DECOMPOSITIONS OF HIERARCHICAL STATE ESTIMATION STRUCTURES
Problems and Strategies

Rogério Bastos Quirino
Department of Electromechanical Technology, CEFET/MD-PR, 85884-000, Medianeira, PR, Brazil

Celso Pascoli Bottura
Department of Machines, Components, and Intelligent Systems, FEEC/UNICAMP, 13083-970, Campinas, SP, Brazil

Keywords: Kalman filters, Hierarchical Structures, Estimation Theory, Distributed Models, Hierarchical Control.

Abstract: This study has three main objectives. First, to point and discuss the principal features, advantages, and limitations of distributed state estimators. Second, to analyze structures and methodologies related to the distributed state estimation problem, with emphasis on the heterarchical one. Finally, to delineate some prospects for future investigations.

1 INTRODUCTION

To provide a complete description of a complex system and its effective control requires a great quantity and variety of sensors. Multiple sensors provide more information and hence a better and more precise understanding of a system and its operation. Multisensor systems have found applications in process control, robotics, navigation, aerospace, meteorology, manufacturing, energy generation and defense systems, among others.

A multisensor system may employ a range of different sensors, with different characteristics, to obtain information about a real system. The diverse and sometimes conflicting information obtained from multiple sensors give rise to the problem of how the information may be combined in a consistent and coherent description of the environment under observation. This is one of the problems of data fusion in multisensor networks that hence requires the elaboration of methods that establish how the information derived from a multitude of sensors can be combined, in order to obtain plausible descriptions of the observed system.

Many data fusion problems in multisensor networks involve a distributed state estimation process. The fusion of information, for example, on the multi-target tracking problem, involves two important phases: distributed state estimation – treated in this article – and data association.

Within this context, considerable attention has been given to the development of distributed and parallel versions of the Kalman filtering algorithm (Kalman,1961), known as the best unbiased linear estimator or the optimal linear estimator under Gaussian assumptions and that serves as basis for our investigation.

The motivation for the material presented in this article derives from two important aspects:
1) From the benefits and importance of the multisensor systems, particularly, the distributed systems of data fusion.
2) From the existence of a narrow gap in the literature to a modest, but constructive synthesis of distributed state estimation methods.

Two main categories of distributed state estimation architectures: hierarchical and heterarchical (Talukdar et al.,1992) are presented and appraised. This serves the purpose of explaining the advantages of heterarchical distribution. A working definition for a heterarchical system is then established and the benefits of such a system outlined. The aim is to show that, although sometimes suboptimal, this kind of distributed state

Bastos Quirino R. and Pascoli Bottura C. (2005).
DECOMPOSITIONS OF HIERARCHICAL STATE ESTIMATION STRUCTURES - Problems and Strategies.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Signal Processing, Systems Modeling and Control*, pages 124-131
DOI: 10.5220/0001165001240131

estimation structure is feasible and can lead to additional advantages, for example, for the case of considering the information space instead of the state space as a starting point for distributed structure generation (Mutambara,1995).

The problems associated with a fully connected topology, and alternative strategies for heterarchical distributed estimation are outlined.

An heterarchical distributed state estimation structure, is defined as a data processing system in which all information is processed locally and there is no centralized or coordinator processing site, in opposition to an hierarchical distributed state estimation. It consists of a network of sensor nodes, each with its own processing facility, which do not require any centralizer or coordinator module. In other words, there is no explicit hierarchy. It is important to highlight that the existence of information exchange among the levels - under centralization - and among the subsystems at the same level – under coordination - in the hierarchical structures depend on mathematical development employed in the *hierarchization of the* Kalman filter discussed in section 2.

2 FORMULATION OF THE HETERARCHICAL DISTRIBUTED STATE ESTIMATION PROBLEM

The general theory of hierarchical systems was and is continuing to be applied to control and estimation. This application involves optimization techniques – minimum variance in the Kalman approach – and concepts of hierarchical structures.

The aim is to construct state estimation architectures with different performance degrees.

In this section, we briefly treat the principles of the hierarchical state estimation theory, presented in (Chong,1979), due to its importance to the comprehension and development of distributed state estimation topologies.

In the sequence, we present and analyze the dynamics of the hierarchical structures to yield distributed state estimation methods.

2.1 Fundamentals

We intend to illustrate the estimation problem of a stochastic state vector x conditioned to the innovations from two observations y^1 and y^2.

Given the local estimates of x, $\hat{x}^1 = E(x/y^1)$, $\hat{x}^2 = E(x/y^2)$ and the error covariances matrices associated, we wish to find the global estimate $\hat{x} = E(x/y^1, y^2)$ and the corresponding error covariance matrix.

The main issue that this well-known formulation leads to the heterarchical distributed state estimation problem is the following: which dynamics of information exchanging in the decomposed structure preferably would satisfy the minimum variance criteria for the global estimate $\hat{x}$?

The necessary and sufficient conditions for the global estimation can be interpreted as follows. If $(\tilde{y}^1, \tilde{y}^2)$, innovation's subspace, and $(\hat{x}^1, \hat{x}^2)$, distributed estimation subspace, are related by invertible transformations, then $(\tilde{y}^1, \tilde{y}^2)$ and $(\hat{x}^1, \hat{x}^2)$ generate the same subspace. In general, $(\tilde{y}^1, \tilde{y}^2)$ and $(\hat{x}^1, \hat{x}^2)$ do not generate the same subspace. This prevents the optimal fusion in the Kalman sense. Therefore, the optimal fusion is possible when the projection of x on $(\tilde{y}^1, \tilde{y}^2)$ lies in the same subspace generated by $(\hat{x}^1, \hat{x}^2)$.

In the case which $\hat{x}$(*optimal*) does not lie in the same subspace generated by $\hat{x}^1$ and $\hat{x}^2$, as shown in Fig. 1, we can adopt the following approaches:

a) To construct a heterarchical distributed state estimation structure based on the exchange of local state estimates and data β resulting from a transformation $\mathrm{T}^{i\beta}$ that establishes an ap*proximate* relationship between the local subsystems and the global system. $\mathrm{T}^{i\beta}$ allows us to yield an heterarchical estimation structure in which the data β are transmitted between the subsystems at an unique level. This estimation structure would be suboptimal in the Kalman sense;

b) Alternatively, to construct an estimation structure to enlarge the subspace generated by $\hat{x}^1$, $\hat{x}^2$ and β, through the incorporation of an

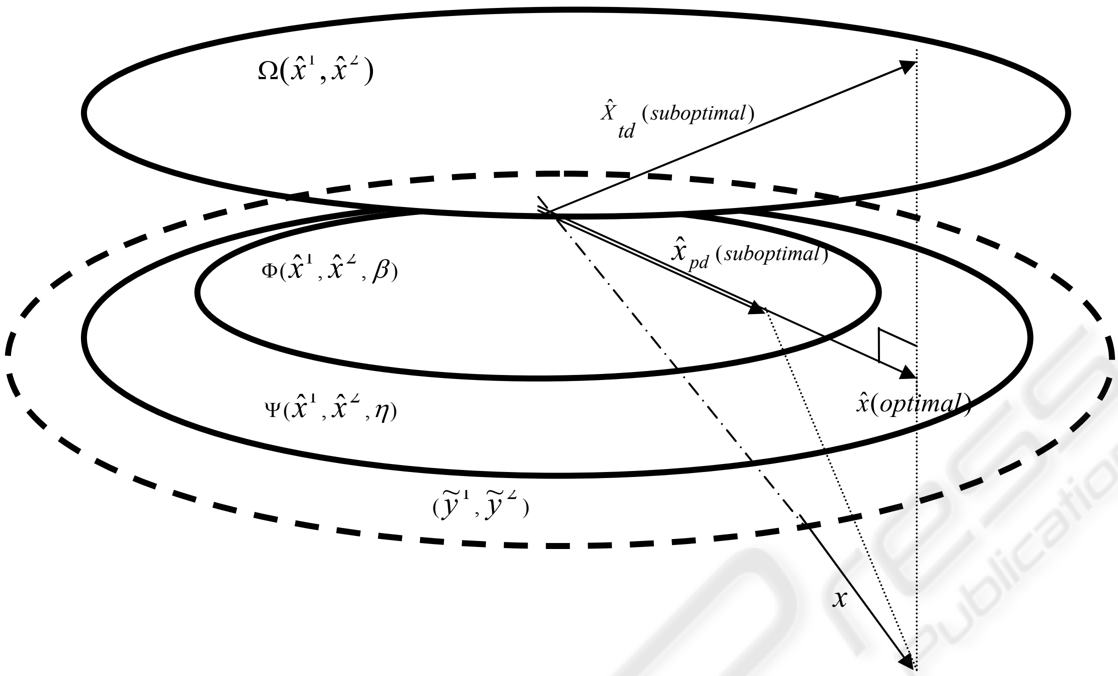

Figure 1: Geometric view of the heterarchical distributed state estimation (**Ω**- totally decoupling subspace: td ; **Φ**- partially decoupling subspace: pd ; ψ- coupling subspace)

information set $\eta > \beta$, where $\eta \supset \beta$, until $\hat{x}_{(optimal)}$ lies in this enlarged subspace. The data $\eta(\tilde{y}^1, \tilde{y}^2)$ resulting from the $\mathrm{T}^{i\eta}$ transformation establishes an exact relationship between the local and global models.

These data are transmitted between the levels of such a hierarchical estimation structure that is coordinated or centralized.

If we wish to minimize the communication in the second approach, without compromising the performance of the estimation structure, the dimension of the data vector η must be reduced as much as possible obeying the restriction $\eta > \beta$.

Once the estimation is generated, the innovation's subspace $\left(\tilde{y}^1, \tilde{y}^2\right)$ becomes equivalent to $(\hat{x}^1, \hat{x}^2, \eta)$, denoted as the optimal distributed state estimation subspace.

If an inherently global system is totally decoupled, the subspaces Ω, Ψ, and Φ constitute an unique subspace within the innovation subspace $(\tilde{y}^1, \tilde{y}^2)$. In this case, the estimation structure would be inherently heterarchical.

3 STRATEGIES OF HETERARCHICAL DISTRIBUTED STATE ESTIMATION

The strategies of decomposition in order to obtain heterarchical distributed state estimation structures can be developed for the prediction stage as well as for the correction stage of the Kalman filter, for instance. The decomposition only of one or both stages will depend on the existence of correlation between the observation and state noises of the system's model. In (Hashemipour & Laub,1987), for example, the strategies of decomposition are developed for both stages.

The problem of heterarchical distributed state estimation formulated in this work deals exactly with the decomposition of the correction stage and is based on the original version of the Kalman filter (Kalman,1961) as well as on its alternative Inverse Covariance form (Anderson & Moore,1979).

3.1 Strategies via Matrix Partitioning

These strategies are based on the Inverse Covariance of the Kalman filter (Anderson & Moore,1979).

Consider the global system model:

$$x_{k+1} = A_k x_k + w_k \tag{1}$$

where w_k is independent of x_0 assumed Gaussian with covariance P_0.

In addition, consider a set of N local observations concerning the global system (1), comprised by the following equations:

$$y_k^i = H_k^i x_k + v_k^i, i = 1,2,..., N \tag{2}$$

where the v^i , measurement noises, with covariance $R^i_{k,}$ are independent among themselves and independent of w_k and x_0.

For the heterarchical distributed state estimation problem we assume that the local processing algorithms are solved based on local models described by:

$$x_k^i = A_k^i x_k^i + w_k^i \tag{3}$$

$$z_k^i = C_k^i x_k^i + v_k^i \tag{4}$$

where i=1,2,..,N.

Consider the global system with the state x, decomposed into two subsystems with states x^1 and x^2. The local observations, y^1 and y^2 , in (1)-(2), based on the knowledge of x, provide an exact representation of the process. On the other hand, the models describing the local subsystems states, x^1 and x^2 , of the global system x, and the local observations based on knowledge of x^1 and x^2 , in (3)-(4), could provide only an *approximate* representation of the global system state x.

Alternatively, x^i might exactly represent x. In this case x^i is a Markovian process identical to x. On the other hand, it may be true that $x^i \neq x$. In this case, could exists a nodal transformation matrix T^i such that $x^i = T^i x$. If T^i does not exist, then or x^i represents a subvector not considered in the global model x, or x^i represents an *approximate* model of this global model. This *approximation* can be reached using reduced order models, derived from relaxation of part of the correlations of the global model.

The global state estimation that will be processed in a centralized node, based on the observation of the global system (2), can be written as follows:

$$\hat{x} = \bar{x} - P\sum_{i=1}^{N} H^{i^t} R^{i^{-1}} H^i \bar{x} + P\sum_{i=1}^{N} H^{i^t} R^{i^{-1}} y^i \tag{5}$$

where $\bar{x} \equiv$ prediction of x.

$P \equiv$ covariance of the estimation error of x.

If there is a transformation T^i that satisfies the relationship between the local and global dynamics such that the measurements y^i and z^i in (2) and (4) become exactly or approximately compatible, then processing at the local nodes solves the following local estimation problem:

$$H^{i^t} R^{i^{-1}} y^i = \Gamma^{i^t} P^{i^{-1}} [\hat{x}^i - (I - H^{i^{-1}} . \\ .R^{i^{-1}} H^i \Gamma^i) \bar{x}^i] \tag{6}$$

where Γ^i is the nodal transformation matrix that satisfies $\Gamma^i = C^{i^\#} . H^i$ and # denotes the pseudo-inverse.

From (5) and (6) we have:

$$\hat{x} = \Lambda \bar{x} + \sum_{i=1}^{N} G^i (\hat{x}^i - \Lambda^i \bar{x}^i) \tag{7}$$

where

$$\Lambda = I - P\sum_{i=1}^{N} H^{i^t} R^{i^{-1}} H^i;$$

$$G^i = P\Gamma^{i^t} P^{i^{-1}}; \tag{8}$$

$$\Lambda^i = I - P^i H^{i^t} R^{i^{-1}} H^i \Gamma^i;$$

It is important to point out that if at least an unique nodal transformation matrix Γ^i provides only just an approximate representation for the i-th sub-state, then the global state estimation based on the gobal reconstructibility will not be optimal in the Kalman sense.

In general, the local estimates $\hat{x}^i$ are not independent. The correlation between these estimates is taken account through the P matrix. The local correction gain G^i given in (8) incorporates the influence of these correlations in the global estimation process represented in (7).

If there is a nodal transformation T^i in (8) that transforms the global model in a feasible local model, such that P, for example, be diagonalizable, then we can construct an heterarchical and suboptimal global estimator based on a set of communicated data of dimension less than $\eta(\hat{x}^i)$, i=1,2,...,N, by *undoing the hierarchy*.

In principle, the strategies via matrix partitioning (Chong,1979) and (Hashemipour & Laub,1988), as well as the strategies via the multiple projections (Hassan et al.,1978), presented in the following subsection, require centralizer and coordinator modules, respectively, in order to fuse the local estimates in such hierarchical estimation structure.

3.2 Strategies via Sucessive Orthogonalizations

These strategies are based on the original version of the Kalman filter (Kalman,1961). In this class of strategies each local node disposes only of its local model that represents exactly a subsystem of the global system. Therefore, the construction of heterarchical distributed structures based on these strategies assumes the existence of a nodal transformation T^i, not explicit, however obvious, that satisfies exact relationships between the global system and the local subsystems.

From this assumption results the requirement of a coordinator module in order to give support to the local estimates processing.

Consider the following representations for the local models:

$$\begin{aligned} x_{k+1}^i &= A_k^i x_k^i + \sum_{\substack{i=1 \\ j\neq i}}^{N} A_k^{ij} x_k^j + w_k^i \\ y_k^i &= H_k^i x_k^i + v_k^i \end{aligned} \tag{9}$$

where the same assumptions made to the noise variables in (1) and (2) are held.

The key idea of the multiple projections method consists in the decomposition of the correction stage of the Kalman filter through the orthogonal projection of the state x^i on the observation vector of the global system. The observation vector is partitioned into N components of local observations.

In this way, the following estimation result is obtained:

$$\hat{x}^i = \bar{x}^i + \sum_{i=1}^{N} E(x^i / \tilde{y}_i^{i-1}) \tag{10}$$

where

$$\begin{aligned} &\sum_{i=1}^{N} \tilde{y}_i^{i-1} \text{ generates the Hilbert subspace:} \\ &\tilde{y}^1{}_{(k/k-1)} \oplus \tilde{y}^1_{2(k/k)} \oplus \tilde{y}^2_{3(k/k)} \oplus \ldots \\ &\oplus \tilde{y}^{N-1}_{N(k/k)}; \end{aligned} \tag{11}$$

$\bar{x}^i = E(x^i / \mathrm{Y}_{k-1})$;

$\mathrm{Y}_{k-1} \equiv$ observation subspace until the (k-1) instant.

The corrections based on the (N-1) nonlocal innovations, described by (10), constitute the coordinated hierarchical nature of the Kalman filter. In this hierarchical structure the important task of incorporating the inherent correlations among the local models, *a priori* partitioned exactly, and the global model, is made by the coordinator. In this way, the optimality of the estimation with coordinated hierarchy, in the Kalman sense, is preserved.

In (Quirino & Bottura, 2001) a nodal transformation $\mathrm{T}^{i\beta}$ on the local state is proposed, that is not explicitly *a priori* transformed by $\mathrm{T}^{i\eta}$ in (Hassan et al.,1978), in order to obviate the incorporation of the (N-1) nonlocal innovations described in (11). Such incorporation that in the original structure is proposed in (Hassan et al.,1978), it must be considered by each one of the local estimators.

4 DISCUSSIONS

The approach taken in this study for distributed state estimation is motivated by important contributions that exist in the literature. These works take the Kalman filter as the starting point to derive the parallel structures for state estimation. However, a gap resides in the fact that the great majority of the proposed structures are concentrated around hierarchical structures and do not sufficiently go beyond them in the sense we explore here.

We believe that in this work we fill part of that gap through the explicit discussion of techniques and strategies of how to undo the hierarchical structures in order to generate the heterarchical distributed structures, as here discussed.

The parallelizations of the Kalman filter equations are achieved for one or more of the different stages: 1) parallelism at the prediction stage; 2) parallelism at the correction stage; and 3) parallelism via segmentation.

Another interesting technique to parallelize the Kalman filter was developed in (Travassos,1980), in which the prediction and correction equations are simultaneously processed. The forced decoupling between these stages is maintained for the interval of one iteration during the whole history of the filter. However, the filter proposed is suboptimal, in the Kalman sense, as proved in (Hashemipour & Laub,1988), through the analysis of the estimation error covariance matrix. This technique will not be considered in this work, and remains open for future investigation.

The principal drawback of the hierarchical structures, usually resides in the fact the coordinator or centralizer module, though undone, still requires great computational and communication efforts for its implementation in a distributed environment.

Hierarchical structures present a low performance from the point of view of communication and synchronization requirements, mainly when the number of partitioned subsystems increases. The bottleneck in processing for hierarchical structures is caused by the centralizer or by the coordinator fusion of the information originating in the lower levels.

As recently commented, though these fusion modules coordinator as well as centralizer can be decomposed, e.g., by strictly computational procedures, generally, they still can generate fully connected structures with equal or great communication and computational requirements than the ones of the original structures. In addition to this, the gain achieved with respect to the communication and synchronization requirements through e.g., merely computational procedures, is not significant, as shown in (Quirino et al.,1988).

In order to minimize the effects of these restrictions we must reflect about the following question: How the proposal of partitioning the subsystems can improve the consistency of the distributed local estimates? It is because, depending on the used partitioning proposal the distributed local estimates could result from *almost purely or purely* local data implying in different performances of these distributed local estimators.

Within this context, there are controversies on the above mentioned questions: e.g., why the inherently hierarchical structures, yet do not present a good performance, if: a) The global estimate derived from local estimates can locally preprocess more data without any loss of global performance?; b) Local filtering may reduce the required bandwidth for transmission of information to a centralizer or coordinator processor?; c) For local models with dimension smaller than for the global models potential advantages can be achieved, e.g., the local processor can be made far less complex than the global processor?

Discussions about these points have been made, e.g., in (Chong,1979; Hashemipour & Laub, 1987, 1988; Hassan et al., 1978; Mutambara, 1995; Quirino et al., 1998; Quirino & Bottura, 2001; Sanders et al., 1978; Shah, 1971; Speyer, 1979; Tackers et al., 1980; and Willsky et al., 1982).

In principle, (7) and (10) can be seen as global solutions to the hierarchical state estimation problem based on the dichotomy among the information filter and the state space Kalman filter representations.

Using (7) and (10) as starting points, a synthetic diagram proposed as support to the development of distributed structures is shown in Fig. 2.

The fully connected topologies resulting from the strictly computational heterarchization of the Kalman filter, as investigated in (Mutambara, 1995) and (Quirino et al., 1998), produce optimal distributed state estimators, due either to the distribution of the coordinator task into the subsystems at the lower level in the hierarchy, class 4 of Fig. 2., or to the complete transfer of the whole coordinator task to the lower level in the hierarchy, class 3 of Fig. 2., respectively.

Such procedures of heterarchization are characterized by being merely derived from the computational distribution of the hierarchical algorithm on the distributed environment.

In spite of not providing significant gain, the distributed topologies achieved by purely computational procedures present important comparative characteristics to be analyzed and compared as scalability, communication, computation, and vulnerability to losses of communication channels.

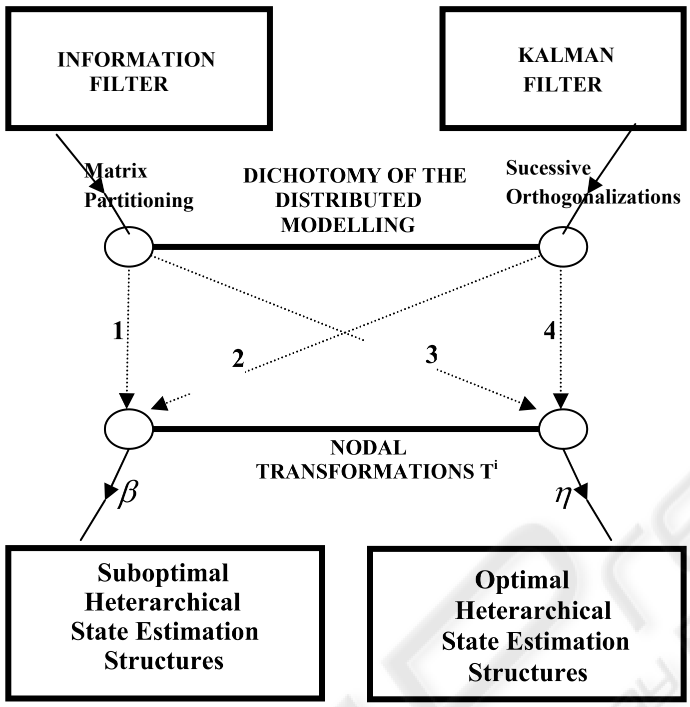

Figure 2: Classes of Strategies for Heterarchical Distributed State Estimation

Distribution strategies based on the multiple projections method as well as on matrix partitioning, lead us to face the question on which model of the local subsystems to adopt considering the global model?

For distributed filters derived via matrix partitioning and the successive orthogonalizations, the nodal transformation matrices, under certain assumptions, can be implicitly modeled in such a way that, the local estimates can be considered *very close* to the optimal estimation. This approximation transforms such structures, in principle strongly coupled into structures partially decoupled of optimal state estimation.

Strongly coupled topologies provide a very restricted practical utilization when dealing with large scale systems. Within this context, generalizations of the state space model as well as of the form how the nodal transformation matrices are obtained, are essential to develop topologies of state estimation weakly or strongly decoupled state estimation.

In (Quirino et al.,1998) an heterarchical distributed estimation topology nonhierarchical, however fully connected, is proposed.

In spite of the developed structure to be fully connected, the communication and synchronization characteristics and the efficiency of the multiprocessing system were substantially improved.

A variance another transformation, a variant of the one obtained in (Quirino et al.,1998), is under investigation using a methodology of nodal transformation proposed in (Mutambara,1995).

The development of distributed state estimation structures expliciting its correlations only in a heuristic way encumbers an enlarged understanding of the relationship among that structure and the existing ones.

Opposite to this perspective, in (Quirino & Bottura,2001) a suboptimal state estimation structure is proposed via an analytical development. This structure is conformed in the class 2 of Fig. 2. of strategies, and its development is based on the hierarchical structure proposed in (Hassan et al.,1978). This analytical development and the form of the *approximate* nodal transformation used in (Quirino & Bottura,2001) based on the SPA (Supplementing Partitioning Approach) technique proposed in (Shah,1971). Also, in (Quirino & Bottura,2001) a theorem that establishes the necessary and sufficient conditions to obtain the heterarchical distributed structure is presented, as well as for the analysis of the conditions heuristically established in (Shah,1971).

Filters conformed in the class 1 of Fig. 2., should be investigated using the same approximate representation used in (Quirino & Bottura,2001).

A study of canonical forms of the nodal transformation matrices, indispensable to formulate the distribution strategies, would be of great value to the design and analysis of distributed efficient structures for the estimation problem.

It is important to highlight that within the context of prospects for future investigations and discussion on the design of distributed state estimators, (Willsky et al.,1982) remains one of the most important references.

5 CONCLUSIONS

In addition to presenting some strategies to construct distributed state estimation algorithms proposed in the literature analyzing some aspects of them, we discuss an heterarchical distributed state estimation algorithm proposed in (Quirino & Bottura,2001).

It is our belief that the material presented in this paper contributes to the development of efficient distributed state estimation algorithms.

REFERENCES

Anderson, B. D. O. and Moore J. B., 1979. *Optimal Filtering*, New Jersey.: Prentice Hall, pp. 138-142.

Chong, C.Y., 1979. Hierarchical Estimation, *Proc.2nd MIT /ONR C3 Workshop*, Monterey, CA.

Hashemipour, R. H. and Laub, A. J., 1987. Decentralized Structures for Kalman Filtering. *IEEE Trans. on Automatic Control,* vol. 33, pp. 88–94.

Hashemipour, R. H. and Laub, A. J., 1988. On the Suboptimality of a Parallel Kalman Filter. *IEEE Trans on Automatic Control*, vol. 33, pp. 214–217.

Hassan, M. F., Salut, G., and Singh, M. G., 1978. A Decentralized Computational Algorithm for the Global Kalman Filter, *IEEE Trans. on Automatic Control,* vol. 23, pp.262-268.

Kalman, R. E., 1961. New Results in Linear Filtering and Prediction Theory, *Basic Engineering. J.*, pp. 95-107.

Mutambara, A.G.O., 1995. Decentralized estimation and Control with Application to a Modular Robot, *Ph.D. Thesis*, Oxford University, UK.

Quirino, R.B., Bottura, C.P., and Costa Filho, J.T., 1998. A Computational Structure for Parallel and Distributed Kalman Filtering, *12o. Congresso Brasileiro de Automática*, Uberlândia, MG, Brazil.

Quirino, R.B. and Bottura, C.P., 2001. An Approach For Distributed Kalman Filtering, *Revista Controle & Automação da Sociedade Brasileira de Automática*, vol. 12, pp.19-28.

Sanders, C.W., Tacker, E.C., and Linton, T.D., 1978. Specific Structures for Large Scale State Estimation Algorithms Having Information Exchange, *IEEE Trans on Automatic Control*, vol. 23, pp. 255-260.

Shah, M.M., 1971. Suboptimal Filtering Theory for Interacting Control Systems, *Ph.D..Thesis*, Cambridge University, UK.

Speyer,J.L., 1979. Computation and Transmission Requirements for a Decentralized Linear Quadratic Gaussian Control Problem, *IEEE Trans .on Automatic Control*, vol. 24, pp. 266-269.

Tacker,E.C., Sanders,C.W.,1980. Decentralized Structures for State Estimation in Large Scale Systems, *Large Scale Systems*, Vol. 1, pp. 39-49.

Travassos,Travassos, R.H., 1983. Application of Systolic Array Technology to Recursive Filtering, Ed. Thomas Kailath, New Jersey: Prentice Hall, 1983, pp.375-388.

Willsky, A.S., Bello, M.G., Castanon, D.A., Levy, B.C., Verghese, G.C., 1982. Combining and Updating of Local Estimates and Regional Maps Along the Sets of One Dimensional Tracks, *IEEE Trans. on Automatic Control,* vol. 27, pp. 799-813.

Talukdar, S., Ramesh, V.C., Quadrel, R., and Christie, R., 1992. Multiagent Organization for Real Time Operations, *Proc. of the IEEE*, vol. 80, N. 5, pp. 765-778, May.