Hierarchical Reinforcement Learning for Optimal Control of Linear Multi-Agent Systems: the Homogeneous Case

Gangshan Jing, He Bai, Jemin George and Aranya Chakrabortty

Abstract—Individual agents in a multi-agent system (MAS) may have decoupled open-loop dynamics, but a cooperative control objective usually results in coupled closed-loop dynamics thereby making the control design computationally expensive. The computation time becomes even higher when a learning strategy such as reinforcement learning (RL) needs to be applied to deal with the situation when the agents dynamics are not known. To resolve this problem, this paper proposes a hierarchical RL scheme for a linear quadratic regulator (LQR) design in a continuous-time linear MAS. The idea is to exploit the structural properties of two graphs embedded in the $Q$ and $R$ weighting matrices in the LQR objective to define an orthogonal transformation that can convert the original LQR design to multiple decoupled smaller-sized LQR designs. We show that if the MAS is homogeneous then this decomposition retains closed-loop optimality. Conditions for decomposability, an algorithm for constructing the transformation matrix, a hierarchical RL algorithm, and robustness analysis when the design is applied to non-homogeneous MAS are presented. Simulations show that the proposed approach can guarantee significant speed-up in learning without any loss in the cumulative value of the LQR cost.

Index Terms—Reinforcement learning, linear quadratic regulator, multi-agent systems, decomposition.

I. INTRODUCTION

Optimal control of multi-agent systems (MASs) has a long-standing literature in both model-based [1]-[6], and model-free [7]-[13] settings. One common challenge in both problems, however, is the high computational cost of the control design that often stems from the large size of typical MASs. Individual agents in the MAS may have decoupled open-loop dynamics, but the control objective is usually cooperative in nature which results in coupled closed-loop dynamics thereby making the control design large-sized. The problem becomes even more critical when the controller needs to be learned in real-time, for example, learning strategies such as reinforcement learning (RL) during situations when the agents dynamics are not known [8], [9].

In this paper, we propose a hierarchical design scheme for infinite-horizon linear quadratic regulator (LQR) optimal control of continuous-time homogeneous MAS to resolve the issue of high computation time. The fundamental idea is to define the $Q$ and $R$ matrices of the LQR objective function over a set of communication graphs $G_1$ and $G_2$, respectively, and find a transformation matrix based on the structure of these two graphs. By utilizing this matrix, the original large-size LQR problem is equivalently converted to multiple decoupled smaller-size LQR problems. The transformation matrix itself is structured in the sense that it partitions the agents into a discrete set of non-overlapping groups, where each group solves a decoupled small-size LQR problem. Due to the decoupling, all of these designs can be run independently and in parallel, thereby saving significant amounts of computational effort and time. The design holds for both model-based and model-free LQR. For the sake of this paper, we only focus on the model-free case, and develop a RL learning strategy that is compatible with the reduced-dimensional LQR designs.

Dimensionality reduction has been used in the past on many occasions to improve the computational efficiency of optimal control, in both model-based [5], [6] and model-free [10]-[12] settings, but the main difference between these approaches and our proposed approach is that the former designs all result in sub-optimal controllers while our controller, when the MAS is homogeneous, retains the optimality of the original LQR problem. We identify specific conditions on the cost function to establish this optimality. Another important difference is that in conventional designs dimensionality reduction usually happens due to grouping, time-scale separation, or spatial-scale separation in the agent dynamics. In our design, however, the reduction occurs due to the structure imposed on the control objective, not due to the plant dynamics. The results are presented in the following way. We first establish the notion of decomposability for optimal control of homogeneous MAS using Definition 1 followed by the derivation of multiple sufficient conditions for decomposability using Theorems 1 through 4. An algorithm (Algorithm 1) for constructing the transformation matrix is proposed, and a corresponding hierarchical RL algorithm (Algorithm 2) is developed to parallelize the control design in situations where the agent models may not be known. Finally, a detailed robustness analysis of the hierarchical controller is presented to encompass stability, performance, and implementation challenges when this controller is applied to a non-homogeneous MAS.

The rest of the paper is organized as follows. Section II introduces the main problem formulation and defines decomposability. Section III presents multiple conditions for decomposability and proposes Algorithm 1 for construction of the transformation matrix. Section IV develops the hierarchical RL algorithm. Section V presents robustness analysis. Section VI shows a simulation example. Section VII concludes the paper.
Notation: Throughout the paper, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denotes an unweighted undirected graph with $N$ vertices, where $\mathcal{V} = \{1, \ldots, N\}$ is the set of vertices, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges; Graph $\mathcal{G}$ is said to be disconnected if there are two nodes $i$ and $j$ with no path between them, i.e., there does not exist a sequence of distinct edges of the form $(i_1, i_2), (i_2, i_3), \ldots, (i_{r-1}, i_r)$ where $i_1 = i$ and $i_r = j$. The $d \times d$ identity matrix is denoted by $I_d$. The Kronecker product is denoted by $\otimes$. Given a matrix $X$, $X \succ 0$ implies that $X$ is positive semi-definite; $\text{im}(X)$ denotes the image space of $X$. We use $\text{diag}(A_1, \ldots, A_N)$ to denote a block diagonal matrix with $A_i$'s on the diagonal. Matrix $G_{ab}(s)$ denotes the transfer function from input $a$ to output $b$. Given a transfer function $G(s)$, its $\mathcal{H}_\infty$ norm is $\|G(s)\|_\infty = \sup_{\omega} \sigma_{\text{max}}(G(j\omega))$, its $\mathcal{H}_2$ norm is $\|G(s)\|_2 = (\frac{1}{2\pi} \int_{-\infty}^{\infty} \text{tr}(G'(j\omega)G(j\omega)) d\omega)^{1/2}$. The Euclidean norm is denoted by $\|\cdot\|$.

II. PROBLEM STATEMENT

Consider the following linear homogeneous MAS:

$$\dot{x}_i = Ax_i + Bu_i, \quad i = 1, \ldots, N$$

(1)

where $x_i \in \mathbb{R}^n$ and $u_i \in \mathbb{R}^m$ denote the state and the control input of agent $i$, respectively. Throughout this paper, we assume that $A$ and $B$ are unknown, but the dimension of $x$ and $u$ are known. Let $x = (x_1^\top, \ldots, x_N^\top) \in \mathbb{R}^{Nn}$ and $u = (u_1^\top, \ldots, u_N^\top) \in \mathbb{R}^{Nm}$, the optimal control problem to be solved in our paper is formulated as

$$\min_u J(x(0), u) = \int_0^\infty (x^\top Qx + u^\top Ru) dt$$

(2)

s.t. $\dot{x} = (I_N \otimes A)x + (I_N \otimes B)u$.

where,

$$Q = G_1 \otimes Q_0, \quad R = G_2 \otimes R_0.$$

(3)

The two graphs $G_1 \succeq 0 \in \mathbb{R}^{N \times N}$ and $G_2 \succ 0 \in \mathbb{R}^{N \times N}$ characterize the couplings between the different agents in their desired transient cooperative behavior, with $Q_0 \succeq 0 \in \mathbb{R}^{n \times n}$ and $R_0 \succ 0 \in \mathbb{R}^{m \times m}$.

RL algorithms for solving LQR control in the absence of $A$ and $B$ have been introduced in [8], [9]. However, naively applying these algorithms to a large-size network of agents would involve repeated inversions of large matrices, making the overall design computational expensive. Depending on the values of $n$ and $N$, the learning time in that case can become unacceptably high. To resolve this problem, in the following we will study the situations when the LQR control problem can be equivalently decomposed into multiple smaller-size decoupled LQR problems. For this, we define the notion of decomposability of (2) as follows.

Definition 1. Problem (2) is said to be decomposable if there exist $r > 1$ functions $J_i(\xi_i(0), v_i)$ such that

$$J(x(0), u) = \sum_{i=1}^r J_i(\xi_i(0), v_i),$$

(4)

$$J_i(\xi_i(0), v_i) = \int_0^\infty (\xi_i^\top Q_i \xi_i + v_i^\top R_i v_i) dt,$$

(5)

and

$$\xi_i = (I_N \otimes A)\xi_i + (I_N \otimes B)v_i, \quad i = 1, \ldots, r,$$

(6)

where $Q_i \in \mathbb{R}^{nN_1 \times nN_1}, R_i \in \mathbb{R}^{mN_1 \times mN_1}, \xi_i \in \mathbb{R}^{nN_1},$ and $v_i \in \mathbb{R}^{mN_1}, \sum_{i=1}^r N_i = N$.

Definition 2. Problem (2) is said to be completely decomposable if it is decomposable with $r = N$.

Let $\xi = (\xi_1^\top, \ldots, \xi_N^\top) \in \mathbb{R}^{Nn}$. It is observed from Definition 1 that the dynamics of $\xi$ is identical to $x$. In fact, $\xi$ is a vector stacking up states of $N_i$ agents, and therefore can be viewed as the state vector of a cluster containing partial agents in the whole group. When problem (2) is completely decomposable, each agent is viewed as a cluster.

III. RECOGNITION OF DECOMPOSABLE OPTIMAL CONTROL PROBLEMS

In this section, we propose several conditions for decomposability of the problem (2), followed by an algorithm to construct a transformation matrix for decomposition. The algorithm can also be used to identify if a given problem is decomposable.

A. Conditions for Decomposability

We start by defining a simultaneously orthogonally block-diagonalizable matrix.

Definition 3. Two matrices $X \in \mathbb{R}^{N \times N}$ and $Y \in \mathbb{R}^{N \times N}$ are simultaneously block-diagonalizable with respect to a disconnected graph $\mathcal{G}$ if there exists an orthogonal matrix $T$ such that $TXT^{-1}$ and $TYT^{-1}$ are both block-diagonal, and $TXT^{-1}, TYT^{-1} \in S(\mathcal{G})$ for graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Here

$$S(\mathcal{G}) \triangleq \{ M \in \mathbb{R}^{N \times N} : M_{ij} = 0 \text{ if } (i, j) \notin \mathcal{E} \}$$

denotes the set of matrices with sparsity patterns similar to that of the graph $\mathcal{G}$.

Since a disconnected graph always has multiple independent connected components, each block on the diagonal of $TXT^{-1}$ and $TYT^{-1}$ corresponds to one or multiple independent connected components in $\mathcal{G}$. Moreover, due to the property of block-diagonal matrices, graph $\mathcal{G}$ in Definition 3 must be undirected and may have self-loops. The following theorem presents a condition for decomposability of (2).

Theorem 1. Problem (2) is decomposable if $G_1$ and $G_2$ are simultaneously block-diagonalizable with respect to some graph $\mathcal{G}$.

Proof. Since $G_1$ and $G_2$ are simultaneously block-diagonalizable, there exists an orthogonal matrix $T \in \mathbb{R}^{N \times N}$ such that

$$TG_1 T^{-1} = \Phi = \text{diag}(\Phi_1, \ldots, \Phi_r),$$

(7)

$$TG_2 T^{-1} = \Psi = \text{diag}(\Psi_1, \ldots, \Psi_r),$$

(8)

where $\Phi_i \in \mathbb{R}^{N_i \times N_i}, \Psi_i \in \mathbb{R}^{N_i \times N_i}$. We will show that the homogeneous MAS (7) can be clustered into $r$ subgroups, with the $i$-th subgroup including $N_i$ agents.
Define $A_i = I_{N_i} \otimes A$, $B_i = I_{N_i} \otimes B$, $T = T \otimes I_{n}$, 
\[\xi = (\xi_1^T, ..., \xi_r^T)^T = (T \otimes I_n)x \in \mathbb{R}^{nN},\]
\[v = (v_1^T, ..., v_r^T)^T = (T \otimes I_m)u \in \mathbb{R}^{mN},\]
where $\xi_i \in \mathbb{R}^{nN_i}$, $v_i \in \mathbb{R}^{mN_i}$. Then the following holds:
\[x^TQx = (T^x)^T(T \otimes I_N)(G_1 \otimes Q_0)(T \otimes I_N)^{-1}Tx = \xi^T(TG_1T^{-1} \otimes Q_0)\xi = \sum_{i=1}^r \xi_i^T(\Phi_i \otimes Q_0)\xi_i.\]
Similarly, we have
\[u^TRu = \sum_{i=1}^r v_i^T(\Psi_i \otimes R_0)v_i,\]
Furthermore,
\[\hat{\xi} = \bar{T}((I_N \otimes A)x + (I_N \otimes B)u) = \bar{T}(I_N \otimes A)\bar{T}^T\xi + (I_N \otimes B)(T^T \otimes I_m)v = (I_N \otimes A)\xi + (I_N \otimes B)v.\]
This completes the proof.

Given a matrix $\Gamma \in \mathbb{R}^{N \times s}$ and matrix $G \in \mathbb{R}^{N \times N}$, $\Gamma$ is said to be $G$-invariant if there exists a matrix $C \in \mathbb{R}^{s \times s}$ such that $G\Gamma = \Gamma C$. Then we have the following result.

**Theorem 2.** Matrices $G_1$ and $G_2$ are simultaneously block-diagonalizable with respect to a disconnected graph if and only if there exists a matrix $\Gamma \in \mathbb{R}^{N \times s}$ with $\Gamma^\top \Gamma = I_s$ and $s < N$ such that $\Gamma$ is both $G_1$-invariant and $G_2$-invariant.

**Proof.** We first prove “necessity”. Let $T \in \mathbb{R}^{N \times N}$ be the orthogonal matrix such that \[1\] and \[8\] hold. We partition $T$ into two parts: $T = [\bar{T}, \bar{T}]^\top$, where $\bar{T} \in \mathbb{R}^{N \times N_i}$ and $\bar{T} \in \mathbb{R}^{N \times (N-N_i)}$. Then $\bar{T}^\top \bar{T} = I_{N_i}$ and $\bar{T}^\top \bar{T} = I_{N-N_i}$.

From \[7\], it holds that $\bar{T}^\top G_1 \bar{T} = \Phi_1$, $\bar{T}^\top G_1 \bar{T} = 0$.

Note that $\bar{T}^\top \bar{T} = 0$ and $[\bar{T}, \bar{T}] = T^\top$ is non-singular. Then $\text{im}(\bar{T})$ is the orthogonal complement space of $\text{im}(\bar{T})$. It follows that $G_1 \text{im}(\bar{T}) \subset \text{im}(\bar{T})$, implying that $\bar{T}$ is $G_1$-invariant. Similarly, it can be proved that $\bar{T}$ is $G_2$-invariant. Therefore, choosing $\Gamma = \bar{T}$ proves the necessity.

Next we prove “sufficiency”. $\Gamma^\top \Gamma = I_s$ implies that any pair of columns of $\Gamma$ are orthogonal to each other. Then there must exist another matrix $\Xi \in \mathbb{R}^{N \times N-s}$ such that $[\Gamma, \Xi]$ is orthogonal. Let $T = [\Gamma, \Xi]^\top$, we have
\[TG_i T^{-1} = \left[\begin{array}{c} \Gamma^\top G_1 \Gamma \\ \Xi^\top G_1 \Gamma \end{array}\right] \Xi, \quad i = 1, 2.\]
Since $\Gamma$ is $G_1$-invariant and $G_2$-invariant, $\Gamma^\top G_i \Xi = 0$ for $i = 1, 2$. Therefore, $G_1$ and $G_2$ are simultaneously block-diagonalizable.

Specifically, when $s = 1$, we obtain the following corollary immediately.

**Corollary 1.** Matrices $G_1$ and $G_2$ are simultaneously block-diagonalizable if they have a common eigenvector.

The following theorem gives another equivalent condition for $G_1$ and $G_2$ to be simultaneously block-diagonalizable.

**Theorem 3.** Matrices $G_1$ and $G_2$ are simultaneously block-diagonalizable with respect to graph $\mathcal{G}$ if and only if there exist orthogonal matrices $E_1$ and $E_2$ such that $E_1G_1E_1^\top$ and $E_2G_2E_2^\top$ are diagonal, and $E_1E_2^\top \in \mathcal{S}(\mathcal{G})$.

**Proof.** We first prove “necessity”. Since $G_1$ and $G_2$ are both symmetric, they are both diagonalizable. It follows that there exist orthogonal matrices $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{N \times N}$ such that $TG_1T^{-1} = U\text{diag}\{\mu_1, ..., \mu_N\}U^\top$, $TG_2T^{-1} = V\text{diag}\{\mu_1, ..., \mu_N\}V^\top$.

Let $\mathcal{G}$ be a disconnected graph such that $TG_1T^{-1} \in \mathcal{S}(\mathcal{G})$ and $TG_2T^{-1} \in \mathcal{S}(\mathcal{G})$. Then there must exist feasible $U$ and $V$ such that $U, V \in \mathcal{S}(\mathcal{G})$.

because such a matrix $U$ can be obtained by designing each block in $U^\top$ as a collection of eigenvectors of a block in $\Phi$, so does $V$. Let $E_1 = U^\top T$ and $E_2 = V^\top T$, then each column of $E_1^\top$ is an eigenvector of $G_1$, $i = 1, 2$. It follows that $UE_1 = V^\top V$.

Due to the fact $U, V \in \mathcal{S}(\mathcal{G})$, we have
\[E_1 E_1^\top = U^\top V \in \mathcal{S}(\mathcal{G}).\]

This implies that $E_1 E_2^\top \in \mathcal{S}(\mathcal{G})$.

Next we prove “sufficiency”. Suppose $E_1$ and $E_2$ exist and \[15\] holds. By setting $U = E_1E_1^\top \in \mathcal{S}(\mathcal{G})$, $V = (E_1E_2^\top)^2$ and $T = UE_1 = V^\top E_2 = E_2 E_1^\top$, we obtain that
\[TG_1T^{-1} = U E_1 G_1 E_1^\top U^\top \in \mathcal{S}(\mathcal{G}),\]
\[TG_2T^{-1} = V E_2 G_2 E_2^\top U^\top \in \mathcal{S}(\mathcal{G}).\]
This completes the proof.

Also note that two diagonalizable matrices are simultaneously diagonalizable if and only if they commute [14, Theorem 1.3.12]. As the cost function is considered to be in a quadratic form, both $G_1$ and $G_2$ are symmetric, thus are diagonalizable. Hence, we have the following theorem.

**Theorem 4.** Problem \[2\] is completely decomposable if $G_1$ and $G_2$ commute.

**Remark 1.** In practice, $G_2$ is usually a diagonal matrix when there are no relative input efforts between different agents to be minimized. In such scenarios, the diagonal elements of $G_2$ can be designed according to the eigenvectors of $G_1$ such that the condition in Theorem \[3\] is satisfied. Specifically, when $G_2 = I_N$, which happens in
many scenarios, the LQR design problem is always completely decomposable because $I_N$ commutes with any $N$-dimensional matrix. In this case, a suitable choice of $T$ is the matrix such that each column of $T^\top$ is an eigenvector of $G_1$. 

B. Construction of the Transformation Matrix

In this subsection, we study how to check if $G_1$ and $G_2$ are simultaneously block-diagonalizable and construct the transformation matrix $T$ (if it exits) such that $TG_1T^{-1}$ and $TG_2T^{-1}$ are simultaneously diagonalized. According to the proof of Theorem 3, to find an appropriate transformation matrix $T$, we only need to find appropriate $E_1$ and $E_2$ such that $E_1E_2^\top \in S(G)$ for some $G$. 

For $l = 1, 2$, let $E_l$ be an orthogonal matrices such that $E_l^\top$ collects eigenvectors of $G_l$. The sparsity pattern of $E_1E_2^\top$ is not always unique due to the following two facts: (i) when $G_1$ has an eigenvalue with its multiplicity more than 1, the corresponding orthonormal eigenvectors lie in a space containing infinite number of vectors; (ii) even when both $G_1$ and $G_2$ have $N$ distinct eigenvalues, different sequences of eigenvectors in $E_1^\top$ and $E_2^\top$ lead to different sparsity patterns of $E_1E_2^\top$.

The problem of finding appropriate $E_1$ and $E_2$ satisfying the condition in Theorem 3 can be formulated as the following feasibility problem:

Find $E_1, E_2$

s.t. $E_1E_2^\top \in S(G)$,

$E_lG_1E_l^\top$ is diagonal, $l = 1, 2$,

$G$ is disconnected. (16)

Problem (16) is nonlinear since both $E_1$ and $E_2$ are variables. It becomes more complicated when $G_1$ or $G_2$ has some eigenvalues with multiplicity more than 1. Moreover, the solution to (16) may not correspond to the graph $G$ with the largest number of connected components. In what follows, we deal with the special case where the following assumption holds:

Assumption 1. Both $G_1$ and $G_2$ have $N$ distinct eigenvalues.

Under Assumption 1, the eigenspace for each eigenvalue has dimension 1. Therefore, we only need to figure out how to order the columns of $E_1$ and $E_2$ such that $E_1E_2^\top$ is block-diagonalizable. We present Algorithm 1 for constructing the transformation matrix $T$. It is worth noting that under Assumption 1, the matrix $T$ obtained by Algorithm 1 leads to the graph $G$ with the largest number of connected components.

Remark 2. Step 2 of Algorithm 1 can be achieved by checking the product of every pair of vectors $p_i$ and $q_j$ for $i, j = 1, ..., N$. To construct $\Pi_1$ and $\Pi_2$, categorize $i$ and $j$ into the same cluster $\{j\}$, and categorize $j$ into cluster $\{i\}$ if there exists an index $j \in \mathcal{V}$ such that $p_i^\top q_j \neq 0$.

1Since both $G_1$ and $G_2$ are symmetric, for each eigenvalue of $G_1$ or $G_2$, its geometric multiplicity is equal to its algebraic multiplicity. Hence, we use the term "multiplicity" to denote either of them.

Algorithm 1 Constructing the Transformation Matrix $T$

Input: $G_1$ and $G_2$ satisfying Assumption 1

Output: $T$.

1. Perform eigenvalue decomposition of $G_1$ and $G_2$, respectively. Obtain $F_1 = [p_1, ..., p_N]^\top$ and $F_2 = [q_1, ..., q_N]^\top$ such that the $i$-th column of $F_i^\top$ is an eigenvector of $G_i$, $l = 1, 2$.

2. Find two partitions $\Pi_1 = \{v_1^1, ..., v_1^l\}$ and $\Pi_2 = \{v_2^1, ..., v_2^m\}$ for $l = 1, 2$ with $\cap_{k=1}^{l} v_1^k = \mathcal{V}$ such that $|v_1^k| = |v_2^k| = N_k \geq 1$ for $k = 1, ..., r$, and for any $i \in v_1^k$ and $j \in v_2^k$, with $k_1 \neq k_2$, it holds that $p_i^\top q_j = 0$. If such a partition cannot be found for $r > 1$, then $G_1$ and $G_2$ are not simultaneously diagonalizable; otherwise go to Step 3.

3. According to the partitions $\Pi_1$ and $\Pi_2$, obtain two index sequences $\{\pi_1(1), ..., \pi_1(N)\}$, $i = 1, 2$ such that

$$v_1^k = \{\pi_1(\sum_{\tau=0}^{k-1} N_\tau + 1), ..., \pi_1(\sum_{\tau=0}^{k} N_\tau)\}, k = 1, ..., r,$$

where $N_0 = 0$. Then construct two permutation matrices $P_l \in \mathbb{R}^{N \times N}$, $l = 1, 2$ such that

$$P_l(i, j) = \begin{cases} 1, & \text{if } \pi_l(i) = j, \\ 0, & \text{otherwise.} \end{cases} (17)$$

Let $E_1 = P_1F_1$ for $l = 1, 2$. Choose $U = E_1E_2^\top$. Matrix $T$ is obtained by

$$T = UE_1 = E_1E_2^\top E_2.$$ 

and $p_i^\top q_j \neq 0$. According to the number of elementary operations, the time complexity of Step 2 is $\mathcal{O}(N^2)$. Although we also need to execute eigenvalue decomposition for two $N \times N$ matrices once, the computational complexity of Algorithm 1 is much smaller than that of the RL algorithm in [8], in which the inversion for a $(nN(nN+1)/2+mnN^2)$-dimensional matrix has to be computed in each iteration of the RL algorithm.

Remark 3. Note that under Assumption 1 when $G_1$ and $G_2$ are simultaneously diagonalizable, there are infinite possibly choices for $T$. More specifically, to construct $T$, matrix $U$ can be any orthogonal matrix in $S(G)$. According to different choices of $U$, one can obtain different $V$ and $T$. For example, in Algorithm 1, based on the choice of $U$, matrix $V$ can be derived from (14) as follows:

$$V = UE_1E_2^\top = E_1E_2^\top E_2^\top .$$

Another feasible choice for $U$, $V$ and $T$ will be

$$U = E_2E_1^\top E_2^\top, V = E_2E_1^\top, T = E_2E_1^\top E_2.$$

It can be observed that in both the above two cases, once $E_2E_1^\top \in S(G)$, it always holds that $U, V \in S(G)$.

IV. HIERARCHICAL RL ALGORITHM DESIGN

In this section, we propose conditions for existence and uniqueness of the solution to each smaller-size problem, and establish the relationship between the lower-dimensional...
optimal controllers and the optimal controller of the original problem \(2\). Based on this relationship, we propose a hierarchical RL algorithm to solve \(2\).

A. Existence and Uniqueness of the Optimal Controller

Let \(A = I_N \otimes A\), and \(B = I_N \otimes B\). To guarantee the existence and uniqueness of the solution to problem \(2\), we make the following assumption.

**Assumption 2.** The pair \((A, B)\) is controllable and \((Q^{1/2}, A)\) is observable.

Assumption 2 implicitly implies that \(G_1\) is non-singular, as proved in the following theorem.

**Theorem 5.** Given problem \(2\) with \(Q\) and \(R\) defined in \(3\), then \((Q^{1/2}, A)\) is observable if and only if \(G_1 > 0\).

**Proof.** The “sufficiency” holds because \(Q = G_1 \otimes Q_0 > 0\). Next we prove “necessity”. Since \(G_1\) is symmetric, there exists an orthogonal matrix \(S \in \mathbb{R}^{N \times N}\) such that 
\[
SG_1 S^{-1} = \text{diag}\{\lambda_1, ..., \lambda_N\}.
\]
It suffices to show \(\lambda_i > 0\) for all \(i = 1, ..., N\). Let \(O\) be the observability matrix corresponding to \((Q^{1/2}, I_N \otimes A)\). Then \(\text{rank}(O) = nN\), and
\[
O = \begin{pmatrix}
G_1^{1/2} \otimes Q_0^{1/2} \\
G_1^{1/2} \otimes (Q_0^{1/2} A) \\
\vdots \\
G_1^{1/2} \otimes (Q_0^{1/2} A^{nN-1})
\end{pmatrix}.
\]
Note that 
\[(SG_1^{1/2} S^{-1})^2 = SG_1^{1/2} S^{-1} SG_1^{1/2} S^{-1} = SG_1 S^{-1},\]
implying that \(SG_1^{1/2} S^{-1} = (SG_1 S^{-1})^{1/2}\). As a result, 
\[
\tilde{O} \triangleq (I_{nN} \otimes (S \otimes I_n)) O (I_{nN} \otimes (S \otimes I_n))^{-1}
\]
\[
= \begin{pmatrix}
(S \otimes I_n)(G_1^{1/2} \otimes Q_0^{1/2})(S^{-1} \otimes I_n) \\
(S \otimes I_n)(G_1^{1/2} \otimes (Q_0^{1/2} A))(S^{-1} \otimes I_n) \\
\vdots \\
(S \otimes I_n)(G_1^{1/2} \otimes (Q_0^{1/2} A^{nN-1}))(S^{-1} \otimes I_n)
\end{pmatrix}
\]
\[
= \begin{pmatrix}
\text{diag}\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_N}\} \otimes Q_0^{1/2} \\
\text{diag}\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_N}\} \otimes (Q_0^{1/2} A) \\
\vdots \\
\text{diag}\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_N}\} \otimes (Q_0^{1/2} A^{nN-1})
\end{pmatrix}.
\]
\[
\text{rank}(O_i) = \text{rank}\begin{pmatrix}
\Phi_i^{1/2} \otimes Q_0^{1/2} \\
\Phi_i^{1/2} \otimes Q_0^{1/2} A \\
\vdots \\
\Phi_i^{1/2} \otimes Q_0^{1/2} A^{nN-1-k}
\end{pmatrix}
\]
for any \(k \geq 0\). Due to observability of \(((\Phi \otimes Q_0)^{1/2}, A)\), we have 
\[
\text{rank}(O) = \sum_{i=1}^{r} \text{rank}(O_i) = nN.
\]
Together with \(\text{rank}(O_i) \leq nN_i\), we have \(\text{rank}(O_i) = nN_i\) for all \(i = 1, ..., r\).

B. Hierarchical RL Algorithm

We next propose a hierarchical RL algorithm to synthesize the optimal controller when \(2\) is decomposable. Let \(u^* = -K^* x\) be the optimal controller for problem \(2\), \(v_i^* = -\kappa_i \xi_i\) be the optimal controller for the \(i\)-th optimal control problem in \(19\). Define \(T_i^T \in \mathbb{R}^{N \times N}\) as a matrix consisting of the \(N_i\) rows in \(T\) corresponding to the \(i\)-th cluster. That is, \(T = \cdots \)
Hierarchical RL Algorithm for Optimal Control

minimum variance problem, it also applies to a deterministic case. Although Algorithm 3 is designed for solving a deterministic problem, it also applies to a deterministic LQR problem by setting the coefficients appropriately.

Algorithm 2 Hierarchical RL Algorithm for Optimal Control of Homogeneous Linear MAS

Input: $G_1$, $G_2$, $T$, $Q_0$, $R_0$.

Output: Optimal control gain $K^*$

1. Obtain scalars $N$, $n$ and $m$ from the dimensions of $G_1$, $Q_0$ and $R_0$. Compute $\Phi_i$ and $\Psi_i$ from (7) and (9), respectively.
2. Run [13, Algorithm 3] to solve the $i$-th LQR problem (19) for $i = 1, ..., r$. Obtain the optimal control gain $\kappa_i$ for the $i$-th LQR problem.
3. Compute the global optimal control gain $K^*$ according to (21).

V. ROBUSTNESS ANALYSIS

In reality, a MAS may not have perfectly homogeneous agents. We next present a thorough robustness analysis of our hierarchical LQR controller when the controller is designed under the assumption of homogeneity as in Algorithm 2, but is implemented in a heterogenous MAS.

Suppose the dynamic model of agent $i$ is given as

$$\dot{x}_i = A^h_i x_i + B^h_i u_i, \quad i = 1, ..., N. \tag{22}$$

The MAS model is written in a compact form as

$$\dot{x} = A^h x + B^h u. \tag{23}$$

Both $A^h$ and $B^h$ are unknown. The transfer function of $A^h$ from $u$ to $x$ is

$$G(s) = (sI_{nN} - A^h)^{-1} B^h. \tag{24}$$

Throughout this section, we make the following assumption:

Assumption 3. $(A^h, B^h)$ is controllable, $Q > 0$.

Note that the optimal control problem for homogeneous MAS, as shown in Theorem 5. If $Q$ has some zero eigenvalue, the condition for implementation of Algorithm 2 is not satisfied.

Let $T \in \mathbb{R}^{N \times N}$ be the transformation matrix such that $TG_1T^T$ and $TG_2T^T$ are simultaneously block-diagonal. Let $K$ be the control gain learned by Algorithm 2. From the proof of Theorem 1, we know that $K$ is the optimal control gain of the following problem:

$$\min_{\hat{J}} \hat{J}(\hat{x}(0), \zeta) = \int_0^\infty (\hat{x}^T Q \hat{x} + \zeta^T R \zeta) dt$$

s.t. $\dot{x} = T^T A^h T \tilde{x} + T^T B^h \tilde{T} \zeta,$

where $\tilde{T} = T \otimes I_n, \tilde{T} = T \otimes I_n$. This implies that

$$\hat{A} = T^T A^h T - T^T B^h K,$$

is Hurwitz. Under Assumption 3, we know $((TQT)^{1/2}, A^h)$ is observable. Then there is a unique positive definite solution $P$ to the following algebraic Riccati equation:

$$A^h P + PA^h + TQT - PB^h R^{-1} T^T B^h P = 0. \tag{26}$$

Let $\hat{P} = T^T P T$. Then we have $K = R^{-1}(T^T B^h \hat{P})^T \hat{P}$. Let $\hat{A} = A^h - T^T A^h T, \hat{B} = B^h - T^T B^h T$. We present stability and performance analysis respectively as follows.

A. Stability Analysis

The following theorem gives a condition on $\hat{A}$ and $\hat{B}$ such that system (23) with control gain $K$ is stable.

Theorem 7. Consider the optimal control of the heterogeneous MAS [23] with the cost function $J(x(0), u)$ in (2). Suppose that $G_1$ and $G_2$ are simultaneously block-diagonalizable. By implementing the optimal control gain learned by Algorithm 2, MAS (23) is asymptotically stable if

$$\dot{\hat{P}} + \hat{P} \hat{A}^T - \hat{B} R^{-1} \hat{B}^T \hat{P} = Q - \hat{P} B^h R^{-1} B^h T \hat{P} < 0. \tag{27}$$

Proof. By applying the controller $u = -K x$, system (23) becomes

$$\dot{x} = \hat{A} x + (\hat{A} - \hat{B} K)x. \tag{28}$$

Consider $V = x^T \hat{P} x$ as a Lyapunov function. It follows that

$$\dot{V} = x^T (\hat{P} \hat{A} + \hat{A}^T \hat{P}) x + x^T (\hat{P} (\hat{A} - B K) + (\hat{A} - \hat{B} K)^T \hat{P}) x$$

$$= x^T (-Q - \hat{P} B^h R^{-1} B^h T \hat{P}) x$$

$$+ x^T (\hat{P} \hat{A} + \hat{A}^T \hat{P} - B B^h R^{-1} B^h T \hat{P} - \hat{B} B^h R^{-1} B^h T \hat{P}) x. \tag{29}$$

Condition (27) implies that $V < 0$ if $x \neq 0$ and $V = 0$ otherwise. Therefore, MAS (23) with control gain $K$ is asymptotically stable.

Theorem 7 indicates stability robustness of our hierarchical LQR by giving a condition on $\hat{A}$ and $\hat{B}$ associated with $\hat{P}$. When the open loop system (23) is stable, the following theorem employs the small-gain theorem to obtain a condition on $\hat{A}$ and $\hat{B}$ associated with the transfer function $G(s)$ of (23) for control gain $K$ to be stabilizing.
Theorem 8. Consider the optimal control of the heterogeneous MAS (23) with the cost function $J(x(0), u)$ in (2). Suppose that $G_1$ and $G_2$ are simultaneously block-diagonalizable and $A^h$ is Hurwitz. By implementing the optimal control gain learned by Algorithm 2, MAS (23) is asymptotically stable if

$$||G_\Sigma(s)||_\infty < ||K(sI_{n,N} - ˜A)^{-1}||_\infty^{-1},$$

where

$$G_\Sigma(s) = ( ˜A - ˜B K) G(s).$$

Specifically, when the MAS is homogeneous, $G_\Sigma(s) = 0$.

Proof. Applying the controller $u = - K x$ to (23), then closed-loop system dynamics can be rewritten as the following two interconnected systems:

$$\Sigma : \begin{cases} ˙x = A^h x + B^h u, \\ e = (A - ˜B K)x, \end{cases}$$

and

$$\Delta : \begin{cases} ˙y = ˜A \eta + e, \\ u = - K \eta. \end{cases}$$

The feedback interconnection between system $\Sigma$ and system $\Delta$ is interpreted in Fig. 1 where $x(0)$ is the initial state of system (23). Accordingly, $\eta(0) = x(0)$, $u(0) = - K \eta(0)$, $e(0) = (A - ˜B K)x(0)$. The transfer function of (32) from $u$ to $e$ is $G_\Sigma(s)$ in (31). On the other hand, the transfer function of (33) from $e$ to $u$ is

$$G_\Delta(s) = - K(sI_{n,N} - ˜A)^{-1}.$$ (34)

We note that the open loop systems of $\Sigma$ and $\Delta$ are both stable. Moreover, (30) implies that

$$||G_\Sigma(s)||_\infty ||G_\Delta(s)||_\infty < 1.$$ (35)

Using the small-gain theorem, we can conclude that the interconnection between $\Sigma$ and $\Delta$ is stable. \hfill \square

B. Performance Analysis

Next we analyze the performance of the heterogenous MAS (23) with controller $u = - K x$. Let $\sqrt{Q}$ and $\sqrt{R}$ be the two matrices such that $\sqrt{Q}^2 = Q$ and $\sqrt{R} = R$. Define

$$y = \begin{pmatrix} \sqrt{Q} \\ -\sqrt{R}K \end{pmatrix} x.$$ It follows that

$$J(x(0), u) = \int_0^\infty y^T(t)y(t)dt|_{u=x(0)\delta(t)} = ||G_{uy}x(0)||_2^2,$$ (36)

where $\delta(t) = \infty$ if $t = 0$, and $\delta(t) = 0$ if $t \neq 0$, and $\int_0^\infty \delta(t)dt = 1$.

For the convenience of analysis, we evaluate robustness with respect to the following $L_2$-norm directly:

$$J_2(x(0), u) = ||y||_2 = ||G_{uy}x(0)||_2.$$ (37)

Let $\tilde{J}_2 = J_2(\tilde{x}(0), -K \tilde{x})$ be the optimal performance for the $\tilde{x}$ dynamics in (25), where $\tilde{x}(0) = x(0)$. $J$ is finite since $K$ is the optimal control gain of problem (25). The theorem below analyzes the performance $J_2$ of the heterogeneous MAS (23) with controller $u = - K x$.

Theorem 9. Consider the optimal control of the heterogeneous MAS (23) with the cost function $J_2(x(0), u)$ in (38). Suppose that $G_1$ and $G_2$ are simultaneously block-diagonalizable. By implementing the optimal control gain learned by Algorithm 2, it holds that

$$J_2 \leq \tilde{J}_2 + \alpha \epsilon,$$ (38)

where $\epsilon = ||G_{\Sigma}(s)||_2$ and $\alpha = ||G_{ey}(I - G_{\Sigma}G_{eu})^{-1}||_2 ||G_{\delta u}||_2$. Specifically, when the MAS is homogeneous, we have $\epsilon = 0$.

Proof. We treat $e = (A - ˜B K)x$ as the disturbance, and consider the following system:

$$\begin{cases} ˙y = ˜A \eta + e + x(0)\delta, \\ u = - K \eta, \\ y = \begin{pmatrix} \sqrt{Q} \\ -\sqrt{R}K \end{pmatrix} \eta. \end{cases}$$ (39)

In the frequency domain, we have

$$\begin{pmatrix} y \\ u \end{pmatrix} = \begin{pmatrix} G_{\delta y} & G_{ey} \\ G_{\delta u} & G_{eu} \end{pmatrix} \begin{pmatrix} \delta \\ e \end{pmatrix}.$$ (40)

It follows that

$$J_2 = ||y||_2 = ||G_{\delta y} + G_{ey}(I - G_{\Sigma}G_{eu})^{-1}G_{\Sigma}G_{\delta u}||_2 \leq ||G_{\delta y}||_2 + ||G_{ey}(I - G_{\Sigma}G_{eu})^{-1}G_{\Sigma}G_{\delta u}||_2,$$ (41)

where $G_\Sigma$ is in form (31). We note that

$$||G_{\delta y}||_2 = \left\| \begin{pmatrix} \sqrt{Q} \\ -\sqrt{R}K \end{pmatrix} (sI_{n,N} - ˜A)^{-1} x(0) \right\|_2 = \tilde{J}_2.$$ (42)

Then we obtain (38). \hfill \square

Theorem 9 establishes an upper bound on $J_2$, which is associated with $||G_\Sigma(s)||_2$. The inequality (38) implies that when $||G_\Sigma(s)||_2$ is small enough, then the performance with respect to controller $u = - K x$ will be close to $\tilde{J}_2$.

VI. NUMERICAL EXAMPLES

We validate the proposed controller using two examples of MAS, one homogeneous and one heterogeneous. When Algorithm 2 is applied, the dynamics of the agents are always considered to be unknown.
Consider a MAS with $N = 100$ agents. Each agent $i$ is a second-order dynamic system with

$$A_i = \begin{pmatrix} 0 & I_2 \\ -\frac{c_i}{m_i}I_2 & 0 \end{pmatrix}, \quad B_i = \begin{pmatrix} 0 \\ \frac{1}{m_i}I_2 \end{pmatrix}, \quad i = 1, \ldots, N.$$ 

Let $G$ be a connected undirected graph with 100 nodes randomly generated in MATLAB, as shown in Fig 2. Matrices $Q$ and $R$ are set as $Q = (0.5I_N + L) \otimes I_2$ and $R = I_{20N}$, where $L$ is the Laplacian matrix of graph $G$. In this case, $G_1 = 0.5I_N + L$, $G_2 = I_N$, $Q_0 = I_n$ and $R_0 = I_m$. This formulation can describe formation control in [1], [12]. Since $G_1$ and $G_2$ commute, according to Theorem 4, the problem is completely decomposable. Matrix $T$ is constructed by collecting the orthonormal eigenvectors of $G_1$.

In both examples, we use $w \in \mathbb{R}^4$ as the common initial state for all the agents, where $w_i$ is randomly chosen from $[0,1]$, $i = 1, \ldots, 4$. Therefore, the initial state for the overall MAS is $x(0) = 1_N \otimes w$. Let $t_{\text{opt}}$ be the computational time for solving the model-based problem 2 via MATLAB, $K_{\text{opt}}$ be the obtained optimal control gain, and $J_{\text{opt}}$ be the corresponding performance. Let $t_{\text{HRL}}$ denote the computational time for implementing Algorithm 2 to solve the model-free problem, $K_{\text{HRL}}$ be the obtained control gain, and $J_{\text{HRL}}$ be the corresponding performance. Also let $t_{\text{RL}}$ be the time of solving the model-free problem via the conventional RL algorithm in [8].

**Example 1.** We first consider a homogeneous MAS where $c_i = m_i = 1$ for $i = 1, \ldots, N$. The simulation results are shown in Table 1 where "-" implies the computational time is more than one hour. This example shows that when the homogeneous MAS has a large size, the computational speed of Algorithm 2 can be even higher than that for the conventional LQR control with known $A$ and $B$, and the obtained control gain is almost optimal.

| $t_{\text{opt}}$ (s) | $t_{\text{HRL}}$ (s) | $t_{\text{RL}}$ (s) | $J_{\text{opt}}$ | $J_{\text{HRL}}$ | $||K_{\text{opt}} - K||$ |
|----------------------|----------------------|----------------------|-----------------|-----------------|-------------------|
| 4.8635               | 0.1234               | -                    | 523.7658        | 523.7662        | 0.0037            |

**Example 2.** We next consider a heterogeneous MAS, where $c_i = c + \alpha_i$, $m_i = m + \beta_i$, $c = m = 1$, $\alpha_i$ and $\beta_i$ are randomly chosen from $[-0.5, 0.5]$. The simulation results are shown in Table 2. Note that $\frac{J_{\text{HRL}} - J_{\text{opt}}}{J_{\text{opt}}} = 1.94\%$, which implies that Algorithm 2 has a strong robustness on the system performance for this example.

| $t_{\text{opt}}$ (s) | $t_{\text{HRL}}$ (s) | $t_{\text{RL}}$ (s) | $J_{\text{opt}}$ | $J_{\text{HRL}}$ | $||K_{\text{opt}} - K||$ |
|----------------------|----------------------|----------------------|-----------------|-----------------|-------------------|
| 7.0832               | 0.1184               | -                    | 538.7285        | 549.1773        | 0.8241            |

**VII. Conclusions**

We introduced the notion of “decomposability” for LQR control of homogeneous linear MAS using which a large LQR design problem can be equivalently transformed to multiple smaller-size LQR design problems. Subsequently, a hierarchical RL algorithm was proposed to solve these small-size designs and synthesize the optimal controller in a model-free way. Robustness analysis was established, followed by simulation results that clearly show that our hierarchical RL strategy is much faster than conventional RL.

REFERENCES

[1] F. Borrelli, and T. Keviczky, “Distributed LQR design for identical dynamically decoupled systems,” IEEE Transactions on Automatic Control, vol. 53, no. 8, pp. 1901-1912, 2008.

[2] G. Foderaro, S. Ferrari, and T. A. Wettergren, “Optimal Control for Multi-Agent Trajectory Optimization,” Automatica, vol. 50, no. 1, pp. 149-154, 2014.

[3] F. L. Lewis, H. Zhang, K. Hengster-Movric and A. Das, “Cooperative control of multi-agent systems: optimal and adaptive design approaches,” Springer Science & Business Media, 2013.

[4] K. H. Movric, and F. L. Lewis, “Cooperative optimal control for multi-agent systems on directed graph topologies,” IEEE Transactions on Automatic Control, vol. 59, no. 3, pp. 769-774, 2013.

[5] D. H. Nguyen, T. Nakaiyio, M. Kawani, and S. Hara, “Hierarchical decentralized robust optimal design for homogeneous linear multi-agent systems,” ArXiv 1607.01848, 2016.

[6] N. Xue, and A. Chakrabortty, “Optimal control of large-scale networks using clustering based projections,” arXiv 1609.05265, 2016.

[7] D. Vrabie, O. Pastravnanu, M. AbuKhalaf, and F. L. Lewis, “Adaptive optimal control for continuous-time linear systems based on policy iteration,” Automatica, vol. 45, no. 2, pp. 477-484, 2009.

[8] Y. Jiang and Z. P. Jiang, “Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics,” Automatica, vol. 48, no. 10, pp. 2699-2704, 2012.

[9] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, “Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers,” IEEE Control Systems Magazine, vol. 32, no. 6, pp. 76-105, 2012.

[10] S. Mukherjee, H. Bai, and A. Chakrabortty, “On Model-Free reinforcement learning of reduced-order optimal control for singularly perturbed systems,” IEEE Conference on Decision and Control, 2018.

[11] T. Sadamoto, A. Chakrabortty, and J. I. Imura, “Fast online reinforcement learning control using state-space dimensionality reduction,” IEEE Transactions on Control of Network Systems, 2020.

[12] G. Jing, H. Bai, J. George and A. Chakrabortty, “Model-Free Optimal Control of Linear Multi-Agent Systems via Decomposition and Hierarchical Approximation,” arXiv 2008.06094, 2020.

[13] G. Jing, H. Bai, J. George and A. Chakrabortty, “Model-free reinforcement learning of minimal-cost variance control,” IEEE Control Systems Letters, vol. 4, no. 4, pp. 916-921, 2020.

[14] R.A. Horn, and C.R. Johnson, “Matrix analysis,” Cambridge University press, 2012.