AGGREGATION AND COHERENCY IN NETWORKS AND MARKOV CHAINS*

F. Delebecque, J. P. Quadrat
INRIA
Domaine de Voluceau - Rocquencourt
78150 Le Chesnay, France

and

P. V. Kokotovic
Coordinated Science Laboratory
University of Illinois
1101 W. Springfield Ave.
Urbana, Illinois 61801, USA

Abstract

A unified treatment of aggregability, lumpability, coherency, reversibility, partial balance, and similar properties of dynamic network models and Markov chains clarifies most of the known and reveals some new conditions for model simplification. Coherency condition, well known in power systems, implies the existence of a finite state filter for Markov chains. Aggregability and coherency yield a new condition for decentralized computation of the invariant measure, which differs from the partial balance condition. Such conditions are indispensable tools in the study of large scale dynamic systems and networks of queues.

*This work was supported in part by the U.S. Department of Energy, Electric Energy Systems Division under Contract 01-81RA50422 with INRIA; in part by the Department of Energy, Electric Energy Systems Division under Contract DE-AC01-81RA-50658 with Dynamic Systems, P.O. Box 423, Urbana, IL 61801; and in part by the National Science Foundation under Grant ECS-79-19396.

1. Introduction

In as diverse fields as economics, fluid dynamics, vibration analysis of structures, models of queues, power systems and many others, concepts of aggregation [1-4,23-30], lumping [5-11], mass condensation [12,13], coherency [14-19], and partial balance [20-22] appear independently, but refer to a common set of properties allowing model simplifications. This paper is an attempt to treat such properties in a unified manner. The unification is possible because of the common decomposition-aggregation pattern. First, in the original n-dimensional system, $N \ll n$ "local" subsystems are identified. Second, each local subsystem is represented by an "aggregate" variable. Third, an N-dimensional aggregate model is formed describing the interaction of the aggregate variables.

From a general algebraic formulation in Section 2 we deduce most of the known and some new conditions for aggregability and coherency of electromechanical models of power systems (Section 3) and for lumpability of Markov chains (Section 4). Coherency is shown to be a weak lumpability property which implies the existence of a finite state filter. In Section 5 we discuss relations between different conditions, such as reversibility and partial balance, under which the computation of the invariant measure can be decentralized and the residual system can be decomposed into N local subsystems.

2. Decomposition and Projection Views of Aggregation and Coherency

Aggregation can be seen as a coordinate transformation into a block-triangular model with the aggregate appearing as an independent diagonal block, or, equivalently, as a projection of the n original variables x to an N-dimensional subspace independent of the phenomena in the rest of the n-dimensional space. To characterize these two aspects of aggregation we introduce (n,N)-matrix B and (N,n)-matrix C and their annihilators $B_\perp$ and $C_\perp$ such that

$$B_\perp B = 0, \quad CC_\perp = 0; \quad (CB)^{-1} \text{ and } (B_\perp C_\perp)^{-1} \text{ exist} \tag{2.1}$$

$$\mathcal{R}(B) \oplus \mathcal{N}(C) = R^n, \tag{2.2}$$

where $\mathcal{R}$ is the range space, $\mathcal{N}$ is the null space and prime denotes a transpose. When C is the aggregation matrix, then the transformation of x into a set of N <u>"aggregate" variables</u> $y = Cx$ and n-N <u>"local" variables</u> $z = B_\perp x$ is

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} C \\ B_\perp \end{bmatrix} x = T_C x \tag{2.3}$$

where x, y, and z are column vectors. The inverse transformation is

$$T_C^{-1} = [B(CB)^{-1} \vdots C_\perp (B_\perp C_\perp)^{-1}]. \tag{2.4}$$

The aggregation is a projection P on $\mathcal{R}(B)$ parallel to $\mathcal{N}(C)$,

$$P = B(CB)^{-1}C \tag{2.5}$$

where obviously $PP = P, PB = B$, and $CP = C$. The projection on $\mathcal{N}(C)$ parallel to $\mathcal{R}(B)$ is

$$\tilde{P} = I - P = C_\perp (B_\perp C_\perp)^{-1} B_\perp \tag{2.6}$$

and, hence, $\tilde{P}\tilde{P} = \tilde{P}$, $P\tilde{P} = \tilde{P}P = 0$, $\tilde{P}C_\perp = C_\perp$, and $B_\perp \tilde{P} = B_\perp$. A system whose (n,n)-matrix in the original basis x is A, is represented in the new basis (2.3) by

$$T_C A T_C^{-1} = \begin{bmatrix} CAB(CB)^{-1} & CAC_\perp (B_\perp C_\perp)^{-1} \\ B_\perp AB(CB)^{-1} & B_\perp AC_\perp (B_\perp C_\perp)^{-1} \end{bmatrix}. \tag{2.7}$$

The so-called "C-aggregate"

$$A_C = CAB(CB)^{-1} \tag{2.8}$$

separates from the rest of the system, that is $T_C A T_C^{-1}$ is lower block triangular iff

$$CAC_\perp = 0 \iff A \text{ is C-aggregable.} \tag{2.9}$$

In coherency $T_C A T_C^{-1}$ is upper block triangular, that is

$$B_\perp AB = 0 \iff A \text{ is B-coherent.} \tag{2.10}$$

If both (2.9) and (2.10) hold then $T_C A T_C^{-1}$ is block-diagonal. A further desirable property would be the "decentralization" of the residual $B_\perp AC_\perp (B_\perp C_\perp)^{-1}$ into N diagonal blocks representing "local subsystems." Although such ideal properties seldom occur in reality, their study helps us to identify classes of systems in which they approximately hold. A unified framework for this study is provided by the following equivalent forms of necessary and sufficient conditions for C-aggregability and B-coherency.

<u>Proposition 2.1</u> (C-aggregability): Each of the following statements is equivalent to (2.9).

2.1.1. $A\,\mathcal{N}(C) = \mathcal{N}(C)$ and, hence, $A'\mathcal{R}(C') = \mathcal{R}(C')$.

2.1.2. There exists projector P such that $\mathcal{N}(P) = \mathcal{N}(C)$ and $PA = PAP$.

2.1.3. There exists projector $\tilde{P}$ such that $\mathcal{R}(\tilde{P}) = \mathcal{N}(C)$ and $A\tilde{P} = \tilde{P}A\tilde{P}$.

2.1.4. There exists A_C such that $CA = A_C C$.

2.1.5. There exists $\tilde{A}_C$ such that $AC_\perp = C_\perp \tilde{A}_C$.

Proof is by straightforward verification. As an illustration let us note from (2.5) that $CPA = CA$ and, from 2.1.2, that

$$CPA = CPAP = CAP = CAB(CB)^{-1}C = A_C C \qquad (2.11)$$

which verifies 2.1.2 $\Leftrightarrow$ 2.1.4 and exhibits the same aggregate matrix A_C as in (2.8).

Proposition 2.2 (B-coherency): Each of the following statements is equivalent to (2.10)

2.2.1. $A\mathcal{R}(B) = \mathcal{R}(B)$ and, hence, $A'\mathcal{N}(B') = \mathcal{N}(B')$.

2.2.2. There exists a projector P such that $\mathcal{R}(P) = \mathcal{R}(B)$ and $AP = PAP$.

2.2.3. There exists a projector $\tilde{P}$ such that $\mathcal{N}(\tilde{P}) = \mathcal{R}(\tilde{P})$ and $\tilde{P}A = \tilde{P}A\tilde{P}$.

2.2.4. There exists A_B such that $AB = BA_B$.

2.2.5. There exists $\tilde{A}_B$ such that $B_\perp A = \tilde{A}_B B_\perp$.

Matrix A is both C-aggregable and B-coherent iff any one of the statements of Proposition 2.1 and any one of the statements of Proposition 2.2 hold simultaneously. In particular, $T_C A T_C^{-1}$ is block-diagonal iff 2.1.2 and 2.2.2 hold, that is iff

$$PA = AP. \qquad (2.12)$$

If the variables are arranged as row-vectors, then we employ B as the aggregation matrix in the transformation of row-vectors

$$T_B = [B \quad C\,], \quad T_B^{-1} = \begin{bmatrix} (CB)^{-1}C \\ (B_\perp C_\perp)^{-1}B_\perp \end{bmatrix}. \qquad (2.13)$$

Then B-aggregability of A is the property that the transformed matrix

$$T_B^{-1}AT_B = \begin{bmatrix} (CB)^{-1}CAB & (CB)^{-1}CAC_\perp \\ (B_\perp C_\perp)^{-1}B_\perp AB & (B_\perp C_\perp)^{-1}B_\perp AC_\perp \end{bmatrix} \tag{2.14}$$

be lower block-triangular, and the C-coherency is that $T_B^{-1}AT_B$ be upper block-triangular. The aggregate is the same as in 2.2.4, that is

$$A_B = (CB)^{-1}CAB. \tag{2.15}$$

The corresponding decomposition

$$\mathcal{R}(C') \oplus \mathcal{N}(B') = R^n \tag{2.16}$$

is dual to (2.2) and the projectors are P' and $\tilde{P}'$. Clearly, if A is C-aggregable then A' is C'-coherent. If A is B-coherent, then A' is B'-aggregable. Therefore Propositions 2.1 and 2.2, respectively, are also necessary and sufficient for C'-coherency and B'-aggregability of A'. The aggregates A_C and A_B are similar matrices

$$A_B = (CB)^{-1}A_C(CB), \quad \text{hence} \quad CB = I \Rightarrow A_B = A_C = \bar{A}. \tag{2.17}$$

Furthermore, (2.12) is necessary and sufficient for the two pairs of properties to hold simultaneously.

Let us give a system theory interpretation of these concepts. In dynamic systems C-aggregability indicates that an output $y = Cx$ is an N-dimensional aggregate of the n-dimensional state x,

$$\frac{dx}{dt} = Ax, \quad y = Cx, \quad CA = A_C C \iff \dot{y} = A_C y, \quad \forall t, \quad \forall x(0) \in R^n \tag{2.18}$$

that is $\mathcal{N}(C)$ is the maximal A-invariant subspace of R^n unobservable from y, [2,3,4]. The less familiar B-coherency concept is implicit in economic literature [1] and more explicit in power system practice when synchronous machines which "swing together" after a disturbance form "coherent areas" [14-19]. If the disturbance is modeled as an N-dimensional input v and the input matrix is B, then B-coherency means that

$$\frac{dx}{dt} = Ax + Bv,\ x(0)\in\mathcal{R}(B),\ AB = BA_B \ \Leftrightarrow\ x(t)\in\mathcal{R}(B),\ \forall v(t)\in R^N,\ \forall t\geq 0 \qquad (2.19)$$

that is $\mathcal{R}(B)$ is the maximal A-invariant subspace of R^n controllable by v.

3. Aggregation and Coherency in Networks

In networks and Markov chains B and C depend on the partition into groups (areas) and the weights with which members of groups enter the aggregates. Let a set $E=\{1,\ldots,n\}$ of n states, machines or similar elements consist of N groups $\mathcal{U}=\{J\,|\,J\subset E,\ \cup J=E,\ \cap J=\phi\}$ and denote by U the (n,N)-partition matrix whose (i,J)-entry is

$$u_{iJ} = \begin{cases} 1 & \text{if} \quad i\in J \\ & \qquad\qquad \forall i\in E,\ \ \forall J\in\mathcal{U}. \\ 0 & \text{if} \quad i\notin J \end{cases} \qquad (3.1)$$

For example, if in the network in Fig. 1 the set of n=5 nodes $E=\{1,2,3,4,5\}$ is partitioned into N=2 groups {1,2,4} and {3,5}, then

$$U = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \Longleftrightarrow \begin{array}{l} i\in\{1,2,4\} \Leftrightarrow J=1 \\ i\in\{3,5\} \Leftrightarrow J=2 \end{array} . \qquad (3.2)$$

We consider aggregation with only one aggregate variable per group and form B and C using U, a weighting matrix W, and a scaling matrix S

$$W = \operatorname{diag}(w_1,\ldots,w_n), \quad S = U'WU = \operatorname{diag}(s_1,\ldots,s_N) \tag{3.3}$$

with all weights w_i positive. It is useful to note that s_J is the sum of weights in group J

$$s_J = \sum_{j\in J} w_j, \quad J = 1,\ldots,N \tag{3.4}$$

and that $S^{-1}U'W$ scales each sum of group weights to 1. Thus for U in (3.2) we have $s_1 = w_1 + w_2 + w_4$, $s_2 = w_3 + w_5$ and

$$S^{-1}U'W = \begin{bmatrix} \frac{w_1}{s_1} & \frac{w_2}{s_1} & 0 & \frac{w_4}{s_1} & 0 \\ 0 & 0 & \frac{w_3}{s_2} & 0 & \frac{w_5}{s_2} \end{bmatrix}. \tag{3.5}$$

For our study of networks and Markov chains we choose

$$B = U, \quad C = S^{-1}U'W \Rightarrow P = UC \tag{3.6}$$

and specialize the aggregability and coherency conditions not only for A, but also for

$$\hat{A} = W^{-1}A'W \quad \Rightarrow \quad \hat{P} = P, \tag{3.7}$$

which will be helpful in subsequent applications. From the equivalent relations

$$\begin{aligned} CA = \bar{A}C &\Longleftrightarrow U'WAW^{-1} = S\bar{A}S^{-1}U' \\ &\Longleftrightarrow \hat{A}U = UA \\ &\Longleftrightarrow \hat{A} \text{ is B-coherent} \end{aligned} \tag{3.8}$$

we make the following conculsion

<u>Corollary 3.1</u>:

3.1.1. A is C-aggregable ⟷ $\hat{A}$ is B-coherent

3.1.2. A is B-coherent ⟷ $\hat{A}$ is C-aggregable.

A further observation is that if $AP = \hat{A}P$, that is if

$$AUC = \hat{A}UC \tag{3.9}$$

then the aggregate matrices $\bar{A} = CAB$ and $\bar{\hat{A}} = C\hat{A}B$ are equal, $\bar{A} = \bar{\hat{A}}$, and the aggregability and coherency imply each other.

As an application we consider networks consisting of n storage nodes connected by branches without storage effects. Typically the state equations of such networks are

$$\text{(a)} \quad \frac{dx}{dt} = Ax \quad \text{or} \quad \text{(b)} \quad \frac{d^2x}{dt^2} = Ax, \qquad A = M^{-1}K \tag{3.10}$$

where $M = \text{diag}(m_1, \ldots, m_2)$ characterizes node storages and K represents the branches. For example, if (a) is an RC network, then the entries of M are capacitances and the entries of K are resistances. If (b) is a mass-spring

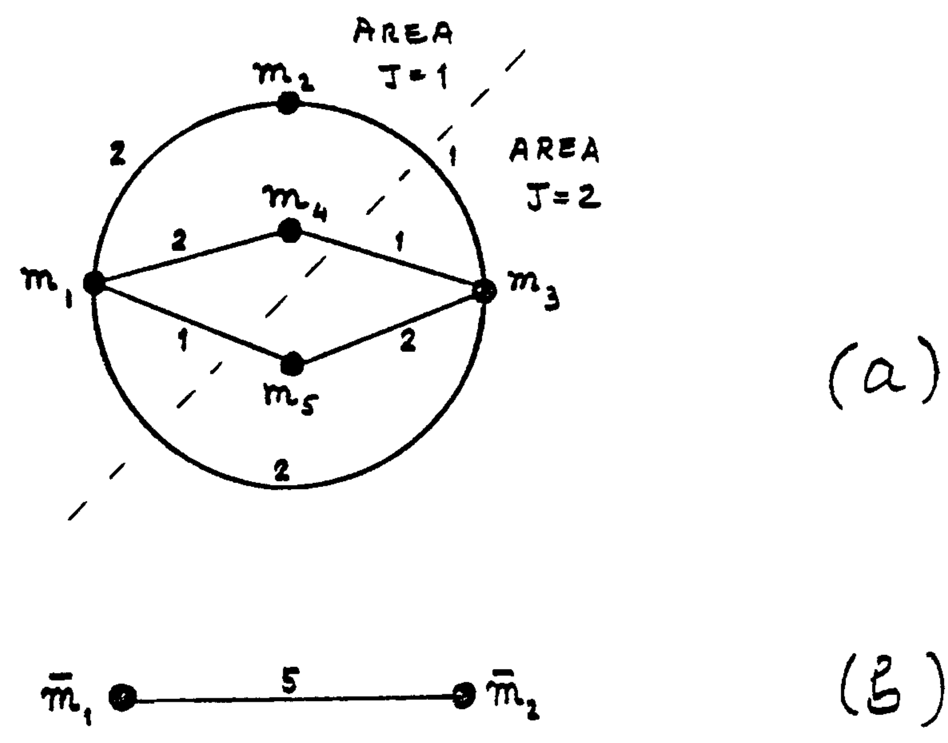

Fig. 1

system then M is the mass matrix and K is the stiffness matrix. When the eigenvalues λ of A are real nonpositive, then the same analysis applies to both (a) and (b), except that instead of R^n and λ in (a), the state space in (b) is R^{2n} and the nonzero modes $\pm\sqrt{\lambda}$ are oscillatory. To be specific, we will consider (b) in (3.10) as a linearized electromechanical model of a power system in which m_i is the inertia of the i-th synchronous machine (node i) and K is the lossless admittance matrix reduced to the machine nodes. An example is the power system in Fig. 1 where

$$M = \mathrm{diag}(5,1,4,1,1), \quad K = \begin{bmatrix} -7 & 2 & 2 & 2 & 1 \\ 2 & -3 & 1 & 0 & 0 \\ 2 & 1 & -6 & 1 & 2 \\ 2 & 0 & 1 & -3 & 0 \\ 1 & 0 & 2 & 0 & 3 \end{bmatrix}. \tag{3.11}$$

Commonly used aggregate variables for such networks are the area centers of inertia

$$y_J = \frac{1}{\bar{m}_J} \sum_{j\in J} m_j x_j, \quad \bar{m}_J = \sum_{j\in J} m_j, \quad j = 1,\ldots,N \tag{3.12}$$

where $\bar{m}_J$ is the aggregated mass of area J. With $B = U$, $M = W$, and hence $S = \bar{M} = U'MU$, the vector form of (3.12) is

$$y = \bar{M}^{-1}U'Mx = Cx. \tag{3.13}$$

Network matrix $A = M^{-1}K$ is aggregable with respect to the centers of inertia (3.12) for the partition $\mathcal{U}$ iff

$$U\bar{M}^{-1}U'K = U\bar{M}^{-1}U'KU\bar{M}^{-1}U'M. \tag{3.14}$$

which is a consequence of 2.1.2. The aggregate network consists of masses $\bar{m}_1,\ldots,\bar{m}_N$ connected with springs whose stiffness matrix is $\bar{K}$, since

$$\bar{A} = CAB = \bar{M}^{-1}\bar{K}, \quad \bar{K} = U'KU. \tag{3.15}$$

It can be readily verified that M and K in (3.11) and Fig. 1a satisfy (3.14) for partition (3.2) and that the resulting network

$$\bar{M} = \begin{bmatrix} 7 & 0 \\ 0 & 5 \end{bmatrix}, \quad \bar{K} = \begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix}, \tag{3.16}$$

represents the two machine system in Fig. 1b. The meaning of (3.14) is clearer from the equivalent condition

$$U'KM^{-1}U_{\perp}' = 0 \tag{3.17}$$

which is a consequence of $CAC_{\perp} = 0$, see (2.9), and $C_{\perp} = M^{-1}U_{\perp}'$, where $U_{\perp}$ is a full rank (n-N,n) matrix such that $U_{\perp}U = 0$. To select $U_{\perp}$ we pick in each group J a "local reference" x_j, $j \in J$, and as local variables z define $x_i - x_j$, $\forall i \in J$, $i \neq j$, that is

$$z = U_{\perp} x = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} x_2 - x_1 \\ x_4 - x_1 \\ x_5 - x_3 \end{bmatrix} \tag{3.18}$$

where we have shown $U_{\perp}$ and z when in (3.11) the reference of {1,2,4} is x_1 and the reference of {3,5} is x_3. From (3.17) it follows that $M^{-1}K$ is aggregable with respect to (3.13) iff

$$\frac{1}{m_j} \sum_{i \in J} k_{ij} = \frac{1}{m_r} \sum_{i \in J} k_{ir}, \quad \forall j,r \in J, \quad \forall J \in \mathcal{U}. \tag{3.19}$$

The local reference form of $U_{\perp}$ is also convenient for an interpretation of coherency. From (2.19) and (3.18) coherency is the property that

$$x(0) \in \mathcal{R}(U) \Leftrightarrow z(t) = 0, \quad \forall t \tag{3.20}$$

which is an analytic expression of the practical observation that "the machines in the same area swing together," since

$$z(t) = 0 \Leftrightarrow x_i(t) = x_j(t); \quad \forall t, \quad \forall i,j \in J, \quad \forall J \in \mathcal{U} . \tag{3.21}$$

A set of scalar coherency conditions similar to the aggregability conditions (3.19) can be obtained from

$$U_{\perp}M^{-1}KU = 0 \tag{3.22}$$

which is a consequence of (2.10). Finally, using (3.17) and (3.28), we see that $M^{-1}K$ is both aggregable and coherent iff

$$U_{\perp}M^{-1}KU = U_{\perp}M^{-1}K'U = 0. \tag{3.23}$$

The special case when K is symmetric and, hence, $A = \hat{A}$ has appeared in the coherency based aggregation of power systems [17-18].

4. Lumpability, Coherency, and Finite State Filters of Markov Chains

In this section we first review the notion of lumpability [5] of a recurrent Markov chain X_t defined on $E = \{1,\ldots,n\}$. Then we show that coherency is a form of weak lumpability and that under the condition of coherency there exists a finite state filter for X_t.

Let $a_{ij} = \mathcal{P}(X_{t+1} = j \mid X_t = i)$ be the (i,j)-entry of the transition probability matrix A of X_t, and p(t) be the row n-vector whose i-th component is $p_i(t) = \mathcal{P}(X_t = i)$. Then

$$p(t+1) = p(t)A, \qquad t = 0,1,2,\ldots \tag{4.1}$$

As in (3.1) we partition E into N groups $\mathcal{U}$ and define the $\mathcal{U}$-valued process Y_t such that

$$Y_t = J \quad \text{whenever} \quad X_t = j \in J. \tag{4.2}$$

We point out that Y_t does not have to be a Markov process. Following [5, page 124] Markov process X_t is called lumpable with respect to partition $\mathcal{U}$ iff Y_t is a Markov process on $\mathcal{U}$, that is iff row N-vector r(t), whose J-th component is $r_J(t) = \mathcal{P}(Y_t = J)$, satisfies

$$r(t+1) = r(t)\bar{A}, \quad t = 0,1,2,\ldots \tag{4.3}$$

for some (N,N) stochastic matrix $\bar{A}$ and all initial probability measures p(0) of X_o. By (4.1), (4.2), and (4.3)

$$r(t+1) = p(t+1)U = p(t)AU = r(t)\bar{A} = p(t)U\bar{A}, \quad t = 0,1,2,\ldots \tag{4.4}$$

holds for all p(0). Hence, X_t is lumpable iff

$$AU = U\bar{A} \tag{4.5}$$

or, equivalently, iff

$$\sum_{k \in K} a_{jk} = \bar{a}_{JK}, \quad \forall j \in J, \quad K = 1,\ldots,N \tag{4.6}$$

that is iff the probability for X_t to go from $j \in J$ to group K depends only on J, and is independent of j.

Corollary 4.1: Markov chain X_t is lumpable with respect to partition $\mathcal{U}$ iff its transition matrix A satisfies Proposition 2.2 with B = U, that is $AU = UA_B$. Then

$$A_B = \bar{A} = CAB = S^{-1}U'WAU = (U'U)^{-1}U'AU \tag{4.7}$$

is independent of the choice of W in (3.3).

Markov chain X_t is called "coherent with weight W" iff, as in Proposition 2.1,

$$C(w)A = \bar{A}C(w), \quad C(w) = S^{-1}U'W. \tag{4.8}$$

In this case the invariant measures q of A and $\bar{q}$ of $\bar{A}$ are related by

$$\bar{q}C(w)A = \bar{q}\bar{A}C(w) = \bar{q}C(w) = q \tag{4.9}$$

which in view of $CU = I$, implies that

$$\bar{q} = qU \tag{4.10}$$

From (4.9) and (4.10) we see that coherency is possible only with the weight W satisfying

$$C_{Jj}(w) = \frac{w_j}{\sum_{j\in J} w_j} = \frac{q_j}{\sum_{j\in J} q_j} = C_{Jj}(q); \quad j\in J, \quad 0 \text{ elsewhere} \tag{4.11}$$

and has the following probabilistic interpretation.

<u>Corollary 4.2</u>: If Markov chain X_t is coherent with weight W then $C(w) = C(q) = q^{\mathcal{U}}$ is the conditional law of q, the invariant measure of A, knowing the partition $\mathcal{U}$. Moreover if the initial probability measure p(0) of X_t is in the row space of C, $p(0) = r(0)C$, then Y_t in (4.2) is a Markov chain with the transition matrix $\bar{A}$ in (4.7) and its invariant measure $\bar{q}$ is qU.

It should be pointed out that coherency and lumpability (i.e., aggregability) are different properties, each being a sufficient condition for the weak lumpability as defined in [5].

An important consequence of coherency is the existence of a finite state filter

$$\pi(t) = \text{conditional law of } X_t \text{ knowing } (Y_o,\ldots,Y_t). \tag{4.12}$$

In general $\pi(t)$, $t = 0,1,\ldots,$ can visit an infinite set of points in the simplex $\mathcal{S}$ of R^n,

$$\mathcal{S} = \{\pi\in R^n : \pi_i \geq 0, \ \sum_{i=1}^{n} \pi_i = 1\} \tag{4.13}$$

In the case of coherency (4.11), however, if there exists r(0) such that $\pi(0) = r(0)C$, then filter $\pi(t)$ visits only N points q^J of $\mathcal{S}$, where q^J is the J-th row of C. Its j-th component q_j^J, $j = 1,\ldots,n$ is the conditional probability for the invariant probability measure of A, of an event to be in j, knowing that it is in J. Thus $\pi(t)$ is a process on $\mathcal{S}$ whose law at time t belongs to the set of probability measures $\mathcal{M}_+^1(\mathcal{S})$ on $\mathcal{S}$. This process is a finite state Markov chain whose states are the points q^J, $J = 1,\dots,N$. The JK-entry of its transition matrix is

$$\mathcal{P}(\pi(t+1) = q^K | \pi(t) = q^J) = q^J A \mathbb{1}^K = c^J A u^K \qquad (4.14)$$

where $\mathbb{1}^K$ is the indicator vector of the set K, c^J is the J-th row of C and u^K is the K-th column of U.

When A is both lumpable and coherent the calculation of its invariant probability measure q is "decentralized." To show this we denote by A^{JK} the submatrix of the transitions from $j \in J$ to $k \in K$. Then the conditions (2.12) for simultaneous aggregability and coherency can be written as

$$q_+^J A^{JK} = \bar{A}_{JK} q_+^K, \quad J,K = 1,\dots,N \qquad (4.15)$$

where q_+^J is the row vector of the entries of q^J for $j \in J$. Thus q_+^J is the invariant measure of A^{JJ} normalized by $\bar{A}_{JJ}$,

$$q_+^J = q_+^J \frac{A^{JJ}}{\bar{A}_{JJ}} . \qquad (4.16)$$

Since

$$\bar{A}_{JJ} = \sum_{k \in J} a_{jk}, \quad j \in J, \quad J = 1,\dots,N \qquad (4.17)$$

we see that q_+^J is obtained by solving the J-th local problem (4.15) with J=K.

Corollary 4.3: Invariant measure q of a Markov chain X_t which is aggregable and coherent with respect to partition $\mathcal{U}$ is given by

$$q_j = \bar{q}_J q_j^J, \quad \forall j \in J, \quad \forall J \in \mathcal{U} \qquad (4.18)$$

where $\bar{q}$ is the invariant measure of the aggregate chain Y_t and q_j^J is obtained from (4.16).

Let us illustrate the notions of aggregability and coherency and the existence of a finite state filter by the simple three state chain in Fig. 2a, whose transition matrix A and invariant measure q are

$$A = \begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, \qquad q = [\tfrac{1}{3} \quad \tfrac{4}{9} \quad \tfrac{2}{9}]. \tag{4.19}$$

This chain is both lumpable and coherent with respect to partition $\mathcal{U}$ defined by

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{4.20}$$

and the transition matrix $\bar{A}$ of the aggregate chain Y_t is

$$\bar{A} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{4.21}$$

and is given in Fig. 2b. Due to coherency the optimal filter based on the observation whether X_t is in class J=1, or in class J=2, is itself a two-state Markov chain. Its transition matrix is the same as in (4.21) and its states are indicated in Fig. 2c, where the simplex $\mathcal{S}$ is the triangular sector $\pi_i \geq 0$, $\pi_1 + \pi_2 + \pi_3 = 1$.

5. Partial Balance and Decentralization

For a Markov chain X_t with invariant measure q and transition matrix A the reversed Markov chain $\hat{X}$ is defined by the transition matrix $\hat{A} = Q^{-1}A'Q$,

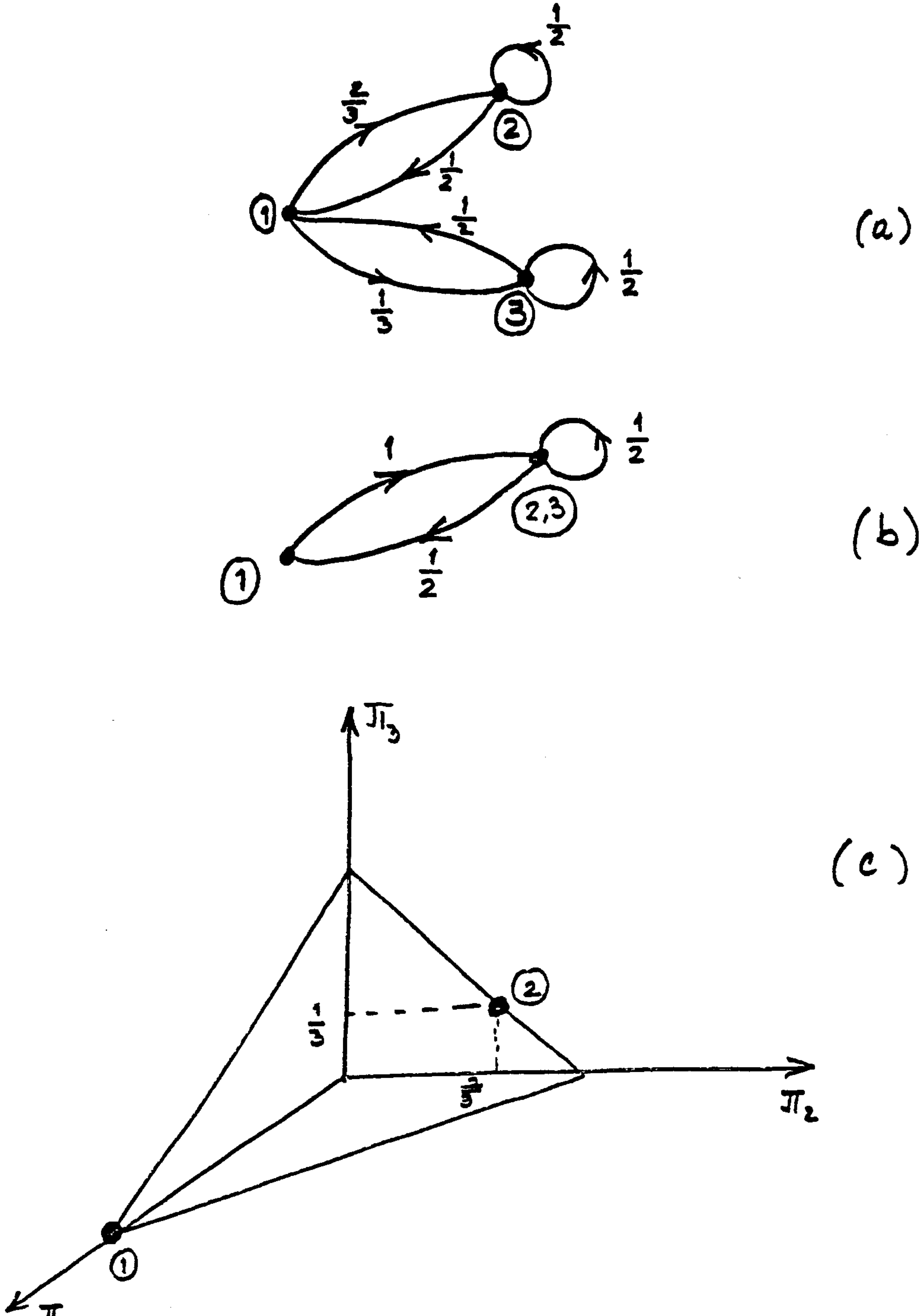

Fig. 2

where $Q = \mathrm{diag}(q_1,\ldots,q_n)$. We note that with $Q = W$ matrix $\hat{A}$ is the same as in (3.7). Chain X_t is called reversible if $A = \hat{A}$. In a reversible chain every possible path $\mathcal{C}_{ij}$ between states i and j,

$$\mathcal{C}_{ij} = [(i,j_1),(j_1,j_2),\ldots,(j_\ell,j)] \tag{5.1}$$

has the property that

$$\frac{q_i}{q_j} = \prod_{(k,k')\in\mathcal{C}_{ij}} \frac{a_{kk'}}{a_{k'k}} \tag{5.2}$$

and, hence, q can be obtained by scalar computations. A property less restrictive than reversibility of A is the so-called partial balance

$$U'QA = U'A'Q \tag{5.3}$$

which for $P = \hat{P} = UC$ is equivalent to

$$AP = \hat{A}P \tag{5.4}$$

or, in scalar form,

$$\sum_{j\in J} q_j a_{jk} = \sum_{j\in J} q_k a_{kj} = q_k \sum_{j\in J} a_{kj}, \quad \forall J,k. \tag{5.5}$$

This condition, introduced by Kelly [20], allows a decentralized computation of the invariant measure q of A:

<u>Proposition 5.1</u>: If the partial balance condition (5.3) is satisfied then $q_j = \bar{q}_J q_j^J$ where $\bar{q}_J$ is defined by

$$\bar{q}_J = \sum_{j\in J} q_j \tag{5.6}$$

which is obtained by solving the aggregate problem

$$\bar{q}\bar{A} = \bar{q}; \tag{5.7}$$

and we can find conditional probabilities

$$q_j^J = \frac{q_j}{\bar{q}_J}, \quad j \in J \tag{5.8}$$

solving N local problems

$$q_+^J A^{JJ} = q_+^J D^J, \qquad J = 1,\ldots,N \tag{5.9}$$

where A^{JJ} and q_+^J are defined as in (4.15) and

$$D^J = \mathrm{diag}(\sum_{k\in J} a_{jk},\ j\in J). \tag{5.10}$$

Although neither aggregability nor coherency have been assumed, the aggregate matrix $\bar{A} = S^{-1}U'QAU$ of (4.7) with W=Q appears in (5.7). To see that this is due only to the partial balance, we multiply (5.3) by U from the right and by the row N-vector $e_N = [1\ 1\ \ldots\ 1]$ from the left,

$$e_N SS^{-1}U'QAU = e_N U'A'QU = e_n QU = \bar{q} \tag{5.11}$$

where $e_N U'A' = e_n$, because A is a stochastic matrix.

The relationship between partial balance, aggregability, and coherency can be deduced from Corollary 3.1 and (5.4). Under the condition of partial balance, aggregability and coherency imply each other and the aggregate chain $\bar{A}$ is reversible

$$\bar{A} = \bar{Q}^{-1}\bar{A}'\ \bar{Q}. \tag{5.12}$$

On the other hand aggregability, coherency, and the reversibility of the aggregate chain imply

$$AP = PA = PAP = P\hat{A}P = \hat{A}P = \hat{P}A \tag{5.13}$$

which is the partial balance property. When applied to queuing networks, the partial balance leads to the "product form" property [21,22] which is one of the most useful, if not the only tool for computing the invariant measure of a large scale network of queues.

A further step in the decentralization is to require this property not only for an eigenvector (invariant measure) of A, but for the whole matrix. We already know that aggregability and coherency guarantee the existence of the block-diagonalizing transformation (2.7), that is the separation of the aggregate $\bar{A} = CAB$ from the residual matrix $B_\perp AC_\perp (B_\perp C_\perp)^{-1}$. We now give necessary and sufficient conditions under which the off-diagonal blocks of the residual matrix are zero.

Proposition 5.2: If A is C-aggregable and B-coherent a necessary and sufficient condition for the residual matrix to consist of N diagonal blocks is that the (L,J) block of A for $L \neq J$; $L,J = 1,\ldots,N$ be of the form

$$A^{LJ} = \bar{a}_{LJ} e_{n_L} \otimes q_+^J \tag{5.14}$$

that is

$$a_{\ell j}^{LJ} = \bar{a}_{LJ} \frac{q_j}{\bar{q}_J}, \qquad \forall \ell \in L, \quad \forall j \in J \tag{5.15}$$

where n_L is the number of the elements in the group L, and $\otimes$ is the tensor product.

Proof: Since $\tilde{P} = I-P$ is block-diagonal we have

$$(\tilde{P}A\tilde{P})^{LJ} = \tilde{P}^{LL}A^{LJ}P^{JJ} \tag{5.16}$$

and, since in view of aggregability and coherency

$$\tilde{P}A\tilde{P} = A\tilde{P} = \tilde{P}A, \tag{5.17}$$

the proof will follow if

$$\tilde{P}^{LL}A^{LJ} = A^{LJ}\tilde{P}^{JJ} = 0, \quad \forall L \neq J. \tag{5.18}$$

For (5.18) to hold A^{LJ} must be of the form

$$A^{LJ} = ke_{n_L} \otimes q_+^J \tag{5.19}$$

where k is a scalar. From coherency

$$q_+^L A^{LJ} = \bar{a}_{LJ} q_+^J \tag{5.20}$$

and hence $k = \bar{a}_{LJ}$.

5. Conclusion

We have developed a unified framework which encompasses various aggregability-lumpability conditions. The aggregation matrix, either C or B', can be chosen to define meaningful aggregate variables such as centers of intertia or probabilities for lumped states. Matrices $B_\perp$ and $C_\perp'$ express common properties shared by the members of the same group, such as coherency. These properties can be weaker than aggregability and allow aggregation for restricted sets of initial conditions.

We have shown that coherency, defined in econometrics and in power systems, also implies the existence of a finite state filter for Markov chains. Aggregability (coherency) of a chain implies coherency (aggregability) of the reversed chained. Aggregability and coherency yield a new condition for decentralized computation of the invariant measure. This condition differs from the well-known partial balance condition. A necessary and sufficient condition is also given for decomposability of the residual system into N local subsystems which are coupled through the N-dimensional aggregate.

Acknowledgment

The authors express their thanks to Professors Bruce Hajek, University of Illinois and Guy Cohen, Ecole des Mines. for helpful criticism and comments.

References

1. H. A. Simon and Ai Ando, "Aggregation of Variables in Dynamic Systems," Econometrica, Vol. 29, pp. 111-138, 1963.

2. M. Aoki, "Control of Large Scale Dynamic Systems by Aggregation," IEEE Trans. on Automatic Control, Vol. AC-13, pp. 246-253, 1968.

3. M. Aoki, "Some Approximation Methods for Estimation and Control of Large Scale Systems," IEEE Trans. on Automatic Control, Vol. AC-23, pp. 173-181, 1978.

4. E. C. Y. Tse, J. V. Medanic, and W. R. Perkins, "Generalized Hessenberg Transformations for Reduced-Order Modeling of Large-Scale Systems," Int. J. Control, Vol. 27, No. 4, pp. 493-152, 1978.

5. J. G. Kemeny and J. L. Snell, Finite Markov Chains, Van Nostrand, Princeton, N.J., 1967.

6. V. G. Gaitsgori and A. A. Pervozvanskii, "Aggregation of States in a Markov Chain with Weak Interactions," Kybernetika, pp. 91-98, May-June, 1975.

7. A. A. Pervozvanskii and V. G. Gaitsgori, Decomposition, Aggregation, and Approximate Optimization (in Russian), Nauka, Moscow, 1979.

8. P. J. Courtois, Decomposability: Queuing and Computer System Applications, Academic Press, New York, 1977.

9. F. Delebecque and J. P. Quadrat, "Optimal Control of Markov-Chains Admitting Strong and Weak Interactions," Automatica, Vol. 17, No. 2, pp. 281-296, 1981.

10. R. G. Phillips and P. V. Kokotovic, "A Singular Perturbation Approach to Modeling and Control of Markov Chains," IEEE Trans. on Automatic Control, Vol. AC-20, Oct. 1981.

11. H. T. Vantilborgh, R. L. Garner, and E. D. Lazowska, "Near-Complete D-composability of Queueing Networks with Clusters of Strongly Interacting Servers," ACM, pp. 81-92, 1980.

12. A. Jennings, "Mass Condensation and Simultaneous Iteration for Vibration Problems," Int. J. for Numerical Methods in Engineering, Vol. 5, pp. 543-552, 1973.

13. G. Strang and J. Fix, An Analysis of the Finite Element Method, Prentice-Hall, 1973.

14. R. Podmore, "Identification of Coherent Generators for Dynamic Equivalents," IEEE Trans. on Power Apparatus and Systems, Vol. PAS-79, pp. 1344-1354, 1978.

15. U. DiCaprio and R. Marconato, "Structural Coherency Conditions in Multimachine Power Systems," Proc. of IFAC 7th Triennial World Congress, Helsinki, Finland, pp. 35-45, 1978.

16. J. Lawler, R. A. Schlueter, P. Rusche, and D. L. Hackett, "Modal-Coherent Equivalents Derived from an RMS Coherency Measure," IEEE Trans. on Power Apparatua and Systems, Vol. PAS-99, pp. 1415-1425, 1980.

17. B. Avramovic, P. V. Kokotovic, J. R. Winkelman, and J. H. Chow, "Area Decomposition of Electromechanical Models of Power Systems," Automatica, Vol. 16, pp. 637-648, 1980.

18. P. V. Kokotovic, B. Avramovic, J. H. Chow, and J. R. Winkelman, "Coherency Based Decomposition and Aggregation," Automatica, Vol. 17, No. 6, 1981.

19. S. Sastry and P. Varaiya, "Coherency for Interconnected Power Systems," IEEE Trans. on Automatic Control, Vol. AC-26, pp. 218-226, 1981.

20. F. P. Kelly, Reversibility and Stochastic Networks, J. Wiley, 1979.

21. C. H. Sauer and K. M. Chandy, "Approximate Solution of Queuing Models," Computer, pp. 25-32, April 1980.

22. C. H. Sauer and K. M. Chandy, Computer Systems Performance Modeling, Prentice-Hall, 1981.

23. V. G. Gaitsgori and A. A. Pervozvanskii, "Aggregation of Linear Control Systems," Automatika i Telemekhanika, No. 8, pp. 88-95, 1980.

24. V. A. Khomyakov, "Generalization of One Proof of Convergence of a Process of Iterative Aggregation for Solving Systems of Linear Equations," Automatika i Telemekhanika, No. 7, pp. 170-173, July 1973.

25. C. Commault, "Optimal Choice of Modes of Aggregation," Automatica, Vol. 17, No. 2, pp. 397-399, 1981.

26. B. I. Vaklnetisky, L. M. Dudkin, and A. A. Ryvkin, "Iterative Aggregation-A New Approach to the Solution of Large Scale Problems," Econometrica, Vol. 47, No. 4, pp. 821-841, 1979.

27. P. Bertrand, G. Michailesco, J. M. Siret, "Sur la synthese de modeles reduits par agregation," RAIRO, Vol. 10, No. 7, pp. 105-112, July 1976.

28. W. Whitt, "Approximations of Dynamic Programs I and II," Math. of Operat. Research, Vol. 3, No. 3, August 1978; Vol. 4, No. 2, May 1979.

29. A. Brandt, "Multilevel Adaptive Computations in Fluid Dynamics," AIAA Journal, Vol. 18, No. 10, Oct. 1980.

30. W. Hackbush, "On the Convergence of Multigrid Iterations," Bertrage zur Numerischan Mathematik, Vol. 9, pp. 213-239, 1981.