A Novel Fuzzy Reinforcement Learning Approach in Two-Level Intelligent Control of 3-DOF Robot Manipulators

Nasser Sadati, Member, IEEE
Intelligent Systems Laboratory
Electrical Engineering Department
Sharif University of Technology
Tehran, Iran
sadati@sina.sharif.edu

Mohammad Mollaie Emamzadeh
Intelligent Systems Laboratory
Electrical Engineering Department
Sharif University of Technology
Tehran, Iran
molaie@ee.sharif.edu

Abstract – In this paper, a fuzzy coordination method based on Interaction Prediction Principle (IPP) and Reinforcement Learning is presented for the optimal control of robot manipulators with three degrees-of-freedom. For this purpose, the robot manipulator is considered as a two-level large-scale system where in the first level, the robot manipulator is decomposed into several subsystems. In the second level, a fuzzy interaction prediction system is introduced for coordination of the overall system where a critic vector is also used for evaluating its performance. The simulation results on using the proposed novel approach for optimal control of robot manipulators show its effectiveness and superiority in comparison with the centralized optimization methods.

I. INTRODUCTION

The optimal control is one of the most important topics in control theory and optimization of large-scale systems. The problems such as complexity, high dimensionality of variables, geographical separation of subsystems, etc., usually are the burdens for solving the overall problem in a centralized fashion. That is way during past three decades, many approaches such as coordination strategies in multi-level systems and decentralized schemes have been proposed by researchers.

In centralized methods, the system is decomposed into several subsystems where their optimization only depends on local variables while the effects and interactions of other subsystems are either ignored, or considered resulting in robust decentralized sub-optimal control schemes.

In the coordination methods, similar to the decentralized approaches, the system is first decomposed into several subsystems, while the effects among them are compensated through a coordinator. In this approach, the control of large-scale systems is done by using the hierarchical multi-level control scheme. So hierarchical multi-level control is a common approach that has been presented as an important and efficient method in control of large scale systems.

The basic principle of hierarchical control is decomposition of a given large-scale system into several smaller scale systems and then coordination of the resulted sub-systems to reach the optimum solution. In an attempt for improving this strategy, Mesarovic et al. presented one of the earliest formal quantitative treatments of hierarchical systems by postulating two coordination principles; Interaction Prediction Principle (IPP) and Interaction Balance Principle (IBP) [1],[2], where the coordination of large-scale systems are mainly based on these two principles.

In [3]-[6], using these two principles (IPP & IBP), two new gradient based coordination schemes are introduced that have much faster convergence rates than the classical methods. In [7],[8], two new neuro-fuzzy reinforcement strategies are introduced for intelligent coordination of large-scale systems based on IPP and IBP, where critic vectors are used for evaluation of their performances. In [9]-[11], using the new gradient based coordination schemes, the optimal control of robot manipulators have also been considered. In this paper, by using the novel strategy [7], the optimal control of robot manipulators is investigated. The simulation results are also presented.

In section II, the problem formulation and control problems are defined. In section III, the dynamic model of a robot arm with three degrees-of-freedom has been formulated. In section IV, decomposition of the overall problem, into \( m \) sub-problems and modeling the corresponding subsystems is done. In section V, the optimization of the first level subsystems are done using the gradient method. In section VI, the proposed fuzzy system is introduced to predict the change of interactions while a critic vector is used to evaluate its performance. In section VII, the proposed fuzzy interaction prediction system has been applied to optimal control of the robot manipulator and the obtained results are compared with the centralized optimal control approach. Finally, in section VIII, some concluding remarks are briefly discussed.

II. STATEMENT OF THE PROBLEM

Let there be given an overall process \( P : X \times U \rightarrow Y \) and a performance function \( G : U \times Y \rightarrow V \) with \( U \) as a set of controls, \( X \) as a set of states, \( Y \) as a set of outputs and \( V \) as a set of performance values. Let also \( g \) be defined on \( U \) by the following equation

\[
g(U) = G(U, P(U))
\]

The goal of the overall control problem which is denoted by \( D \), is to find a control action \( \hat{U} \) which minimizes \( g \) over \( U \). Such a control action will be referred to as the overall optimum.

Let \( U = U_1 \times \cdots \times U_m \), \( Y = Y_1 \times \cdots \times Y_m \), and \( X = X_1 \times \cdots \times X_m \). For each \( i = 1, 2, \ldots, m \), the \( i \)-th subsystem is given by
$P_i: U_i \times Z_i \rightarrow Y_i$, where $Z_i$ is the set of interactions of the other subsystems. Now the $i$-th initial control problem can be formulated in terms of an objective function $g_i$, given on $U_i \times Z_i$, in terms of the $i$-th subsystem. Also a performance function $G_i: U_i \times Z_i \times Y_i \rightarrow V$ can be given by the following equation

$$g_i([U_i, Z_i], p_i([U_i, Z_i]))$$

(2)

Let $Z_p$ in $Z=Z_1 \times \cdots \times Z_m$ be the predicted input for interference subsystems that is formulated by $P_i (Z_p) = P_i ([U_i, Z_p])$; $1 \leq i \leq m$. Then for each $Z_p$ in $Z$, the initial control problem $D_i[Z_p]$ is to find a control $U_i$ in $U_i$ such that

$$g_i([U_i, Z_p]) = \min_{U_i} g_i([U_i, Z_p])$$

(3)

and the minimization is only over the set $U_i$ of local controls.

Let $Z_p = (Z_1, \ldots, Z_m)$ be the predicted interference inputs and let $Z = (Z_1, \ldots, Z_m)$ be the actual interference inputs occurring when the sub-optimal control $\hat{U}_i([Z_p]) = \hat{U}_i([Z_p]) \cdots \hat{U}_i([Z_p])$ is implemented. The overall optimum is then achieved if the predicted interference inputs are correct (i.e. $e_i = 0$ where $e_i = Z_i - Z_p$ and $1 \leq i \leq m$). Alternatively, if $e_i$ can not be made to be zero, the supremal control problem can be defined as minimization of an appropriate function of the errors $e_i, e_1, \ldots, e_m$ in the second level.

In Fig. 1, the block diagram of the coordination of two subsystems using the interaction prediction principle is shown.

![Fig. 1. Coordination of two subsystems using the interaction prediction principle.](image)

**III. ROBOT MODELING**

The dynamic model of the robot arm with three degrees-of-freedom is shown in Fig. 2.

![Fig. 2. The robot arm with three degrees-of-freedom.](image)

The dynamic model of the robot arm can be obtained by the Lagrangian method as follows

$$M(q) \ddot{q} + H(q, \dot{q}) + G(q) = \tau$$

(4a)

where $q_i$ is the angle of joint $i$, $M(q)$ is the symmetric and positive definite inertia matrix, $H(q, \dot{q})$ is the coriolis and centrifugal vector, $G(q)$ is the gravity vector and $\tau$ is the torque vector.

Now using the following definitions

$$p_1 := l_1 + m_2 l_2 + m_2 l_2^2 + m_2 l_2^3, \quad c_{q_1} := \cos (q_1)$$

$$p_2 := l_2 + m_2 l_2^2 + m_2 l_2^3, \quad c_{q_2} := \cos (q_2 + q_1)$$

$$p_3 := m_2 l_2^3 + m_2 l_2^4, \quad s_{q_3} := \sin (q_3 + q_2 + q_1)$$

$$p_4 := m_2 l_2^3 + m_2 l_2^4, \quad c_{q_4} := \cos (q_4 + q_3 + q_2 + q_1)$$

the parameters of equation (4a) can be written in the following forms

$$M(q) := \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, \quad H = \begin{bmatrix} -h_2 (p_1^2 + 2p_2^2) - h_3 (p_3^2 + 2p_4^2) + 2q_1 q_2 \\ h_2 (p_3^2 + 2p_4^2) - h_3 (p_1^2 + 2p_2^2) + 2q_2 q_3 \\ h_3 q_2 q_3 + 2q_2 q_3 + 2q_2 q_3 \end{bmatrix}$$

(4b)

IV. DECOMPOSITION OF THE ROBOT MODEL

For the purpose of decomposition, here the system of robot arm with three degrees-of-freedom is decomposed into three subsystems where each joint is assumed as one subsystem. This is given below

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}, \quad \tau = \tau_1 + \tau_2 + \tau_3$$

(5a)

and $F_i$ is continuously double differential analytical function of the $i$-th subsystem, and $X_i, U_i$ and $Z_i : i = 1, 2, 3$ are defined as follows

$$\begin{bmatrix} X_i[k] \\ Z_i[k] \end{bmatrix} = \begin{bmatrix} x_i[k] \\ z_i[k] \end{bmatrix}, \quad U_i[k] = u_i[k] = \tau_i[k]$$

(5b)

\[
X[k] = \begin{bmatrix}
    x_1[k] \\
    x_2[k] \\
    \vdots \\
    x_m[k]
\end{bmatrix},
\quad
U[k] = u_1[k] + r_1[k]
\]

where \( X \) is the state vector, \( U \) is the control action and \( F \) is continuously double differential analytical function. Also the initial state \( x_0 \) is assumed to be known.

Now, the control problem is to find \( U \) which minimizes the cost function given by the following equation

\[
J(x, U) = g_{i+1}(x_0 + 1) + \sum_{i=1}^{m} g_i(x_i, U_i)
\]

(7)

By decomposition of the overall system into \( m \) subsystems, each subsystem can be described by non-linear state space equation of the following form

\[
\begin{align*}
    x_{i}[k+1] &= f_i(x_i[k], U_i[k], Z_i[k]) \\
    x_i[k] &= X_{io}, \quad 1 \leq i \leq m
\end{align*}
\]

The cost function can also be decomposed as follows

\[
J(x, U) = \sum_{i=1}^{m} J_i(x_i, U_i, Z_i)
\]

(9a)

where

\[
J_i(x, U) = g_{i+1}(x_i[k+1], Z_i[k+1]) + \sum_{k=0}^{\infty} g_i(x_i, U_i[k], Z_i[k])
\]

(9b)

Now, the necessary conditions for optimality of each subsystem can be written in terms of the Lagrangian \( L_i \), which is in the following form

\[
L_i = L_i + \frac{\partial G_i}{\partial X_i} (x_i[k+1], Z_i[k]) + \sum_{k=0}^{\infty} \frac{\partial G_i}{\partial U_i}[x_i[k], U_i[k], Z_i[k]] + \lambda[k]
\]

(10)

So the first order necessary conditions become

\[
\begin{align*}
    \frac{\partial L_i}{\partial x_i}[x_i[k+1], Z_i[k]] & = 0 \\
    \frac{\partial L_i}{\partial U_i}[x_i[k+1], Z_i[k]] & = \lambda[k]
\end{align*}
\]

(11)

VI. FUZZY PREDICTION

The prediction of the change of interactions will be done using a fuzzy system. For evaluating its performance, a critic vector is used as shown in Fig. 3, to develop a method of training for this fuzzy system. The proposed training approach is based on minimizing the energy of the critic vector. In this reinforcement approach, the error of prediction and its rate of changes are used in such a way that they increase the speed of convergence of the algorithm.

A. Designing the Critic Vector

The critic vector is defined in the following forms

\[
e_i[k] = R(e_i[k], d_i[k]) \quad 1 \leq i \leq m, \quad 0 \leq k \leq n
\]

(16)

where

\[
d_i[k] = d_i[k] - d_i[k-1]
\]

(17)

\[
e_i[k] = Z_i[k] - Z_{Ri}[k]
\]

(18)

\[
Z_i = [z_i, \ldots, Z_i, \ldots, Z_m]^T, \quad Z_i = [z_i[k], \ldots, Z_i[k], \ldots, Z_m[k]]
\]

(19)

also \( Z_i \) is the real interactions and \( Z_{Ri} \) is the predicted value given by the coordinator. Now, the critic vector is assumed to be formulated as follows

\[
e_i[k] = R(e_i[k], d_i[k]) = e_i[k]^2 + \lambda d_i[k]^2
\]

(20)

where \( \lambda \) is a positive constant parameter \( \lambda \leq 0 \leq \lambda \leq 1 \).

B. Designing the Fuzzy Prediction System

The predicted values for the interactions are assumed to be updated in the following form

\[
Z_{Ri}[k] = Z_{Ri}[k-1] + \Delta Z_{Ri}[k]
\]

(21)

To predict the changes of interactions, a fuzzy system based on fuzzy inference is used that denotes the variation of the predicted values as follows

\[
\Delta Z_i[k] = s_i[k] = s(e_i[k], d_i[k])
\]

(22)

where \( s \) is a fuzzy system based on the Takagi-Sugeno-Kang (TSK) model, that also has been used extensively in fuzzy...
modeling, and in this case, is defined by the following fuzzy sets and rules:

\[
\begin{align*}
&\text{if } e[k]\text{ is } A_1 \text{ and } d[k]\text{ is } B_1 \text{ then } s_1 = a_1 \cdot e[k] + b_1 \cdot d[k] + c_1 \\
&\text{if } e[k]\text{ is } A_2 \text{ and } d[k]\text{ is } B_2 \text{ then } s_2 = a_2 \cdot e[k] + b_2 \cdot d[k] + c_2
\end{align*}
\]  

(23)

In this paper, \( a_i \) and \( b_i \) are assumed to be zero. So the fuzzy sets and rules can be obtained easily, in the following forms:

\[
\begin{align*}
&\text{if } e[k]\text{ is } A_1 \text{ and } d[k]\text{ is } B_1 \text{ then } s_1 = c_1 \\
&\text{if } e[k]\text{ is } A_N \text{ and } d[k]\text{ is } B_N \text{ then } s_N = c_N
\end{align*}
\]  

(24)

where \( \mu_{A_i} \) and \( \mu_{B_i} \) are the membership functions of \( A_i \) and \( B_i \), respectively. Also \( c_j \) are the real constant parameters.

Now to summarize, the relation of \( s_i[k] \) with \( e[k] \) and \( d[k] \) can be given by the following fuzzy inference:

\[ s_i[k] = \sum_{j=1}^n \alpha_{ij}[e[k], d[k]] \cdot c_j \]

(25)

C. Training the Fuzzy Predictor

The goal of training is to minimize the energy of a critic vector related to the system parameters \( c_j \), where

\[ E = \frac{1}{2} \sum_{i=1}^n \alpha_{ij}^T \cdot r[k] \]

(26)

Also

\[ r[k] = \begin{bmatrix} r_1[k] \\ r_2[k] \\ \vdots \\ r_m[k] \end{bmatrix} \]

(27)

Now to update the parameters of the proposed fuzzy system, the following updating rule can be used:

\[ \Delta c_j = -\eta \sum_{i=1}^n \frac{\partial E}{\partial c_j} \cdot r_i[k] = -\eta \sum_{i=1}^n \frac{\partial E}{\partial c_j} \cdot r_i[k] \]

(28)

where \( \eta \) is a step size which has an effect on the rate of training and convergence of the fuzzy system parameters, during learning. Now, by using the chain rule, we have

\[ \frac{\partial E}{\partial c_j} = \frac{\partial E}{\partial r_i[k]} \cdot \frac{\partial r_i[k]}{\partial c_j} \]

(29)

To calculate the right side of this equation, we need to calculate; \( \frac{\partial r_i[k]}{\partial c_j} \) and \( \frac{\partial E}{\partial c_j} \). Now using (20), we can write

\[ \frac{\partial r_i[k]}{\partial c_j} = 2 \cdot e_i[k] \quad \text{and} \quad \frac{\partial E}{\partial c_j} = 2 \cdot d_i[k] \]

(30)

Moreover, the gradient of the prediction errors related to the system parameters can be given by

\[ \frac{\partial E}{\partial c_j} = \frac{\partial E}{\partial Z_p[k]} \cdot \frac{\partial Z_p[k]}{\partial c_j} = \frac{\partial E}{\partial Z_p[k]} \cdot \frac{\partial Z_p[k]}{\partial c_j} = \frac{\partial E}{\partial Z_p[k]} : = \frac{\partial E}{\partial c_j} \]

(31)

Now, in order to calculate \( T_i[j] \), using (22) and (25), we have

\[ T_i[j] = -\frac{\partial Z_p[k]}{\partial c_j} = \frac{\mu_{A_i}(e[k]) \cdot \mu_{B_i}(d[k])}{\sum_{j=1}^n \alpha_{ij} \cdot \mu_{A_i}(e[k]) \cdot \mu_{B_i}(d[k])} \]

(32)

Also from (17) and (31), we can formulate \( \frac{\partial E}{\partial c_j} \) as follows:

\[ \frac{\partial E}{\partial c_j} = \frac{\partial d_i[k]}{\partial c_j} + \frac{\partial E}{\partial c_j} \]

(33)

So the training of the fuzzy predictor can be formulated in the following form:

\[ c_j^{t+1} = c_j^{t} + \Delta c_j \]

(34a)

where

\[ \Delta c_j = -\eta \sum_{i=1}^n \frac{\partial E}{\partial c_j} \cdot r_i[k] = -\eta \sum_{i=1}^n \frac{\partial E}{\partial c_j} \cdot r_i[k] \]

(34b)

\[ \frac{\partial E}{\partial c_j} = 2 \cdot e_i[k] \cdot T_i[j] \cdot k \]

(34c)

VII. SIMULATION RESULTS

The proposed method is used to obtain the optimal control of a 3DOF robot manipulator. The parameters of the robot model are assumed to have the following values [12]

| Table 1. The parameters of the robot manipulator. |
|---|---|---|
| Mass (m_kg) | Length (l_m) | Moment of Inertia (l_kg.m^2) |
| Joint 1 | 1.2 | 0.5 | 0.25 |
| Joint 2 | 1.5 | 0.4 | 0.25 |
| Joint 3 | 3.0 | 0.3 | 0.15 |

| m_{load} = m_{end} = 0 Kg |

m_{load} = 0.01 Kg |

g = 9.8 N/Kg

Now using (9), the cost function can be written as:

\[ G_{ax}(\lambda) \cdot \frac{1}{2} \left[ \left( x_{i+1} - x_i \right) \cdot \left( x_{i+1} - x_i \right) \right] + k \cdot \left( x_{i+1} - x_i \right) \cdot \left( x_{i+1} - x_i \right) \]

(35)

where \( k_{ax} \), \( k_a \), and \( k \) are the parameters of the cost function and defined by the conditions and limitations of the problem.

In the simulation presented in this paper, they are defined to be:

| Table 2. The parameters of the cost function. |
|---|---|---|---|---|---|---|
| \( k_{ax} \) | \( k_a \) | \( k_{ax} \) | \( k_a \) | \( T \) | \( x_0 \) | \( x_f \) |
| 10^{-4} | 1.5 | 12 | 10^{-4} | 0 | 0.5 | 5 |

In critic vector, \( \lambda \) is assumed to be 0.2. The fuzzy membership functions of \( \mu_{A_i} \) or \( \mu_{B_i} \) are experimentally defined by three triangular membership functions as shown in Figure 4. By this number of membership functions, the fuzzy system has \( N = 3^2 = 9 \) rules.
The proposed approach is applied to the robot manipulator to minimize the cost function defined by (9) and (35), where the desired target trajectories are assumed to be in three sinusoid forms and the initial conditions are assumed to be zero [12].

The simulations are performed for two values: \( k_u = 0 \) and \( k_u = [1.5 \ 3 \ 12] \times 10^{-3} \), after 70 iterations of coordination, where in each iteration the optimization of each subsystem is performed using the gradient method in 200 steps. For \( k_u = 0 \), the goal of minimization of cost function defined by (9) and (35) is only minimization of an energy function of tracking errors, so there are no limitations on the norm of the input actions. But for \( k_u = [1.5 \ 3 \ 12] \times 10^{-3} \), the goal of minimization of cost function is minimization of both energy functions of tracking errors and input actions.

In Fig. 5, the calculated optimal control actions are presented for both values: \( k_u = 0 \) and \( k_u = [1.5 \ 3 \ 12] \times 10^{-3} \). As shown in this figure, for \( k_u = [1.5 \ 3 \ 12] \times 10^{-3} \), the obtained control actions have smaller values (norm) than the obtained control actions corresponding to \( k_u = 0 \). Also in Fig. 6, the obtained optimal trajectories of all joint angles are shown where for \( k_u = 0 \), the optimized trajectories of all joint angles have tracked the target trajectories with acceptable tracking errors. In contrast to obtained results using \( k_u = 0 \), the optimized trajectories of joint angles using \( k_u = [1.5 \ 3 \ 12] \times 10^{-3} \) have not be able to track the target trajectories with acceptable tracking errors.

In Figs. 7, 8 and 9, respectively, the sum-squared interaction prediction errors of control actions \( [\tau_1(k) \ \tau_2(k) \ \tau_3(k)] \), joint angles \( [q_1(k) \ q_2(k) \ q_3(k)] \), and the joint angular velocities \( [\dot{q}_1(k) \ \dot{q}_2(k) \ \dot{q}_3(k)] : 1 \leq k \leq 50 \) are shown. These figures present the convergence of the interactions predicted by fuzzy system in 50 iterations of coordination.

The simulation results of both methods; the centralized optimization approach (in this approach, the whole problem is solved in one shot, using a typical gradient optimization method), and the proposed strategy, are plotted together in Figs. 10 and 11. The centralized gradient based optimization is performed in 10000 steps for optimization of 150 variables; \( U_1[1] \ldots U_1[50] \), \( U_2[1] \ldots U_2[50] \) and \( U_3[1] \ldots U_3[50] \). In the proposed strategy, the coordination is done in 50 iterations and in each iteration, the optimization of each subsystem is performed in 200 steps, where each subsystem has 50 variables: \( U_i[1] \ldots U_i[50] : i = 1, 2, 3 \). As is shown in Figs. 10 and 11, the proposed approach has faster convergence rate rater than the centralized optimization approach.
Fig. 9. The sum-squares of the interaction prediction errors of the joint angular velocities.

Fig. 10. The optimized cost function with $\dot{k}_u = 0$, using two approaches: the centralized optimization, and the proposed fuzzy coordination method.

Fig. 11. The optimized cost function with $k_u = 1.5 \times 10^{-3}$, using two approaches; the centralized optimization, and the proposed fuzzy coordination method.

VIII. CONCLUSION

In this paper, a new approach for optimal control of robot manipulators, considered as two-level large-scale systems, is presented. It is shown how the specific feature of intelligent control including decision making, adaptation, planning and other characteristics of fuzzy logic and classical optimization methods, could help to design a unified hybrid approach for intelligent control of robot manipulators.

The fuzzy interaction prediction system is used to coordinate the overall system and also resolve the weakness of conventional coordination approaches in convergence. This fuzzy coordinator uses a critic vector to evaluate its own operation. The fuzzy coordinator learns its dynamics (parameters) through the minimization of an energy function relating to the critic vector. The minimization process, to train the fuzzy coordinator, is done by using the gradient of the energy function. The training of fuzzy predictor can be done in an off-line or on-line manner. Here, the parameters of fuzzy predictor are first calculated by an off-line manner and then, they are updated in each iteration of coordination in an on-line manner. It should be noted that in addition to the convergence of prediction, both critic vector and also the fuzzy interaction prediction system use the error and its variations, to predict the change of interactions for coordination of the overall system.

The obtained results present the performance and convergence of the proposed approach. By simulation of a robot manipulator, the superiority of the proposed method in comparison to the centralized method is shown.

The proposed approach could also be extended and used for coordination and cooperation of multi-agent systems.

REFERENCES