A STUDY OF CO-OPERATIVE INVENTORY CONTROL WITH THE HELP OF DYNAMIC PROGRAMMING

Saloni Srivastava*

Dr. R.K. Shrivastava**

ABSTRACT

Communication and coordination among the retailers motivated by the possibility of sharing set up costs in the Multi-Retailer Inventory Control. In this paper we solve the problem of searching the sub optimal distributed reordering policies which minimize set up, ordering, storage and shortage costs, which is incurred by the retailers over a finite horizon. The computational complexity of the solution algorithm from exponential reduces to polynomial on the number of retailers through Neuro-Dynamic Programming.

*Department of Applied Science and Humanities, Sachdeva Institute of Technology, Farah, Mathura.

**Department of Mathematics, Agra College, Agra

1. INTRODUCTION

Consider a two echelon, one-warehouse multi-retailer inventory system. It is to be noted that each day, a stochastic demand materializes at each node and unfulfilled demand is backlogged. To fulfill the expected demand, retailers observe their own inventory level, communicate and make decisions whether to reorder or not from ware house. At each store the ordered quantities and the inventory may not exceed storage capacity. Reordering occurs by means of a single track which serves all the retailers,

and set up costs are shared among all retailer who reorders also called ACTIVE RETAILERS. This motivates a certain coordination of reordering policies. It is explained in the below figure:

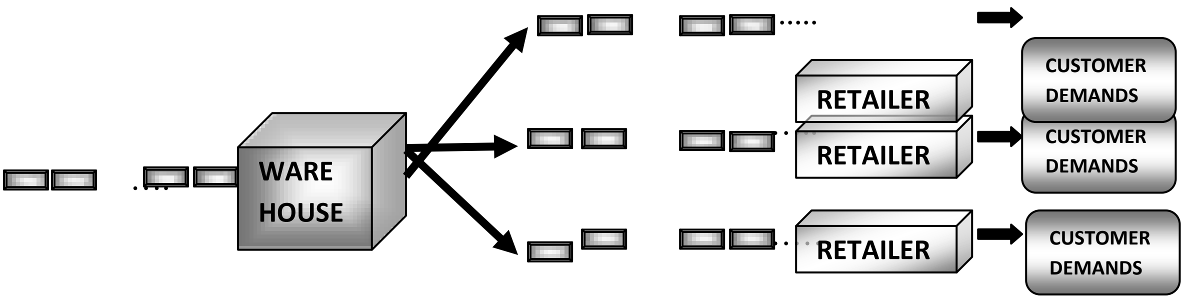

Fig. 1. One-warehouse multi-retailer inventory system

Decentralization of policies under partial information is mainly focused by Fransoo, Wouters and de Kokin 2001. Yu, Yan and Cheng in 2002 analyze the benefits of the information sharing on the performance of the entire chain. In 2001, Axsater discussed the issues, regarding the use of different kinds of penalties, transfer prices and cost sharing schemes to improve the coordination of policies optimized on a local basis. In 2003, Bauso, Giarre and Pesenti introduced a static context (for fixed day and fixed inventory levels), Saber and Murray in 2003 distributed a consensus protocol for estimating the number of active retailers and coordinating the reordering policies. Each retailer is assumed to choose a fixed threshold policy, with threshold t_k on the number of active retailers, i.e., one defines its intention to reorder only if at least other $t_k - 1$ retailers are willing to do the same.

In this paper we extend the aforementioned results to a dynamic inventory control context in which each day, the inventory levels changes. We prove that a optimal policy, for each i-th retailer, is to order only in conjunction with at least other $t_k - 1$ retailers and also that the threshold t_k can be computed locally by the k-th retailer depending on the current inventory level and expected demand, which is possible by implementing a distributed Neuro-Dynamic

Programming (NDP) algorithm polynomial on the number of retailers, which avoid the curse of dimensionality and reduces errors due to model uncertainties.

In this paper, for a cooperative inventory control problem we develop a hybrid model. Further we prove that the cost-function is k-convex and hence can be efficiently computed in a reduced number of points and also we show that threshold policies on the number of active retailers are optimal. We also introduced the NDP algorithm.

2. HYBRID MODEL

We introduced a novel hybrid model for the multi-retailer inventory system, which is explained in Figure2.

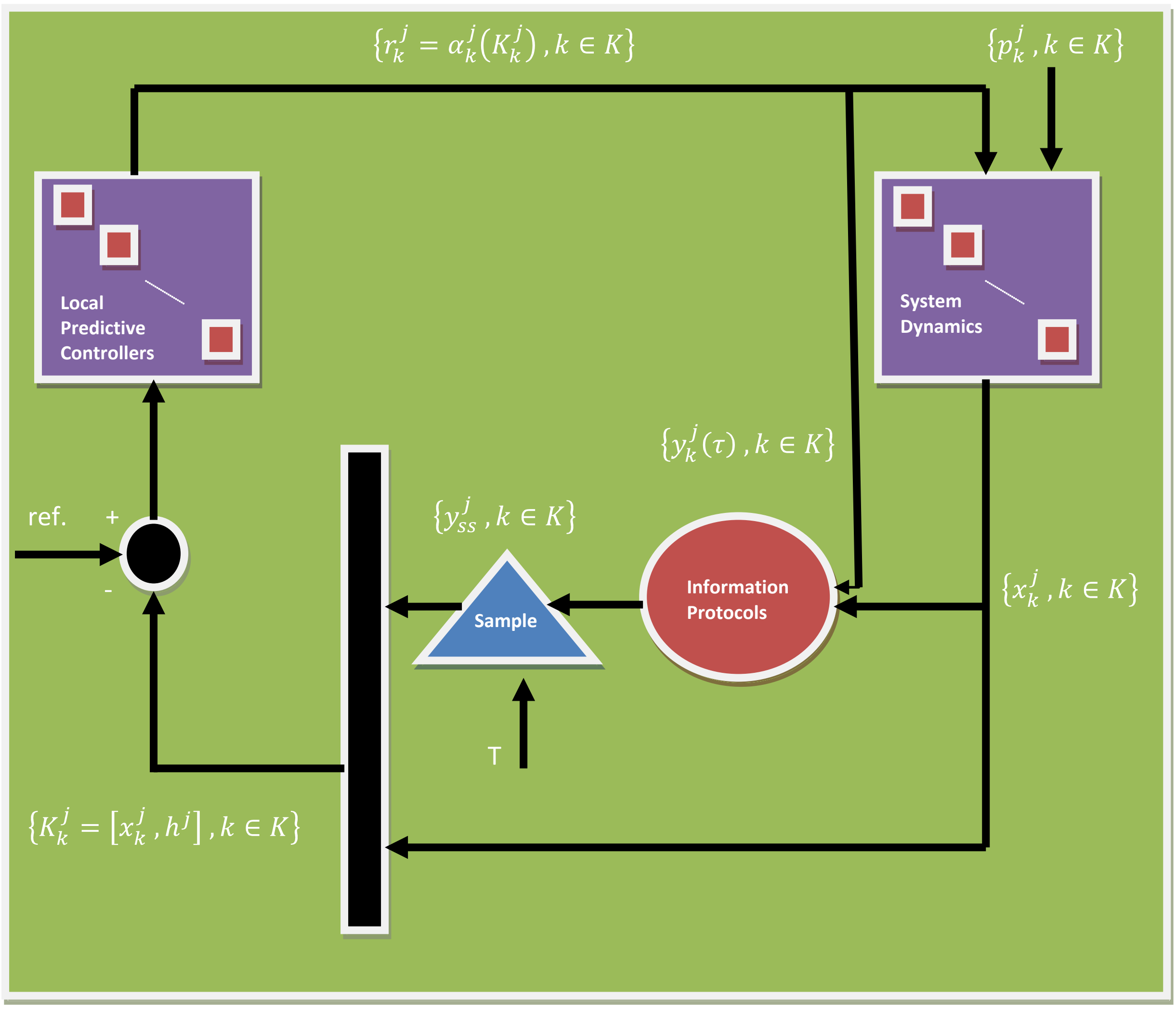

Fig 2. Block Diagram of the closed loop inventory system.

Under the SYSTEM DYNAMICS, we model the n decoupled inventory subsystems. Under the CONSENSUS PROTOCOLS, we model the information flow among the subsystems and under the LOCAL PREDICTIVE CONTROLERS; we introduce the structure of the local controllers.

2.1. SYSTEM DYNAMICS

Let us consider a network $G = (V, E)$, where each $v_i \in V$ is a node representing retailer for $i = \{1,2, \dots, n\}$ and each edge $e = (v_i, v_j) \in E$ representing communication link for $i, j = \{1,2, \dots, n\}$. Let $n = |S|$, where $|S|$ is the cardinality of the set S. The model input r_k^j is the quantity of inventory ordered by the kth retailer at each stage $j = 0,1,2, \dots \dots.., n-1$. We model with p_k^j the stochastic demand faced by the kth retailer. The kth inventory subsystem is a finite-state discrete-time model, that for all $k \in \{1,2, \dots.., n\}$ takes on the form $z_k^{j+1} = z_k^j + r_k^j - p_k^j$.

The inventory at hand plus inventory ordered may not exceed storage capacity (S_c) as is expressed in the following equation:

$z_k^j + r_k^j \leq x_k^j$, referred to as sensed information is $z_k^j = y_k^j$, which means each retailer observes only his inventory level.

2.2. CONSENSUS PROTOCOLS

Through a distributed protocol, the information flow is managed and is given as:

$$\prod = \{(\Phi_k, \Psi_k, \Upsilon_k) : \forall\, k \in V\}$$

$$y_k^j(\tau) = \Phi_k\big(x_k^j(\tau), \forall\, i \in n_k\big)\,; 0 \leq \tau \leq T \tag{1}$$

$$y_k^j(0) = \Psi_k\big(y_k^j\big) \tag{2}$$

$$h_j = \Upsilon_k\big(y_{ss}^j\big) \tag{3}$$

Where $\Phi_k : \Re^n \to \Re$, describes the dynamics of the transmitted information of the kth node as a function of the information both available at the node itself and transmitted by the other nodes eq(1). $\Psi_k : \mathbb{Z} \to \Re$ generates a new transmitted information vector given its output at the stage j eq(2). And

$\Upsilon_k : \Re \to \mathbb{Z}$ estimates, that which is based on current information and having the aggregate information eq(3).

$n_k = \{i \in (1,2, \dots.., n) : (v_i, v_j) \in \mathcal{E}\} \cup \{k\}$ is the neighborhood of the kth retailer, which means that the set of all the retailers i that are connected to k and k itself an

$$y_{ss}^j = \lim_{\tau \to T^-}\{y_k(jT + \tau)\}, \forall\, k \in \{1,2, \dots.., n\} \tag{4}$$

represents the steady state value assumed by $y_k^j(\tau)$ within the interval $[jT, (j+1)T]$. For a full state vector $z^j = [z_k^j, \forall\, k \in \{1,2,\ldots,n\}]$ the converging value of the transmitted information, h_k^j , plus the sensed information, x_k^j constitute the partial information vector, $K_k^j = [x_k^j, h_k^j]$ available to the kth retailer.

2.3. LOCAL PREDICTIVE CONTROLLER

Over a finite horizon, the local controllers compute the following cost

$$B_k(\widehat{K}_k^j, r_k^j) = E_c\left\{f_k(\widehat{K}_k^n) + \sum_{\hat{j}=j}^{n-1}\left(\delta^{\hat{j}} f_k(\widehat{K}_k^{\hat{j}}, r_k^{\hat{j}})\right)\right\} \qquad (5)$$

where $\widehat{K}_k^j$ is the predicted information and $\delta^{\hat{j}}$ is the discount factor at stage j. The stage cost $h_k(\widehat{K}_k^j, r_k^j, j)$ is defined as:

$$f_k(\widehat{K}_k^j, r_k^j, j) = \frac{J}{h^j}\beta(r_k^j) + p_c r_k^j + p_s E_c\{\max(0, -\hat{x}_k^{j+1})\} + \Psi E_c\{\max(0, \hat{x}_k^{j+1})\} \qquad (6)$$

Where J represents the set up cost, p_c is the purchase cost per unit stock, p_s is the penalty on storage, Ψ is the penalty on shortage and $\beta(r_k^j)$ is zero if the kth retailer does not reorder, and one if he reorders.

The algorithm that is used for solution is to taken due to the consideration of simulation-based tunable predicator of the form:

$$\widehat{K}_k^{j+1} = \begin{bmatrix}\hat{x}_k^{j+1}\\ \hat{h}_k^{j+1}\end{bmatrix} = \begin{bmatrix}z_k^j + r_k^j - \hat{p}_k^j\\ f_k(h_k^j, r_k^j)\end{bmatrix} \qquad (7)$$

The set up cost is equally shared among the active retailers, we assumed this in eq(6).

Under consideration we give the formalization of the problem. For a given set of reviewed retailers as dynamic agents of a network with topology G = (V, E).

Problem : (Local Controllers Synthesis) For each kth retailer, determine the reordering policy

$r_k^j = \alpha(K_k^j)$, that minimizes the n-stage individual pay off defined in (3).

Sub Problem : (Protocols Design) Determine a distributed protocol Π which maximizes the set of active retailers AR_Π .

3. DYNAMIC PROGRAMMING APPROACH

For the upcoming days, to fulfill exactly the expected demand, the inventory must be ordered in quantity. Let $J^j = \frac{J}{h^j}$ be the set up cost charged to each retailer that reorders at stage j , and $l_k^j = z_k^j + r_k^j$ be the instantaneous inventory position (the inventory level just after the order has been issued). Let us consider that if the set up cost J^j decreases with time retailers place short term orders. Optimal policies are multi period policies (s^j, S^j) , with a unique lower and upper threshold which can be easily seen in figure 3.

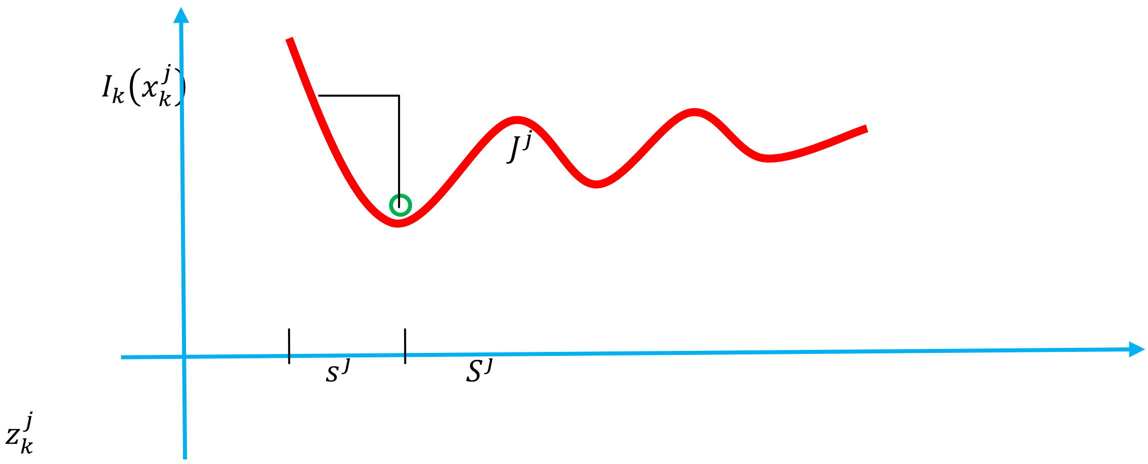

Fig. 3. Intuitive plot of the cost

On the contrary, if the set up cost J^j increases with time, retailers place long-term orders. Optimal policies are multi period policies (s^j, S^j) with multiple thresholds at different inventory levels which can be easily seen in figure 4.

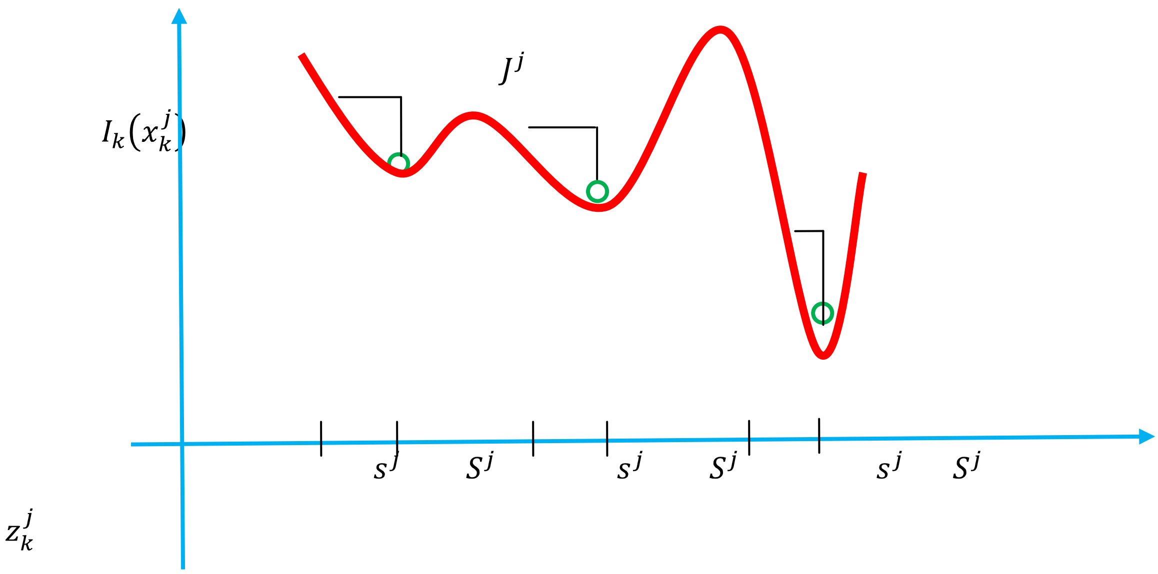

Fig. 4. Intuitive plot of the cost when the set up cost increases with time : multiple thresholds (s^j, S^j)

3.1. SEARCHING FOR STRUCTURE

We first apply the Dynamic Programming (DP) algorithm given by eq (6) and (7) to minimize the cost (5) which shows that the individual objective functions $I_i \in \{1,2,\ldots.,n\}$ have atmost n local minima. By defining a new function $F(\cdot)$ in which $\bar{J}_k$ is the maximum set up cost incurred by the kth retailer over the horizon which is obtained by rearranging the Bellman's equation eq(4) can also be arranged as:

$$f_k\left(l_k^j, h^j\right) = J^j\beta\left(r_k^j\right) + p_c r_k^j + p_s E_c\left\{max\left(0, -\left(l_k^j - p_k^j\right)\right)\right\} + \Psi E_c\left\{max\left(0, \left(l_k^j - p_k^j\right)\right)\right\}$$

With the help of DP algorithm, we get

$$B_k^n(K_k^n) = 0 \qquad (8)$$

$$B_k^j\left(K_k^j\right) = \min_{r_k^j \in U}\left\{f_k\left(l_k^j, h^j\right) + h^{j+1}E_c\left[B_k^{j+1}\left(K_k^{j+1}\right)\right]\right\} \qquad (9)$$

Consider a new function as:

$$F_k^j\left(l_k^j, h^{j+1}\right) = p_c l_k^j + E_c\left\{p_s\, max\left(0, -\left(l_k^j - p_k^j\right)\right) + \Psi\, max\left(0, l_k^j - p_k^j\right) + B_k^{j+1}\left(K_k^{j+1}\right)\right\}$$

Eq.(7) i.e., Bellman's equation can also be written as:

$$B_k^j\left(K_k^j\right) = -p_{c_k} z_k^j + \min_{l_k^j \geq z_k^j}\left\{J^j + F_k^j\left(l_k^j, h^{j+1}\right), F_k^j\left(z_k^j, h^{j+1}\right)\right\} \qquad (10)$$

It is to be noticed that Bellman's eq.(8) has a unique minimize if we can show that B_k^{j+1} is J^j-convex then F_k^j is also J^j-convex. This shows a sufficient condition that guarantees optimality of multi period $\left(s_k^j, S_k^j\right)$ order-up-to policies. As we know that s_k^j represents the minimum threshold on inventory level below which retailers reorder to restore level S_k^j and also that S_k^j minimize $F_k^j\left(\cdot, h^{j+1}\right)$ and threshold s_k^j verifies $F_k^j\left(s_k^j, h^{j+1}\right) = F_k^j\left(S_k^j, h^{j+1}\right) + J^j$

Let us consider the following:

- $\underline{s}_k$as the threshold computed if all retailers would share equally the set up costs $J_k^j = \frac{J}{n}$, namely one nth of the entire cost J.
- $\bar{s}_k^j$ as the threshold which corresponds to the assumption that the kth retailer is charged the whole set up cost; we have $J_k^j = J$; $k \in \{1,2,\ldots,n\}$.
- In the function $s_k^j\left(\frac{J}{h^j}\right)$, we explicit dependence of threshold s_k^j on set up cost J_k^j for which it holds $\underline{s}_k^j \leq s_k^j(\cdot) \leq \bar{s}_k^j$.

  We also have $F_k^j\left(l_k^j, h^j\right) = \min_{x_j^k \geq z_k^j}\left[J^j + F_k^j\left(l_k^j, h^{j+1}\right), G_k^j\left(z_k^j, h^{j+1}\right)\right]$

On considering h^j parameter, we show that the individual objective function, $I_k^j(z_k)$; $k \in \{1,2,\dots,n\}$ which is a non-convex and has all local minima coincide with the demand summed over one or more days.

Theorem 3.1.1: Solution of the Bellman's eq.(9) are at most n – j different multi period policies

(s_k^j, S_k^j), where $S_k^j \in \{\sum_{i=j}^{m} p_j^k ; m = j, j+1, \dots, n\}$ and threshold s_k^j verifies $F_k^j(s_k^j, h^{j+1}) = F_k^j(S_k^j, h^{j+1}) + J^j$. Policy are associated to different intervals of inventory levels.

Proof: Since the cost is piecewise linear, the Bellman's eq. where the cost I_k^j is the summation of a piecewise linear stage cost f_k^j having unique global minima at p_k^j and a piecewise linear future cost having potential local minima at $p_k^j + S_k^{j+1}$.

3.2. THRESHOLD REORDERING POLICIES : Now we have to show that on the number of retailers interested in reordering, Nash equilibrium reordering policies have a threshold structure. For this we have the following mentioned lemma:

Lemma 3.2.1 : (Single-Stage Optimization) ∃ a threshold $t_k \in \{1,2,\dots,n\}$, for each inventory level z_k such that the reordering policy

$$\alpha_k(K_k) = \begin{cases} S_k - z_k & if\ h \geq t_k \\ 0 & if\ h < t_k \end{cases} \tag{11}$$

is a Nash equilibrium for the single-stage formulation of the Multi-retailer Inventory Control Problem.

Proof: We have a unique multi period policy (s_k, S_k) from the theorem 3.1.1. Then the retailers make decision according to the following equation:

$$\alpha_k(K_k) = \begin{cases} S_k - z_k & if\ z_k < S_k \\ 0 & if\ z_k \geq S_k \end{cases} \tag{12}$$

To find the minimum value of t_k , for given z_k , that verifies the condition $z_k < S_k$. This is straight forward for the two limit cases of "low" inventory level $(z_k < \underline{s}_k)$ and "high" inventory level $(z_k \geq \overline{s}_k)$. This lemma can be extended to the multi-stage formulation.

Theorem 3.3.2 : (Multi-Stage Optimization) ∃ a threshold $t_k^j \in \{1,2,\dots,n\}$ for each inventory level z_k^j such that the reordering policy is

$$\alpha(K_k^j) = \begin{cases} S_k^j - z_k^j & if\ h^j \geq t_k^j \\ 0 & if\ h^j < t_k^j \end{cases} \quad (13)$$

is a Nash equilibrium for the multi-stage formulation of the Multi-retailer Inventory Control problem.

Proof : Proof for this theorem is the same as for the single-stage inventory problem in lemma 3.2.1. From theorem 3.1.1, we have at most n – j different multi period policy (s_k , S_k), each one associated to a different interval of inventory levels. After repeating the above argument for each interval, we get the proof. Optimizing the multi-retailer inventory control problem over a multi-stage horizon leads to Nash equilibrium reordering policies with threshold structure on the number of active retailers.

3.3. LOCAL ESTIMATION WITH CONSENSUS PROTOCOLS

Here we discuss the solution of the sub problem on protocol design and concentrate on consensus protocols to estimate the number of active retailersh^j. Also for a given vector $t = \{t_k\}$, collecting the optimal thresholds, each retailer makes the decision "do not reorder" if his local estimation is lower than his threshold, as expressed in eq.(13). We assume that the current estimate of the percentage of retailers who are interested in reordering is the transmitted information. At the beginning of each time interval $[jT , jT + 1]$ based on the current inventory level z_k^j , the current estimate $y_k(\cdot)$ is reinitialized to {0, 1}. Also if the corresponding threshold t_k does not exceed the network size n that is the kth inventory level is "low", then the retailer is willing to reorder, got no information yet except his observed inventory level thus he assumes that all other retailers are in the same circumstances and set $y_k^j = 1$, which signifies that today everyone is interested in reordering. On the other hand, if the inventory level is "high" i.e., if t_k exceed n then he is not willing to join the group to order and set $y_k^j = 0$, which signifies that no one is in need to reorder. Hence we have

$$y_k^j = \begin{cases} 0 & t_k(z_k^j) > n \\ 1 & otherwise \end{cases} \quad (14)$$

Here each retailer updates the estimate on-line on the basis of new estimates data received from neighbors. At any time T_k whenever the number of retailers interested in reordering, h^j, goes below his threshold, t_k the kth retailer communicates his decision to "give up" to reorder by activating an exogenous impulse signal, $\beta_k(T - T_k)$. This exogenous impulse can be activated only one time and only when the all local estimates have reached consensus on a final value. This occurs every T_Φ, where T_Φ is an estimate of the worst case possible setting time of the protocol dynamics. For a given eq.(14), the continuous-time average-consensus protocol takes on the form:

$$\Psi_k(z_k^j) = t_k(z_k^j) \leq n$$

$$\Phi_k\left(y^j(\tau)\right) = -L_k y^j(\tau) + \beta_k(T - T_k) \cdot r_k^j$$

$$\gamma\left(y_k^j(\tau)\right) = n\left(\lim_{T \to T^-} z_k^j(\tau)\right)$$

where L is the Laplacian matrix of the communication network topology; t_k is in turn the time instant where the current estimate converges to a value below the threshold; it can be defined by the following logic conditions t_k such that:

$$[t_k(z_k^j) > n] \;\; OR \;\; \left[\left(t_k(z_k^j) \leq n\right) \;\; AND \;\; (n\, z_k(T_k) < t_k) \;\; AND \;\; (T_k = k\, T_\Phi, k \in \mathbb{N})\right]$$

4. NDP SOLUTION ALGORITHM

Here we consider the hybrid model in the field of neuro-dynamic programming.

4.1 CONSENSUS ON FEATURES_h_k^j

The NDP architecture based on feature extraction displayed in figure 5.

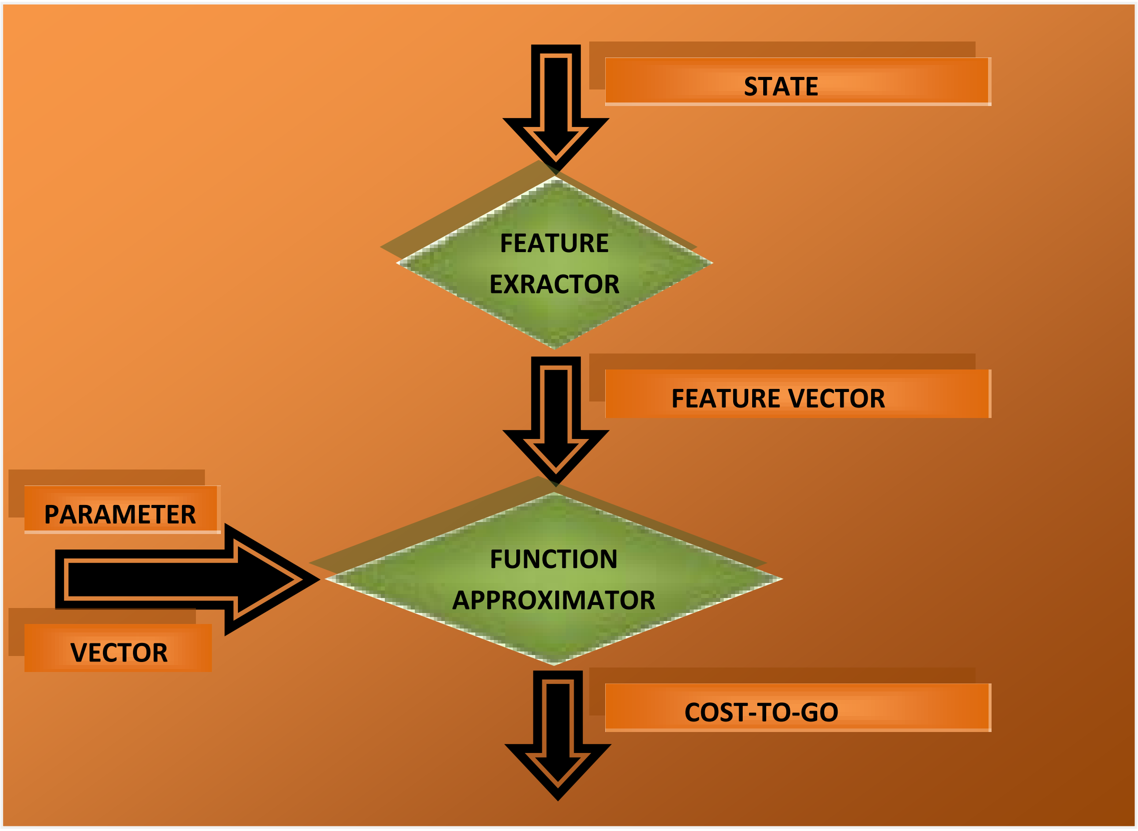

Fig. 5: The information flow management uses consensus protocols to extract the features.

The block diagram of the Hybrid Model is displayed in figure 6.

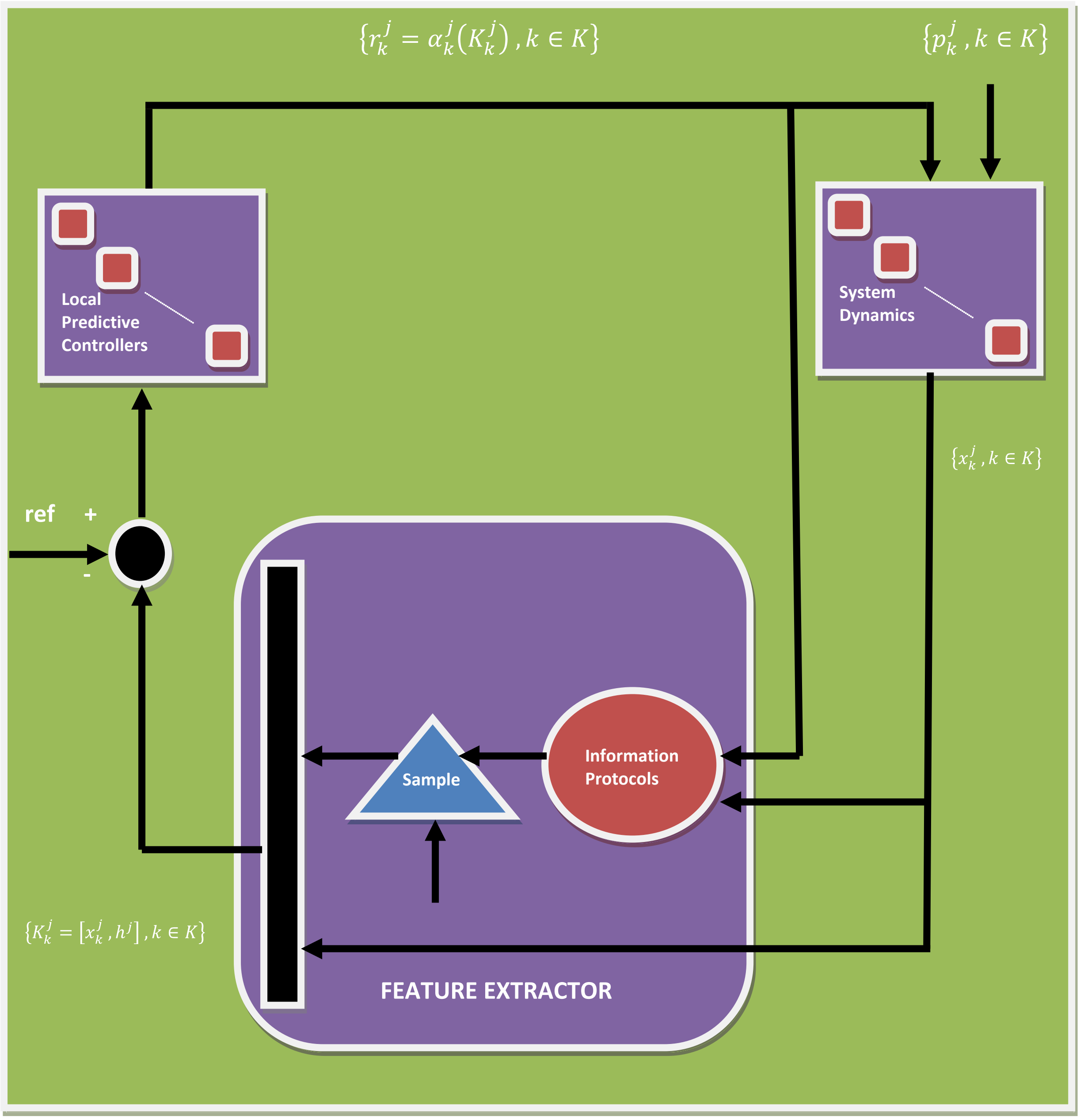

Fig. 6 : Block Diagram of the closed loop system

It is to be noticed that the full state vector of the hybrid model, z^j in the approximation architecture, the input to the feature extractor. The information flow management block can be reviewed as the feature extractor. The full state vector reduces to the partial information vector $K_k^j = [x_k^j, h_k^j]$ available to the kth retailer. Each local controller implement a function approximation, which receive the partial information vector and returns the individual cost-to-go $\tilde{I}_k^j(K_k^j, q)$ over the horizon.

4.2. LINEAR ARCHITECTURE

Let us consider that the probability distribution over all potential values assumed by h^j propogates according to the linear dynamics $h^{j+1} = h^j \chi^j$ where $\chi^j = \chi_{ki}^j\,; k, i \in \{1,2,\dots,n\}$. For this we have a matrix of weight q that coincides with the transition probability matrix of the predictor i.e. $q = \chi = \{\chi^j,\ j = 1,2,\dots,n\}$ and basis functions $\tilde{I}_i^{j+1}(K_k^{j+1}, h^{j+1})$ representing different future costs associated to different h^{j+1}. Association of all possible behavior of the other retailers over the horizon, the approximation architecture linearly parameterizes the future costs as:

$\sum_{h^{j+1}=1}^{|z|} \chi_{h^j,h^{j+1}}^j\ \tilde{I}_k^{j+1}(K_k^{j+1}, h^{j+1}) = \chi_{h^j}^j\, \hat{I}_k^{j+1}(K_k^{j+1}, \cdot)^T$, where, $\chi_{h^j}^j$ is the row of the transition probabilities from h^j to all possible h^{j+1} and $\hat{I}_k^{j+1}(K_k^{j+1}, \cdot)^T$ is the transposed row of the associated future costs.

4.3. THE NDP ALGORITHM

The NDP algorithm consists of two parts. In the first part, the retailers compute the set of admissible decisions U_k^j and reachable states V_k^j over the horizon and the second part involving three steps:

STEP 1: Policy Improvement

For a given prediction, we improve the policy with the stochastic Bellman's equation backwards in time

$$\alpha_k^j(K_k^j) = \arg \min_{r_k^j \in U_k^j(z_k^j)} \left[f_k(K_k^j, r_k^j, j) + \delta^{j+1}\, \chi_{h^j}\, \hat{I}_k^{j+!}(K_k^{j+1}, \quad \cdot)\right]$$

STEP 2 : Value Iteration

The improved policy is valued through repeated Quasi-Monte Carlo simulations. Active exploration guarantees that initial states are sufficiently spread over the local minima. During the value iteration we compute and store the number of times a transition φ_{ki} occurs during the repeated finite length simulations. At the end of each simulation, the protocol runs over the horizon and returns the training set for the next step.

STEP 3. Temporal Difference

We use the training set to update the transition probabilities of the predictor.

Above mentioned three steps are iteratively repeated to get the convergence of policies.

Lemma 4.3.1: **Each iteration of the NDP algorithm, for given initial state z^0, has computational complexity polynomial on the number of retailers, i.e., $O(n^2, N, R^2)$.**

Proof: The proof starts from considering that the complexity of the algorithm depends essentially on the complexity of the second part. Here, we write the Bellman's equation considering the set of feasible decisions U_k^j, for each retailer $k \in \{1,2,\ldots,n\}$, for each stage $j = 1, 2, \ldots, N$ and for each decomposed state $K_k^j \in \left(R_k^j \times \{1,2,\ldots,n\}\right)$. Thus the complexity is $O(n^2, N, R^2)$.

Example 1: Let us consider a group of three retailers and parameters $K = 24$, $p = 8$, $h = 1$, and $c = 2$. Retailers face a stochastic poissonian demand with expected values over the horizon of ten days as in Table I.

ω_1	**4**	**8**	**6**	**5**	**7**	**8**	**4**	**5**	**6**	**8**
ω_2	**0**	**0**	**1**	**7**	**8**	**0**	**6**	**2**	**1**	**4**
ω_3	**0**	**3**	**2**	**0**	**3**	**1**	**1**	**3**	**3**	**0**

Table 1. Expected demand for the upcoming ten days

At the first iteration, no communication has occurred among the retailers and the "policy improvement" returns the uncoordinated reordering policies displayed in Fig. 7.

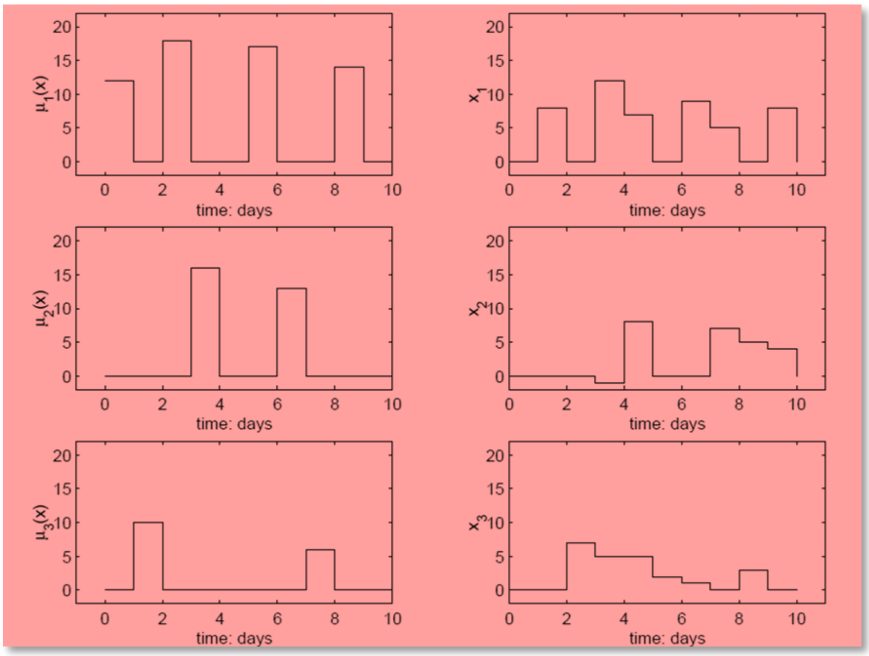

Fig. 7. Uncoordinated reordering policies

The "value iteration" consists in 12 simulations of the inventory system under the improved reordering policies. The set of initial states is a stochastic sequence extracted from a poissonian distribution with mean value respectively, equal to 25, 10, and 6 for the 1st, 2nd, and 3rd retailer. Indeed, we know from deterministic simulation results that J_1 has potential local minima at 18, 23, 30, J_2 at 1, 8, 16, and J_3 at 8, 10 as displayed in Figure 9 (solid and dotted lines). Here, the costs associated to the 1st, 2nd, 3rd and 4th policy improvements when demand is deterministic are represented by four lines of different colors (blue, red, magenta, and red). At the end of each simulation the retailers run a consensus protocol returning a^k over the horizon. Based on this new aggregate information, during the "temporal difference" the retailers update the transition probabilities of the predictor and a new iteration starts. In this example, the algorithm eventually converges to Nash equilibrium in six iterations returning a coordinated distribution of reorders

over the horizon as shown in Fig. 8. We see from Fig. 9 that the costs-to-go at the 4th and 5th iteration (green and red crosses) draw much near to the cost-to-go of the deterministic problem. We may conclude that the NDP algorithm possesses satisfying learning capabilities.

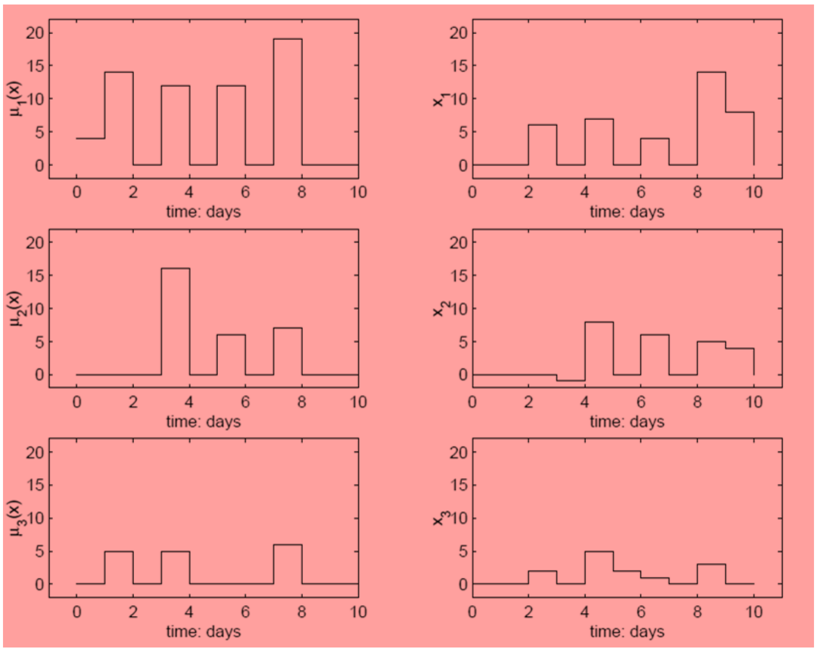

Fig. 8. Coordinated reordering reordering policies

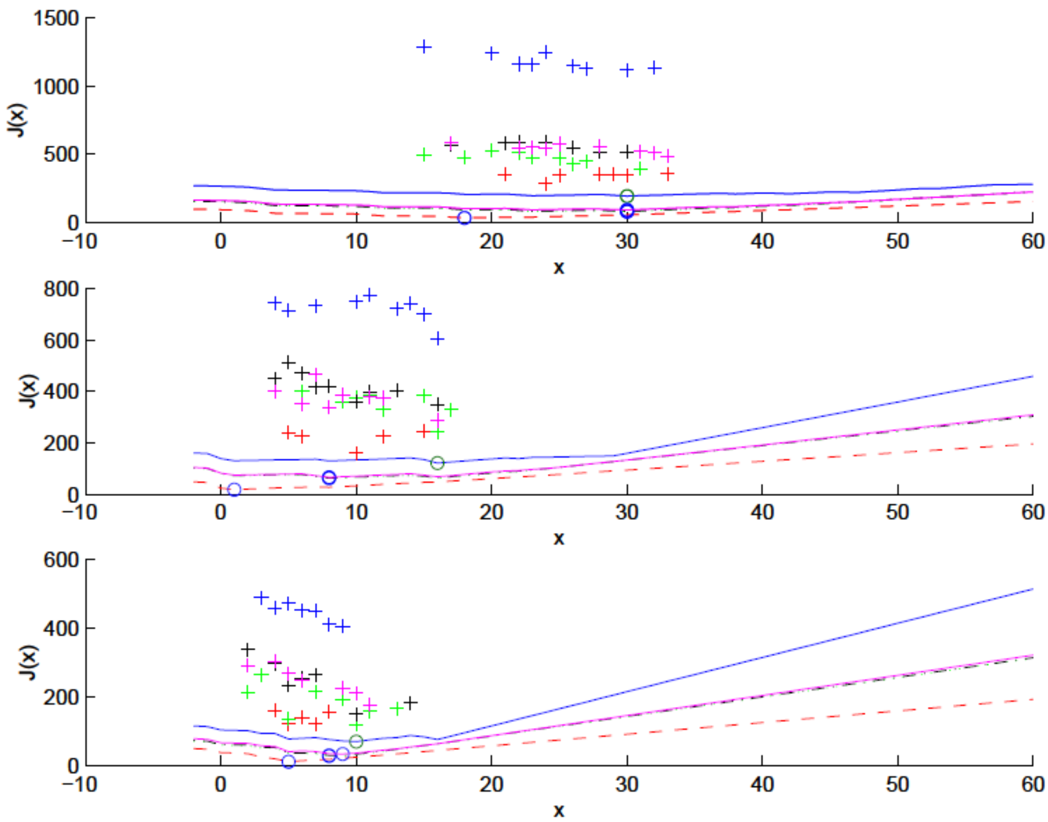

Fig. 9. Costs versus inventory : deterministic (colored lines) and stochastic demand (colored crosses)

5. RESULT

We propose an NDP approach to coordinate the reordering policies of a group of retailers. Coordination is motivated by the possibility of sharing set up cost when orders are placed in conjunction, therefore we develop a hybrid model to describe the inventory subsystems and the information flow and designed consensus protocols for the information flow to presented a scalable and suboptimal NDP algorithm.

REFERENCES

[1]. S. Axs¨*a*ter, "A framework for decentralized multi-echelon inventory control", *IIE Transactions*, vol. 33, no. 1, 2001, pp. 91-97.

[2]. D. Bauso, "Cooperative Control and Optimization: a Neuro-Dynamic Programming Approach", *Ph. D. Thesis* Universit`a di Palermo, Dipartimento di Ingegneria dell'Automazione e dei Sistemi, Dec. 2003.

[3]. D. Bauso and L. Giarr`e and R. Pesenti, “Distributed Consensus Protocols for Coordinating Buyers”, *Proc. of the IEEE Conference on Decision and Control*, Maui, Hawaii, Dec. 2003.

[4]. D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Bellmont, MA: Athena, 1995.

[5]. D. P. Bertsekas and J. N. Tsitsiklis,“Neuro-Dynamic Programming”, *Athena Scientific*, Bellmont, MA, 1996.

[6]. J. C. Fransoo, M. J. F. Wouters and T. G. de Kok, “Multiechelon multi-company inventory planning with limited information exchange”, *Journal of the Operational Research Society*, vol. 52, no. 7, Jul. 2001, pp. 830-838.

[7]. R. Olfati Saber and R. M. Murray, “Consensus Protocols for Networks of Dynamic Agents”, *Proc. of American Control Conference*, Denver, Colorado, Jun. 2003.

[8]. H. E. Scarf, “Inventory Theory”, *Operations Research*, vol.50, no.1, Jan-Feb 2002, pp.189-191.

[9]. H. E. Scarf, “The Optimality of (*s; S*) Policies in the Dynamic Inventory Problem”, *Mathematical Methods in the Social Sciences*, Stanford University Press, Stanford, CA, 1995.

[10]. Z. Yu, H. Yan and T. C. E. Cheng, “Modelling the benefits of information sharing-based partnerships in a two-level supply chain”, *Journal of the Operational Research Society*, vol. 53, no. 4, Apr. 2002, pp. 436-446.