi(xx')_{[v]^\circ}$. Then $((R^{(n)}_{[v\rangle}, \Lambda^{(n)}_{[v]}) \mid v \in V)$ is invariant wrt G, $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \mathbf{L}_v(R^{(n)}_{[v\rangle}, \Lambda^{(n)}_{[v]}) = \infty$$

and hence

$$\lim_{n\to\infty} \prod\big(\mathbf{L}_v\big(R^{(n)}_{[v\rangle}, \Lambda^{(n)}_{[v]}, \mathbf{L}_v\big(R^{(n)}_{[v\rangle}, \Lambda^{(n)}_{[v]}\big) \mid v \in V\big) = \infty.$$

This hereby proves the "existence" part of the theorem.

For the uniqueness of $\widehat{\Sigma}$ exists; that is, the matrices $\psi(xx')_{[v]^\circ}$, $v \in V$, are all nonsingular. Then first assume that for each $v \in V$, the equation

$$\psi(xx')_{[v\rangle} = R_{[v\rangle}\psi(xx')_{\langle v\rangle} \tag{7.17}$$

has a unique solution $\widetilde{R}_{[v\rangle}$ for $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$. Defining $\widehat{R}$ as in (7.14), then for $v \in V$, $\widehat{R}_{[v\rangle}$ is a solution to (7.17) [cf. (7.16)], and since $\widehat{R}_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$ we must have $\widetilde{R}_{[v\rangle} = \widehat{R}_{[v\rangle}$ by the uniqueness assumption. In particular, $(\widetilde{R}_{[v\rangle} \mid v \in V)$ is invariant wrt G.

Conversely, assume that $(\widetilde{R}_{[v\rangle} \mid v \in V)$ is a family of solutions to the equations (7.17), $v \in V$, which is unique among all families of solutions in $\times(\mathrm{M}([v] \times \langle v\rangle) \mid v \in V)$ that are invariant wrt G. Then for each $v \in V$, $\widetilde{R}_{[v\rangle}$ is a unique solution to (7.17), for $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$. To show this, let $u \in V$, and choose any $R_{[u\rangle} \in \mathrm{M}_{G_u}([u] \times \langle u\rangle)$ such that $\psi(xx')_{[u\rangle} = R_{[u\rangle}\psi(xx')_{\langle u\rangle}$. Then define

$$\widehat{R}_{[gu\rangle} = \rho(g)_{[gu,\,u]} R_{[u\rangle} \rho(g)'_{\langle gu,\,u\rangle}, \tag{7.18}$$

for $g \in G$ and let

$$\widehat{R}_{[v\rangle} = \widetilde{R}_{[v\rangle}, \tag{7.19}$$

for $v \in V$ where $v \neq gu$ for all $g \in G$. First notice that (7.18) is well defined; for $g_1, g_2 \in G$ where $g_1 u = g_2 u$ we have $g_2^{-1} g_1 \in G_u$ and hence by (6.4) and (6.8),

$$\begin{aligned} R_{[u\rangle} &= \rho(g_2^{-1} g_1)_{[u]} R_{[u\rangle} \rho(g_2^{-1} g_1)'_{\langle u\rangle} \\ &= \rho(g_2^{-1})_{[u,\,g_2 u]} \rho(g_1)_{[g_2 u,\,u]} R_{[u\rangle} \rho(g_1)'_{\langle g_2 u,\,u\rangle} \rho(g_2^{-1})'_{\langle u,\,g_2 u\rangle}, \end{aligned}$$

which by assumption and (6.6) implies that

$$\rho(g_2)_{[g_2 u,\,u]} R_{[u\rangle} \rho(g_2)'_{\langle g_2 u,\,u\rangle} = \rho(g_1)_{[g_1 u,\,u]} R_{[u\rangle} \rho(g_1)'_{\langle g_1 u,\,u\rangle}.$$

The family $(\widehat{R}_{[v\rangle} \mid v \in V)$ defined by (7.18) and (7.19) is clearly invariant wrt G. From (6.8), (6.9), (6.11) and (6.12), and the fact that $\psi(xx') \in \mathrm{P}_G(I)$, it follows that

$$\begin{aligned} \widehat{R}_{[gu\rangle}\psi(xx')_{\langle gu\rangle} &= \rho(g)_{[gu,\,u]} R_{[u\rangle} \rho(g)'_{\langle gu,\,u\rangle} \psi(xx')_{\langle gu\rangle} \\ &= \rho(g)_{[gu,\,u]} R_{[u\rangle} \psi(xx')_{\langle u\rangle} \rho(g)'_{\langle gu,\,u\rangle} \\ &= \rho(g)_{[gu,\,u]} \psi(xx')_{[u\rangle} \rho(g)'_{\langle gu,\,u\rangle} = \psi(xx')_{[gu\rangle}, \end{aligned}$$

for all $g \in G$, and hence we must have $\widehat{R}_{[v\rangle} = \widetilde{R}_{[v\rangle}$, $v \in V$. In particular $\widetilde{R}_{[u\rangle} = R_{[u\rangle}$.

From the fact that (7.17) has a unique solution for $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$, if and only if the equation $R_{[v\rangle}\psi(xx')_{\langle v\rangle} = 0$ only has the null-solution for $R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle)$, we have thus proved the last part of the theorem. □

REMARK 7.1. According to Theorem 7.1, there is a one-to-one correspondence between the MLE $\widehat{\Sigma}$ and families $(\widehat{R}_{[v\rangle} \mid v \in V)$ of solutions to (7.8). It should be pointed out that in some cases there could in fact be several families of solutions to (7.8) such that the MLE exists but is not unique.

REMARK 7.2. It follows from Theorem 7.1 and (7.10) that the maximum of the likelihood function for the observation $x \in \mathbb{R}^I$ is

$$\prod(\det(\psi(xx')_{[v]^\circ})^{-1/2} \mid v \in V) \exp\left(-\frac{|I|}{2}\right). \tag{7.20}$$

□

REMARK 7.3. The explicit expression for $\widehat{\Sigma}(x)$ may be obtained from (7.9) by means of the reconstruction algorithm given in AP (1998), Section 5.

COROLLARY 7.2. *In the model* (7.1), *the maximum likelihood estimator* $\widehat{\Sigma} = \widehat{\Sigma}(x)$ *of* $\Sigma \in \mathrm{P}_{G,D}(I)$ *for the observation* $x \in \mathbb{R}^I$ *exists and is unique if the matrices* $\psi(xx')_{[v]\dot{\cup}\langle v\rangle}$, $v \in V$, *all are positive definite.*

In this case, $\widehat{\Sigma}$ *is determined by*

$$\widehat{\Sigma}_{[v\rangle}\widehat{\Sigma}_{\langle v\rangle}^{-1} = \psi(xx')_{[v\rangle}\psi(xx')_{\langle v\rangle}^{-1}, \qquad \widehat{\Sigma}_{[v]\cdot} = \psi(xx')_{[v]\cdot}, \tag{7.21}$$

$v \in V$.

PROOF. Let $v \in V$. If $\psi(xx')_{[v]\dot{\cup}\langle v\rangle}$ is positive definite, then $\psi(xx')_{\langle v\rangle}$ is positive definite and then the equation $R_{[v\rangle}\psi(xx')_{\langle v\rangle} = 0$ implies $R_{[v\rangle} = 0$. Furthermore $\psi(xx')_{[v]^\circ} = \psi(xx')_{[v]\cdot}$ is positive definite. □

REMARK 7.4. Note that Theorem 7.1, Remarks 7.2 and 7.3, and Corollary 7.2 are generalizations of Theorem 3.1, Remarks 3.2 and 3.3, and Corollary 3.1, respectively, in AM (1998).

REMARK 7.5. In many cases, the condition for existence and uniqueness of the ML estimator in Corollary 7.2 is also necessary. In particular, this is always the case when we consider the situation where the representation of G_v on $\mathbb{R}^{[v]}$ is trivial for all $v \in V$ (cf. Remark 6.4), since in this case $R_{[v\rangle}$ is unrestricted for all $v \in V$. The condition in Corollary 7.2 may typically not be fulfilled in situations where both the GS restrictions and the CI restrictions imply independence relations between some of the variables considered.

We shall prove (cf. Proposition 7.2 below) that either the ML estimator exists and is unique with probability 1 wrt all $N(\Sigma)$, $\Sigma \in \mathrm{P}_{G,D}(I)$, or else it will not exist or it will not be unique for any $x \in \mathbb{R}^I$. For this, define for $v \in V$ the sets

$$\Omega_{\mathrm{ex}}(v) = \{x \in \mathbb{R}^I \mid \det(\psi(xx')_{[v]^\circ}) \neq 0\}$$

and

$$\Omega_{\mathrm{un}}(v) = \{x \in \mathbb{R}^I \mid \forall\, R_{[v\rangle} \in \mathrm{M}_{G_v}([v] \times \langle v\rangle) \colon R_{[v\rangle}\psi(xx')_{\langle v\rangle} = 0 \Rightarrow R_{[v\rangle} = 0\},$$

respectively, and let

$$\Omega = \cap(\Omega_{\mathrm{ex}}(v) \cap \Omega_{\mathrm{un}}(v) \mid v \in V). \tag{7.22}$$

Thus, from Theorem 7.1, it follows that the ML estimator $\widehat{\Sigma} = \widehat{\Sigma}(x)$ for $x \in \mathbb{R}^I$ exists and is unique if and only if $x \in \Omega$.

PROPOSITION 7.2. *The set* Ω *is either empty or else the complement* $\mathbb{R}^I \backslash \Omega$ *has Lebesgue measure zero.*

PROOF. Let $v \in V$ and $x \in \mathbb{R}^I$. It follows that $x \in \Omega_{\mathrm{un}}(v)$ if and only if the linear mapping

$$\begin{aligned} \mathrm{M}_{G_v}([v] \times \langle v \rangle) &\to \mathrm{M}_{G_v}([v] \times \langle v \rangle), \\ R_{[v\rangle} &\mapsto R_{[v\rangle} \psi(xx')_{\langle v \rangle} \end{aligned} \tag{7.23}$$

is bijective. This mapping clearly is the restriction of

$$\begin{aligned} \mathrm{M}([v] \times \langle v \rangle) &\to \mathrm{M}([v] \times \langle v \rangle), \\ R_{[v\rangle} &\mapsto R_{[v\rangle} \psi(xx')_{\langle v \rangle}, \end{aligned} \tag{7.24}$$

which is represented by the Kronecker product matrix $I_{[v]} \otimes \psi(xx')_{\langle v \rangle}$ when considered as a mapping $\mathbb{R}^{[v] \otimes \langle v \rangle} \to \mathbb{R}^{[v] \otimes \langle v \rangle}$. Since each entry of $\psi(xx')_{\langle v \rangle}$ is a polynomial in the coordinates of x, each entry of the matrix representing (7.23) in any basis for $\mathrm{M}_{G_v}([v] \times \langle v \rangle)$ is also a polynomial, and hence, the determinant of (7.23) is a polynomial. Thus if $\Omega_{\mathrm{un}}(v)$ is nonempty; the complement $\mathbb{R}^I \backslash \Omega_{\mathrm{un}}(v)$ consists of the zeros of a nonnull polynomial, and hence this set has Lebesgue measure zero [see, e.g., Bourbaki (1963), Chapter VII, Section 3, number 3, Lemma 9].

Now assume that $x \in \Omega_{\mathrm{un}}(v)$. Since each entry of the matrix representing (7.23) is a polynomial in the coordinates of x, this is also the case for the unique solution $\widehat{R}_{[v\rangle}$ to (7.8). Therefore the same statement holds for $\psi(xx')_{[v]^\circ}$, and hence $\det(\psi(xx')_{[v]^\circ}) = p_v(x)$ where $p_v \colon \mathbb{R}^I \to \mathbb{R}$ is a polynomial.

From the definition of $\Omega_{\mathrm{ex}}(v)$ it then follows that

$$\Omega_{\mathrm{ex}}(v) \cap \Omega_{\mathrm{un}}(v) = \Omega_0(v) \cap \Omega_{\mathrm{un}}(v),$$

where $\Omega_0(v) = \{x \in \mathbb{R}^I \mid p_v(x) \neq 0\}$. If $\Omega_{\mathrm{ex}}(v) \cap \Omega_{\mathrm{un}}(v) \neq \varnothing$ then both $\Omega_0(v) \neq \varnothing$ and $\Omega_{\mathrm{un}}(v) \neq \varnothing$. In this case $\mathbb{R}^I \backslash \Omega_0(v)$ and $\mathbb{R}^I \backslash \Omega_{\mathrm{un}}(v)$ have Lebesgue measure zero since both sets consist of the zeros of a nonnull polynomial. Hence the complement $\mathbb{R}^I \backslash (\Omega_{\mathrm{ex}}(v) \cap \Omega_{\mathrm{un}}(v)) = (\mathbb{R}^I \backslash \Omega_0(v)) \cup (\mathbb{R}^I \backslash \Omega_{\mathrm{un}}(v))$ has Lebesgue measure zero. The proposition now follows from (7.22). □

REMARK 7.6 (Independent repetitions). Let $n \in \mathbb{N}$ and define $N = \{1, \dots, n\}$. Now consider n independent repetitions of the GS-ADG model (7.1). In a similar way to that of the GS-LCI models in AM (1998), Section 5.1, it follows that (except for a reparametrization) this model is a GS-ADG model on the sample space $\mathbb{R}^{I \times N}$. As a consequence, Theorem 7.1 and Corollary 7.2 hold when xx' is replaced with the normed empirical covariance matrix $S = (1/n)yy'$, where $y \in \mathbb{R}^{I \times N}$ is the full observation matrix, the columns independent and identically distributed according to the GS-ADG model (7.1). Similarly Proposition 7.2 holds when replacing $\mathbb{R}^I$ with $\mathbb{R}^{I \times N}$ and $\Omega_{\mathrm{ex}}(v)$, $\Omega_{\mathrm{un}}(v)$ with

$$\left\{ y \in \mathbb{R}^{I \times N} \mid \det\left(\psi\left(\frac{1}{n} yy' \right)_{[v]^\circ} \right) = 0 \right\},$$

and

$$\left\{ y \in \mathbb{R}^{I\times N} \mid \forall\ R_{[v\rangle} \in M_{G_v}([v] \times \langle v\rangle)\colon R_{[v\rangle}\psi\left(\frac{1}{n}yy'\right)_{\langle v\rangle} = 0 \Rightarrow R_{[v\rangle} = 0 \right\},$$

respectively, $v \in V$.

REMARK 7.7. For the GS-LCI models, AM (1998) gives explicit conditions for the existence and uniqueness of the ML estimator with probability 1, expressed in terms of the dimensions of the representation of G on each of the multivariate vertices $\mathbb{R}^{[v]}$, $v \in V$; the so-called *structure constants*. (Recall that the GS-LCI models correspond to GS-ADG models where the graph D is transitive and where the symmetry conditions given by G are allowed only to operate inside each multivariate vertex.) Several examples considered (see Examples 7.1–7.7 below) suggest that such numerical conditions exist for the GS-ADG models as well; however, it remains as an open question in what way these conditions should be expressed in a general form.

We shall now continue the seven examples given in Section 6. We use the notation from Remark 7.6; that is, $y \in \mathbb{R}^{I\times N}$ is the full observation matrix of n i.i.d. random observations from the model (7.1), and $S = (1/n)yy'$ denotes the normed empirical covariance matrix. In each of the six examples we first determine the smoothing function $\psi(S)$, and second, we give an expression of the likelihood function such that numerical conditions for the existence and uniqueness of the ML estimator with probability 1 (cf. Remark 7.7), will follow from the standard theory of MANOVA models or from the theory of GS models [see AM (1998), Appendix A]. In all six examples the ML estimators of the D-parameters then can be obtained from Corollary 7.2.

EXAMPLE 7.1 (Continuation of Example 6.1). The smoothing function $\psi(S)$ is given by

$$\begin{aligned}
\psi(S)_{[11]} &= S_{[11]},\\
\psi(S)_{[12]} &= \psi(S)_{[13]} = \tfrac{1}{2}(S_{[12]} + S_{[13]}),\\
\psi(S)_{[22]} &= \psi(S)_{[33]} = \tfrac{1}{2}(S_{[22]} + S_{[33]}),\\
\psi(S)_{[23]} &= \tfrac{1}{2}(S_{[23]} + S_{[32]}).
\end{aligned}$$

The likelihood function can be rewritten as

$$\begin{aligned}
&\mathbf{L}(R, \Lambda, \Lambda_1)\\
&\quad = \det(\Lambda)^{-n} \exp\left(-\tfrac{1}{2}\mathrm{tr}(\Lambda^{-1}((y_{[2]}, y_{[3]}) - R(y_{[1]}, y_{[1]}))(\cdots)')\right)\\
&\qquad \times \det(\Lambda_1)^{-n/2} \exp\left(-\tfrac{1}{2}\mathrm{tr}(\Lambda_1^{-1}(y_{[1]}y_{[1]}'))\right).
\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $2n \geq |J| + |L|$ and $n \geq |J|$.

EXAMPLE 7.2 (Continuation of Example 6.2). The smoothing function $\psi(S)$ is given by

$$\psi(S)_{[11]} = S_{[11]}, \qquad \psi(S)_{[12]} = S_{[12]}, \qquad \psi(S)_{[22]} = S_{[22]},$$
$$\psi(S)_{[13]} = \psi(S)_{[14]} = \tfrac{1}{2}(S_{[13]} + S_{[14]}),$$
$$\psi(S)_{[23]} = \psi(S)_{[24]} = \tfrac{1}{2}(S_{[23]} + S_{[24]}),$$
$$\psi(S)_{[33]} = \psi(S)_{[44]} = \tfrac{1}{2}(S_{[33]} + S_{[44]}),$$
$$\psi(S)_{[34]} = \tfrac{1}{2}(S_{[34]} + S_{[43]}).$$

The likelihood function can be rewritten as

$$\begin{aligned}\mathbf{L}&(R, \Lambda, \Lambda_1, \Lambda_2)\\ &= \det(\Lambda)^{-n} \exp\left(-\tfrac{1}{2}\mathrm{tr}\left(\Lambda^{-1}\left((y_{[3]}, y_{[4]}) - R\begin{pmatrix} y_{[1]} & y_{[1]} \\ y_{[2]} & y_{[2]}\end{pmatrix}\right)(\cdots)'\right)\right)\\ &\quad \times \det(\Lambda_1)^{-n/2}\exp\left(-\tfrac{1}{2}\mathrm{tr}(\Lambda_1^{-1}(y_{[1]}y_{[1]}'))\right)\\ &\quad \times \det(\Lambda_2)^{-n/2}\exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_2^{-1}(y_{[2]}y_{[2]}'))).\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $2n \geq 2|J| + |L|$ and $n \geq 2|J|$.

EXAMPLE 7.3 (Continuation of Example 6.3). The smoothing function $\psi(S)$ is given by

$$\psi(S)_{[11]} = \psi(S)_{[22]} = \tfrac{1}{2}(S_{[11]} + S_{[22]}),$$
$$\psi(S)_{[12]} = \tfrac{1}{2}(S_{[12]} + S_{[21]}),$$
$$\psi(S)_{[13]} = \psi(S)_{[24]} = \tfrac{1}{2}(S_{[13]} + S_{[24]}),$$
$$\psi(S)_{[14]} = \psi(S)_{[23]} = \tfrac{1}{2}(S_{[14]} + S_{[23]}),$$
$$\psi(S)_{[33]} = \psi(S)_{[44]} = \tfrac{1}{2}(S_{[33]} + S_{[44]}),$$
$$\psi(S)_{[34]} = \tfrac{1}{2}(S_{[34]} + S_{[43]}).$$

The likelihood function can be rewritten as

$$\begin{aligned}\mathbf{L}&(A, B, \Lambda_{34}, \Lambda_1, \Lambda_2)\\ &= \det(\Lambda_{34})^{-n} \exp\left(-\tfrac{1}{2}\mathrm{tr}\left(\Lambda_{34}^{-1}\left((y_{[3]}, y_{[4]}) - (A, B)\begin{pmatrix} y_{[1]} & y_{[2]} \\ y_{[2]} & y_{[1]}\end{pmatrix}\right)(\cdots)'\right)\right)\\ &\quad \times \det(\Lambda_{12})^{-n}\exp\left(-\tfrac{1}{2}\mathrm{tr}(\Lambda_{12}^{-1}(y_{[1]}, y_{[2]})(\cdots)')\right).\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $2n \geq 2|J| + |L|$ and $n \geq |J|$.

EXAMPLE 7.4 (Continuation of Example 6.4). The smoothing function $\psi(S)$ is given by

$$\psi(S)_{[11]} = \psi(S)_{[22]} = \tfrac{1}{2}(S_{[11]} + S_{[22]}), \qquad \psi(S)_{[12]} = \tfrac{1}{2}(S_{[12]} + S_{[21]}),$$
$$\psi(S)_{[33]} = \psi(S)_{[44]} = \tfrac{1}{2}(S_{[33]} + S_{[44]}), \qquad \psi(S)_{[34]} = \tfrac{1}{2}(S_{[34]} + S_{[43]}),$$
$$\psi(S)_{[13]} = \psi(S)_{[14]} = \psi(S)_{[23]} = \psi(S)_{[24]} = \tfrac{1}{4}(S_{[13]} + S_{[14]} + S_{[23]} + S_{[24]}).$$

The likelihood function can be rewritten as

$$\begin{aligned}\mathbf{L}(B, \Lambda_{34}, \Lambda_{12}) &= \det(\Lambda_{34})^{-n} \exp\bigl(-\tfrac{1}{2}\mathrm{tr}\bigl(\Lambda_{34}^{-1}((y_{[3]}, y_{[4]}) - B(y_{[1]} + y_{[2]}, y_{[1]} + y_{[2]}))(\cdots)'\bigr)\bigr) \\ &\quad \times \det(\Lambda_{12})^{-n} \exp\bigl(-\tfrac{1}{2}\mathrm{tr}\bigl(\Lambda_{12}^{-1}(y_{[1]}, y_{[2]})(\cdots)'\bigr)\bigr).\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $n \geq |J|$ and $2n \geq |J| + |L|$.

EXAMPLE 7.5 (Continuation of Example 6.5). Let $\binom{y_{l1}}{y_{l2}}$ be the decomposition of $y_{[l]}$ according to (6.19), $l = 1, \ldots, 4$, and let

$$S_{[lm]} = \begin{pmatrix} S_{lm}^{11} & S_{lm}^{12} \\ S_{lm}^{21} & S_{lm}^{22} \end{pmatrix},$$

such that $S_{lm}^{ij} = y_{li} y'_{mj}$, $i, j = 1, 2$, $l, m = 1, \ldots, 4$, $l \leq m$. The smoothing function $\psi(S)$ is then given by

$$\psi(S)_{[lm]} = \begin{pmatrix} \frac{1}{2}(S_{lm}^{11} + S_{lm}^{22}) & \frac{1}{2}(S_{lm}^{12} + S_{lm}^{21}) \\ \frac{1}{2}(S_{lm}^{12} + S_{lm}^{21}) & \frac{1}{2}(S_{lm}^{11} + S_{lm}^{22}) \end{pmatrix},$$

where $l, m = 1, \ldots, 4$, $l \leq m$.

The likelihood function can be rewritten as

$$\begin{aligned}\mathbf{L}(R_3, \Lambda_3, R_4, \Lambda_4, \Lambda_1, \Lambda_2) &= \det(\Lambda_3)^{-n/2} \exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_3^{-1}((y_{[3]} - R_3 y_{\langle 3\rangle})(\cdots)'))) \\ &\quad \times \det(\Lambda_4)^{-n/2} \exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_4^{-1}((y_{[4]} - R_4 y_{\langle 4\rangle})(\cdots)'))) \\ &\quad \times \det(\Lambda_1)^{-n/2} \exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_1^{-1}(y_{[1]} y'_{[1]}))) \\ &\quad \times \det(\Lambda_2)^{-n/2} \times \exp(-\tfrac{1}{2}\mathrm{tr}(\Lambda_2^{-1}(y_{[2]} y'_{[2]}))),\end{aligned}$$

where the parameters $R_3, \Lambda_3, R_4, \Lambda_4, \Lambda_1, \Lambda_2$ have the additional restrictions (6.20). From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $n \geq \max(|J_1| + |J_2| + |J_3|, |J_3| + |J_4|)$.

EXAMPLE 7.6 (Continuation of Example 6.6). Let $\binom{y_1}{y_2}$ be the decomposition of $y_{[2]}$ according to (6.21) and let

$$S_{[22]} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad S_{[23]} = \begin{pmatrix} S_{13} \\ S_{23} \end{pmatrix}, \qquad S_{[24]} = \begin{pmatrix} S_{14} \\ S_{24} \end{pmatrix},$$

where $S_{ij} = y_i y'_j$, $i, j = 1, 2$ and $S_{ij} = y_i y'_{[j]}$, $i = 1, 2$, $j = 3, 4$. The smoothing function $\psi(S)$ is given by

$$\begin{aligned}
\psi(S)_{[22]} &= \begin{pmatrix} \frac{1}{2}(S_{11} + S_{22}) & \frac{1}{2}(S_{12} + S_{21}) \\ \frac{1}{2}(S_{12} + S_{21}) & \frac{1}{2}(S_{11} + S_{22}) \end{pmatrix}, \\
\psi(S)_{[23]} &= \begin{pmatrix} \frac{1}{2}(S_{13} + S_{24}) \\ \frac{1}{2}(S_{14} + S_{23}) \end{pmatrix}, \\
\psi(S)_{[24]} &= \begin{pmatrix} \frac{1}{2}(S_{14} + S_{23}) \\ \frac{1}{2}(S_{13} + S_{24}) \end{pmatrix}, \\
\psi(S)_{[33]} &= \psi(S)_{[44]} = \tfrac{1}{2}(S_{[33]} + S_{[44]}), \\
\psi(S)_{[34]} &= \tfrac{1}{2}(S_{[34]} + S_{[43]}).
\end{aligned}$$

The likelihood function can be rewritten as

$$\begin{aligned}
&\mathbf{L}(A, B, \Lambda, \Phi, \Delta) \\
&\quad = \det(\Lambda)^{-n} \exp\left(-\tfrac{1}{2}\mathrm{tr}\left(\Lambda^{-1}\left((y_{[3]}, y_{[4]}) - (A, B)\begin{pmatrix} y_1 & y_2 \\ y_2 & y_1 \end{pmatrix}\right)(\cdots)'\right)\right) \\
&\quad\quad \times \det\begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix}^{-n/2} \exp\left(-\tfrac{1}{2}\mathrm{tr}\left(\begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix}^{-1} y_{[2]} y'_{[2]}\right)\right).
\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $n \geq |J|$ and $2n \geq 2|J| + |L|$.

EXAMPLE 7.7 (Continuation of Example 6.7). We use the same notation as in Example 7.6. The smoothing function $\psi(S)$ is then given by

$$\begin{aligned}
\psi(S)_{[22]} &= \begin{pmatrix} \frac{1}{2}(S_{11} + S_{22}) & \frac{1}{2}(S_{12} + S_{21}) \\ \frac{1}{2}(S_{12} + S_{21}) & \frac{1}{2}(S_{11} + S_{22}) \end{pmatrix}, \\
\psi(S)_{[23]} &= \psi(S)_{[24]} = \begin{pmatrix} \frac{1}{4}(S_{13} + S_{14} + S_{23} + S_{24}) \\ \frac{1}{4}(S_{13} + S_{14} + S_{23} + S_{24}) \end{pmatrix}, \\
\psi(S)_{[33]} &= \psi(S)_{[44]} = \tfrac{1}{2}(S_{[33]} + S_{[44]}), \\
\psi(S)_{[34]} &= \tfrac{1}{2}(S_{[34]} + S_{[43]}).
\end{aligned}$$

The likelihood function can be rewritten as

$$\begin{aligned}
&\mathbf{L}(A, \Lambda, \Phi, \Delta)\\
&\quad = \det(\Lambda)^{-n} \exp\Big(-\tfrac{1}{2}\mathrm{tr}(\Lambda^{-1}((y_{[3]}, y_{[4]}) - A(y_1 + y_2, y_1 + y_2))(\cdots)')\Big)\\
&\qquad \times \det\begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix}^{-n/2} \exp\bigg(-\tfrac{1}{2}\mathrm{tr}\bigg(\begin{pmatrix} \Phi & \Delta \\ \Delta & \Phi \end{pmatrix}^{-1} y_{[2]}y'_{[2]}\bigg)\bigg).
\end{aligned}$$

From this expression, it can be seen that $\widehat{\Sigma}$ exists and is unique with probability 1 if and only if $n \geq |J|$ and $2n \geq |J| + |L|$.

8. Further discussion. Within the theory of the GS-ADG models, the following problems are currently under investigation by the author.

Definition of structure constants. This problem is discussed in Remark 7.7.

Testing problems. Let $C = (W, F)$ be a second ADG with an associated partitioning $(J_w \mid w \in W)$ of I; that is, $I = \dot{\cup}(J_w \mid w \in W)$, and let $\varphi\colon D \to C$ be a surjective ADG homomorphism such that

$$I_v = \dot{\cup}(J_w \mid w \in W,\ \varphi(w) = v),$$

for all $v \in V$. (In many applications, $V = W$, $E \subset F$ and φ is the identity mapping.) Furthermore, let H be a subgroup of G. By AP (1998), Proposition 3.1(i), $\mathrm{P}_D(I) \subseteq \mathrm{P}_C(I)$, and it is easy to see that $\mathrm{P}_G(I) \subseteq \mathrm{P}_H(I)$. It then follows that $\mathrm{P}_{G,D}(I) \subseteq \mathrm{P}_{H,C}(I)$, and hence the problem of testing the model

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_{G,D}(I)) \tag{8.1}$$

versus the model

$$(N(\Sigma) \mid \Sigma \in \mathrm{P}_{H,C}(I)) \tag{8.2}$$

is well defined [cf. AM (1998), Section 4, and AP (1998), Section 9].

It may be shown that if the model (8.2) is regular; that is, the ML estimator exists and is unique with probability 1 (cf. Proposition 7.2), then the model (8.1) is also regular [cf. AM (1998), Proposition 4.1]. Thus if the model (8.2) is regular, it follows from (7.20) that the likelihood ratio test statistic Q for testing (8.1) against (8.2) exists with probability 1 and is given by

$$Q(x) = \left(\frac{\prod(\det(\psi^H(xx')_{[w]^\circ}) \mid w \in W)}{\prod(\det(\psi^G(xx')_{[v]^\circ}) \mid v \in V)}\right)^{1/2},$$

where $x \in \mathbb{R}^I$. Furthermore, it is possible to show that under the model (8.1), Q and the family of ML estimators $(\widehat{\Sigma}_{[v]^\circ} \mid v \in V)$ are mutually independent [cf. AM (1998), Theorem 4.1, and AP (1998), Proposition 9.2]. A characterization of the central distribution of Q generalizing that of the GS-LCI models [cf. AM (1998), Section 4], may then be possible. Note however, that the genralization is not straightforward since in the case of the GS-ADG models, the ML estimators $\widehat{\Sigma}_{[v]^\circ}$, $v \in V$, are not mutually independent.

Nonzero mean-value hypotheses. The assumption that the expectation of the normal distributions in the model (7.1) is zero can be removed. Thus if $L \subseteq \mathbb{R}^I$ is an $\mathrm{M}_{G,D}(I)$-*subspace*, that is, $\mathrm{M}_{G,D}(I)L = L$, ML estimators for the model

$$(N(\xi, \Sigma) \mid (\xi, \Sigma) \in L \times \mathrm{P}_{G,D}(I))$$

can be derived explicitly [cf. AP (1998), Definition 6.2(iii) and Section 7].

Acknowledgments. The author thanks Steen Andersson for valuable discussion and a referee for careful reading of the manuscript. Also thanks are due to Hans Brøns and Søren Tolver Jensen for comments and discussion.

REFERENCES

ANDERSEN, H. H., HØJBJERRE, M., SØRENSEN, D. and ERIKSEN, P. S. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution*. *Lecture Notes in Statist.* Springer, New York.

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995a). On the relation between conditional independence models determined by finite distributive lattices and directed acyclic graphs. *J. Statist. Plann. Inference* **48** 25–46.

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995b). A graphical characterization of the lattice conditional independence models. *Ann. Math. Artificial Intelligence* **21** 27–50.

ANDERSSON, S. A. and MADSEN, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** 525–572.

ANDERSSON, S. A. and PERLMAN, M. D. (1998). Normal linear regression models with recursive graphical Markov structure. *J. Multivariate Anal.* **66** 133–187.

BOURBAKI, N. (1963). Íntegration. In *Éléments de Mathématique*, Chapitres 7, 8. Hermann, Paris.

HYLLEBERG, B., JENSEN, M. and ØRNBØL, E. (1993). Graphical symmetry models. M.S. thesis, Aalborg Univ.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF COPENHAGEN
5 UNIVERSITETSPARKEN, DK-2100
COPENHAGEN Ø
DENMARK
E-MAIL: jmadsen@math.ku.dk