Model Reduction of Consensus Network Systems via Selection of Optimal Edge Weights and Nodal Time-Scales

Ralph Sabbagh and Dany Abou Jaoude

Abstract—This paper proposes model reduction approaches for consensus network systems based on a given clustering of the underlying graph. Namely, given a consensus network system of time-scaled agents evolving over a weighted undirected graph and a graph clustering, a parameterized reduced consensus network system is constructed with its edge weights and nodal time-scales as the parameters to be optimized. $H_\infty$- and $H_2$-based optimization approaches are proposed to select the reduced network parameters such that the corresponding approximation errors, i.e., the $H_\infty$- and $H_2$-norms of the error system, are minimized. The effectiveness of the proposed model reduction methods is illustrated via a numerical example.

I. INTRODUCTION

High-dimensional models of dynamical network systems are increasingly studied in the context of multi-agent systems across various applications ranging from network optimization [1] to synchronization and distributed coordination [2], [3]. In [4], [5], $H_\infty$- and $H_2$-based performance analyses and network designs are performed for consensus network systems using the edge variant representation of the consensus protocol. As network systems grow in size and complexity, it becomes desirable to find lower-order approximant network models that exhibit similar input-output behavior. Structure-preserving model reduction techniques for interconnected systems based on the balanced truncation method are presented in [6], [7], [8]. In [6], a generalized balanced truncation approach is proposed in which the resulting reduced order model is amenable to a network interpretation. However, no clear relation between the original and reduced network structures is established. In fact, the reconstructed reduced network system is found to evolve exclusively over a complete graph structure. In [7], [8], spatial states are used to represent the interconnections between the agents of the network system. The balanced truncation methods proposed therein apply for systems that are strongly stable, i.e., systems that possess appropriately structured generalized gramians. In [9] and [10], coprime factors reduction methods are proposed as extensions to the balanced truncation methods in [7] and [8] for systems that are not strongly stable but are strongly stabilizable and strongly detectable. The method in [9] is illustrated on a consensus network system example.

In the present paper, we propose model reduction methods for finding a reduced consensus network system that approximates the input-output behavior of a given consensus network system, based on a given clustering of the underlying graph. In clustering-based model reduction techniques, major emphasis is placed on finding an optimal clustering, after which the reduced network system is constructed via projection using the corresponding clustering characteristic matrix. A survey of clustering-based model reduction techniques is found in [11]. Instead of focusing on cluster selection, the works of [12], [13] propose an $H_2$-based model reduction technique for consensus network systems that optimizes the edge weights of the reduced network model for a given predefined graph clustering. Namely, the parameterized edge weights are tuned iteratively via an optimization algorithm to minimize the $H_2$-norm of the error system. The proposed algorithm was shown therein to be effective, i.e., to improve the reduction error, when the edge weights are initialized as the outcome of standard clustering-based projection methods.

In this paper, we extend the scope of the work in [12], [13] in two ways. First, we consider reduced order systems that are parameterized using both edge weights and nodal time-scales. That is, in comparison with the works of [12] and [13], we allow for additional degrees of freedom in the reduced order system, thereby allowing to further reduce the reduction error. Second, in addition to extending the $H_2$-based model reduction method in [12] along with the optimization approach in [13] to further account for optimizing the nodal time-scales, we propose a novel $H_\infty$-based model reduction method for the selection of both edge weights and nodal time-scales. Our proposed $H_\infty$-based model reduction method is analogous, yet complementary, to the methods in [12] and [13] in that it optimizes a different norm of the error system. Our $H_\infty$-based model reduction method remains novel even when the edge weights are the only parameters considered for the reduced order system. To conclude, we note that the model reduction approach in [13] allows for dealing with directed graphs, whereas this paper and the work in [12] focus on undirected graphs. The main contributions of our paper are summarized as follows:

- We propose an $H_\infty$-based model reduction method for consensus network systems that finds an approximant reduced network system via iterative tuning of its parameterized edge weights and nodal time-scales.
- We extend the $H_2$-based model reduction method of [12] by parameterizing the nodal time-scales in the reduced order system in addition to the edge weights.
- We show the efficacy of our methods via an example, wherein our algorithms are initialized using the outcome of the Petrov-Galerkin Projection (PGP) paradigm [14].

The addition of extra parameters, i.e., the nodal time-scales, in the reduced order system increases the complexity of the optimization problems to be solved. To solve the formulated problems, we follow the approach adopted in [13], which is based on the work of [15] that proposes an iterative method for handling bilinear matrix inequality constraints whose left-hand-side admits a convex-concave decomposition. Specifically, our proposed approaches optimize the edge weights for fixed nodal time-scales and the nodal time-scales for fixed edge weights. While the weights optimization
problems are readily expressed in a form that can be handled by the method of [15], as will be seen, the time-scales optimization problems require more involved manipulations to be expressed in such a form.

The paper is structured as follows. In Section II, the notation, preliminaries, and problem setup of the paper are presented. In Section III, the $\mathcal{H}_{\infty}$- and $\mathcal{H}_2$-based model reduction problems are formulated; and algorithms are proposed to solve them. The effectiveness of the proposed methods is demonstrated in Section IV via a numerical example. The paper concludes with Section V.

II. PRELIMINARIES AND PROBLEM SETUP

This section presents the notation, preliminaries, and problem setup of the paper.

A. Notation

The cardinality of a set $\mathcal{M}$ is denoted by $|\mathcal{M}|$. $\mathbb{R}^n$ and $\mathbb{R}^+_{n \times m}$ denote the sets of $n$-dimensional vectors with real and positive real entries, respectively. We drop the superscripts for $n = 1$. $\mathbb{R}^{n \times m}$ denotes the set of $n \times m$ matrices with entries in $\mathbb{R}$. $\mathbb{S}^n$, $\mathbb{S}^+_n$, and $\mathbb{S}^+_{n \times m}$ denote the sets of $n \times n$ symmetric, positive semi-definite, and positive definite matrices, respectively. $\mathbb{D}^+_{n \times n}$ denotes the set of diagonal matrices in $\mathbb{S}^+_{n \times n}$. $0$ denotes a matrix of zeros of appropriate dimensions. $1_n$ and $I_n$ denote the vectors of ones of length $n$ and the identity matrix of size $n \times n$, respectively. For a given vector $v \in \mathbb{R}^n$, $\text{diag}(v)$ denotes a diagonal matrix with the elements of $v$ on its diagonal. For a given matrix $V \in \mathbb{R}^{n \times n}$, $V^T$ and $\ker(V)$ denote the transpose and nullspace of $V$, respectively. If $n > m$ and $V$ is full column rank, then $V^T$ denotes a left inverse of $V$, i.e., $V^TV = I_m$. If $n = m$, then $\text{tr}(V)$ denotes the trace of $V$. If $n = m$ and $V$ is nonsingular, then $V^{-1}$ denotes the inverse of $V$. For $n = m$, we write $V \succeq 0$ and $V \succ 0$ to mean that $V \in \mathbb{S}^+_n$ and $V \in \mathbb{S}^+_{n \times m}$, respectively. Given matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{q \times n}$, the triple $(A, B, C)$ denotes a state-space representation of a strictly proper, continuous-time, linear time-invariant (LTI) system $\Sigma$ of order $n$. The transfer function matrix of $\Sigma$ is given by $G(s) = C(sI_n - A)^{-1}B$, where $s$ denotes the Laplace variable. The system $\Sigma$ is bounded input-bounded output (BIBO) stable if and only if the poles of $\Sigma$ have negative real parts. If $\Sigma$ is BIBO stable, then $|\Sigma|_{H_{\infty}}$ and $|\Sigma|_{H_2}$ denote the (finite) $\mathcal{H}_{\infty}$-norm and $\mathcal{H}_2$-norm of $\Sigma$, respectively.

B. Graph Theory

$\mathcal{G}(\mathcal{V}, \mathcal{E}, E, W)$ denotes a simple, undirected, weighted, and connected graph $\mathcal{G}$ with a set of nodes $\mathcal{V} = \{1, 2, \ldots, n\}$ and a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The elements of $\mathcal{E}$ consist of uniquely labeled, unordered pairs $(i, j)$ that indicate the existence of an edge with label $l$ between nodes $i$ and $j$, where $i, j \in \mathcal{V}$ and the label $l$ is in $\{1, 2, \ldots, |\mathcal{E}|\}$. The nodal time-scales matrix $E \in \mathbb{D}^+_{n \times n}$ is defined as $E := \text{diag}(e_1, e_2, \ldots, e_n)$, where $e_i$ denotes the time-scale associated with node $i \in \mathcal{V}$. Likewise, the edge weights matrix $W \in \mathbb{D}^+_{|\mathcal{E}|}$ is defined as $W := \text{diag}(w_1, w_2, \ldots, w_{|\mathcal{E}|})$, where $w_l$ denotes the edge weight associated with edge $l \in \{1, 2, \ldots, |\mathcal{E}|\}$. The incidence matrix $D \in \mathbb{R}^{n \times |\mathcal{E}|}$ characterizes the incidence relation between the nodes and edges of $\mathcal{G}$. $D$ is defined such that $[D]_{ij} = +1$ if edge $l$ is directed from node $i$, $-1$ if edge $l$ is directed towards node $i$, and 0 otherwise. For undirected graphs, the incidence matrix is generated by assigning arbitrary orientations to the edges of $\mathcal{G}$. The (unscaled) Laplacian matrix $L$ of $\mathcal{G}$ is defined as

\[ L := DWD^T \in \mathbb{S}^+_n \]

satisfies $\ker(L) = \text{span}\{1_n\}$, $L_{1n} = 0$, and $I^T_nL = 0$, where $\text{span}(v)$ denotes the set of vectors of the form $\alpha v$ with $\alpha \in \mathbb{R}$ [11]. Also, for all $i, j \in \mathcal{V}$, $[L]_{ij} \leq 0$ if $i \neq j$ and $[L]_{ii} > 0$ otherwise.

A graph clustering of $\mathcal{G}$ is a partition of $\mathcal{V}$ into $r$ nonempty, mutually exclusive subsets or “clusters” $C_1, C_2, \ldots, C_r$ such that $C_1 \cup C_2 \cup \cdots \cup C_r = \mathcal{V}$. The graph clustering can be characterized by a binary matrix $\Pi \in \mathbb{R}^{r \times |\mathcal{E}|}$ such that $[\Pi]_{ij} = 1$ if $i \in C_j$ and 0 otherwise. By definition, it follows that $\Pi 1_r = 1_n$. A useful property of the clustering matrix $\Pi$ is that the matrix $L = DWD^T \in \mathbb{S}^+_n$, where $D$ is constructed using $D$ and $\Pi$, can be interpreted as a Laplacian matrix that characterizes a simple, undirected, weighted, and connected reduced graph $\hat{\mathcal{G}}(\hat{\mathcal{V}}, \hat{\mathcal{E}}, \hat{E}, \hat{W})$ that has fewer nodes and edges than $\mathcal{G}$ [11]. In particular, $\hat{\mathcal{V}}$ is the reduced set of nodes such that $|\hat{\mathcal{V}}| = r$, where $r \leq n$ by definition, and the matrices $\hat{E} \in \mathbb{D}^+_{r \times r}$, $\hat{W} \in \mathbb{D}^+_{r \times |\hat{\mathcal{E}}|}$, and $\hat{D} \in \mathbb{R}^{r \times |\hat{\mathcal{E}}|}$ are the reduced nodal time-scales, reduced edge weights, and reduced incidence matrices of $\hat{\mathcal{G}}$, respectively. The matrices $\hat{E}$ and $\hat{W}$ can be assigned independently of the choice of the clustering matrix $\Pi$, which is a feature exploited in the present work for optimizing the reduced order network system. The reduced graph $\hat{\mathcal{G}}$ is obtained from $\mathcal{G}$ as follows. First, the edges between the nodes within the same cluster are removed and the nodes within each cluster are aggregated into a single node. Second, if there is at least one edge between any pairs of nodes in different clusters, then a single edge between the corresponding clusters is retained. Otherwise, no edge exists between the two clusters. $\hat{D}$ is algebraically obtained by removing all duplicate and zero columns from $\Pi^T D$.

C. Problem Setup

Consider a consensus network system $\Sigma$ consisting of time-scaled single integrator agents with weighted interconnections evolving over a network described by a simple, undirected, weighted, and connected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, E, W)$. The agents and the interconnections between them are represented by the nodes and edges of $\mathcal{G}$, respectively. The dynamics of this network system are given by

\[ \Sigma : \begin{cases} \dot{E}x(t) = -Lx(t) + Fu(t), \\ y(t) = Hx(t), \end{cases} \]

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, and $y(t) \in \mathbb{R}^q$ are the state, input, and output vectors at time $t \geq 0$, respectively. $L = DWD^T \in \mathbb{S}^+_n$ and $D \in \mathbb{R}^{n \times |\mathcal{E}|}$ are the Laplacian and incidence matrices of $\mathcal{G}$, $F \in \mathbb{R}^{n \times p}$ and $H \in \mathbb{R}^{q \times n}$ are the input and output matrices of $\Sigma$, respectively.

Given a clustering of the graph $\mathcal{G}$ with corresponding characteristic clustering matrix $\Pi \in \mathbb{R}^{n \times r}$, the problem addressed in this paper is to find a reduced consensus network system $\hat{\Sigma}$ that 1) consists of $r$ single integrator agents whose dynamics evolve over the network described by the reduced, simple, undirected, weighted, and connected graph $\hat{\mathcal{G}}(\hat{\mathcal{V}}, \hat{\mathcal{E}}, \hat{E}, \hat{W})$ and 2) approximates the input-output behavior of the original network system $\Sigma$. The dynamics of the reduced network system are given by

\[ \hat{\Sigma} : \begin{cases} \hat{E}\hat{x}(t) = -\hat{L}\hat{x}(t) + \beta\hat{F}u(t), \\ \hat{y}(t) = \alpha\hat{H}\hat{x}(t), \end{cases} \]

where $\hat{x}(t) \in \mathbb{R}^r$ and $\hat{y}(t) \in \mathbb{R}^q$ are the state and output vectors of the reduced system, respectively. $\hat{L} = \hat{D}\hat{W}\hat{D}^T \in \mathbb{S}^+_n$ and $\hat{D} \in \mathbb{R}^{r \times |\hat{\mathcal{E}}|}$ are the Laplacian and incidence matrices.
of the reduced graph $\hat{G}$. As mentioned in Section II-B, $\hat{D}$ is obtained by removing the duplicate and zero columns from $\Pi^T \Pi$. $\hat{F} = \Pi^T F \in \mathbb{R}^{r \times p}$ and $\hat{H} = \Pi \Pi^T \in \mathbb{R}^{q \times r}$ are the reduced system input and output matrices, respectively. The reduced order system $\hat{\Sigma}$ defined in (2) is parameterized by the reduced edge weight matrix $\hat{W}$ and the reduced nodal time-scale matrix $\hat{E}$. The $H_{\infty}$-norm and the $H_2$-norm of the error system $\Sigma - \hat{\Sigma}$ will be used to quantify the reduction error. Specifically, the diagonal entries in the matrices $\hat{W}$ and $\hat{E}$ are unknown parameters that will be chosen to minimize the approximation errors $\|\Sigma - \hat{\Sigma}\|_{H_{\infty}}$ and $\|\Sigma - \hat{\Sigma}\|_{H_2}$. Finally, $\alpha$ and $\beta$ are scalars that will be chosen appropriately to ensure that the error system $\Sigma - \hat{\Sigma}$ is BIBO stable for all $\hat{W} \in \mathbb{D}_{++}^{|E|}$ and $\hat{E} \in \mathbb{D}_{++}^{|E|}$. We formalize the problems addressed in this paper as follows.

Given the network system $\Sigma$ defined in (1) and the parameterized reduced network system $\hat{\Sigma}$ defined in (2) for a predetermined clustering matrix $\Pi \in \mathbb{R}^{n \times r}$, solve the following optimization problems:

$$\min_{\hat{E} \in \mathbb{D}_{++}^{|E|}, \hat{W} \in \mathbb{D}_{++}^{|E|}} \|\Sigma - \hat{\Sigma}\|_{H_{\infty}},$$  
(3)

$$\min_{\hat{E} \in \mathbb{D}_{++}^{|E|}, \hat{W} \in \mathbb{D}_{++}^{|E|}} \|\Sigma - \hat{\Sigma}\|_{H_2},$$  
(4)

Although the PGP paradigm is applied in most clustering-based model reduction techniques, applying this paradigm limits the flexibility of constructing a more accurate reduced model once a clustering has been chosen [12]. Namely, in this paradigm, the reduced system matrices are directly determined via projection using the characteristic clustering matrix $\Pi$, i.e., $\hat{E} = \Pi^T \Pi$ and $\bar{L} = \Pi^T \Pi \Pi$. In [12], Problem (4) is solved for fixed $\hat{E}$, with $\hat{E} = \Pi^T \Pi$, whereby the $H_2$-norm of the error system is minimized by optimizing over the matrix $\hat{W}$. The contribution of this paper is thus to address both Problems (3) and (4), wherein both $\hat{E}$ and $\hat{W}$ are free parameter matrices to be optimized. The setup in this paper is limited to undirected graphs; and future work will look at treating both problems for the case of directed graphs, see also the related work in [13].

### III. MAIN RESULTS

This section presents the main results of the paper. In Section III-A an appropriate choice of the scalars $\alpha$ and $\beta$ in (2) is found that ensures that the error system $\Sigma - \hat{\Sigma}$ is BIBO stable for all $\hat{W} \in \mathbb{D}_{++}^{|E|}$ and $\hat{E} \in \mathbb{D}_{++}^{|E|}$. In Section III-B and Section III-C $H_{\infty}$- and $H_2$-based optimization algorithms are proposed to search for suboptimal edge weights and nodal time-scales for $\hat{\Sigma}$ that minimize $\|\Sigma - \hat{\Sigma}\|_{H_{\infty}}$ and $\|\Sigma - \hat{\Sigma}\|_{H_2}$, respectively.

#### A. BIBO Stability of $\Sigma - \hat{\Sigma}$

Consider the error system $\Sigma_e = \Sigma - \hat{\Sigma}$. A realization for $\Sigma_e$ is given by the triple $(A_e, B_e, C_e)$, where

$$A_e = \begin{bmatrix} -E^{-1}L & 0 \\ 0 & -E^{-1} \bar{L} \end{bmatrix}, \quad B_e = \begin{bmatrix} E^{-1}F \\ \beta \bar{E}^{-1} \bar{F} \end{bmatrix}, \quad C_e = \begin{bmatrix} H & -\alpha \hat{H} \end{bmatrix}.$$  

The corresponding $q \times p$ transfer function matrix is given by $\Sigma_e(s) = C_e(s I_{(n+1)} - A_e)^{-1} B_e$. By construction, and from the properties of the weighted and scaled Laplacian $E^{-1}L$ of $\hat{G}$ and the reduced weighted and scaled Laplacian $E^{-1}L$ of $\hat{G}$ derived in [4], the spectrum of $A_e$ contains two zero eigenvalues, with the remaining eigenvalues being non-negative real. The following lemma provides a condition on $\alpha$ and $\beta$ to ensure that the error system $\Sigma_e$ is BIBO stable, thereby guaranteeing that the approximation errors $\|\Sigma_e\|_{H_{\infty}}$ and $\|\Sigma_e\|_{H_2}$ are bounded for all $\hat{W} \in \mathbb{D}_{++}^{|E|}$ and $\hat{E} \in \mathbb{D}_{++}^{|E|}$.

**Lemma 1:** Consider the network system $\Sigma$ defined in (1) and the reduced network system $\hat{\Sigma}$ defined in (2) for a given clustering matrix $\Pi \in \mathbb{R}^{n \times r}$. If $\alpha \beta = \text{tr}(\hat{E})/\text{tr}(E)$, then $\Sigma_e = \Sigma - \hat{\Sigma}$ is BIBO stable for all $\hat{W} \in \mathbb{D}_{++}^{|E|}$ and $\hat{E} \in \mathbb{D}_{++}^{|E|}$.

**Proof:** A similar realization $(T^{-1}A_eT, T^{-1}B_e, C_eT)$ for the error system $\Sigma_e$ is first found that isolates the zero eigenvalues of $A_e$ from the remaining ones. Then, it is shown that the poles corresponding to the zero eigenvalues of $A_e$ cancel out in the matrix transfer function of the error system when $\alpha \beta = \text{tr}(\hat{E})/\text{tr}(E)$. As a result, in this case, the matrix transfer function of the error system contains poles lying exclusively in the left-half of the $s$-plane for all $\hat{W} \in \mathbb{D}_{++}^{|E|}$ and $\hat{E} \in \mathbb{D}_{++}^{|E|}$, which ensures the BIBO stability of the error system. The sought nonsingular similarity transformation matrix $T$ and its inverse are constructed as follows:

$$T = \begin{bmatrix} \frac{1}{\text{tr}(E)} I_n & 0 & E^{-1} S_n & 0 \\ 0 & \frac{1}{\text{tr}(E)} I_r & 0 & \hat{E}^{-1} S_r \end{bmatrix},$$  
$$T^{-1} = \begin{bmatrix} I_n & 0 & E & 0 \\ 0 & I_r & 0 & \hat{E} \end{bmatrix},$$

where the matrices

$$S_n = -I_{n-1} I_n^{r-1}, \quad S_n^L = (S_n^T E^{-1} S_n)^{-1} S_n^T E^{-1},$$

satisfy $I_n^T S_n = 0$ and $S_n^L S_n = I_{n-1}$, and the matrices $S_c$ and $S_e^L$ are defined similarly. The similarity transformation $T$ is now applied. The resulting expression for $\Sigma_e(s)$ obtained from the realization $(T^{-1}A_eT, T^{-1}B_e, C_eT)$ is given by

$$\Sigma_e(s) = C_e T (s I_{(n+1)} - T^{-1} A_e T)^{-1} T^{-1} B_e = \frac{\hat{C}_e}{\hat{C}_e} \begin{bmatrix} (s I_{1}^{1})^{-1} 0 \\ 0 (s I_{m-r} - \hat{A}_e)^{-1} \hat{B}_e \end{bmatrix} = -\frac{\hat{C}_e \hat{B}_e}{s} + C_e (s I_{m-r} - \hat{A}_e)^{-1} \hat{B}_e = \frac{1}{s} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} H_1 n_1^{T} F - \frac{\alpha \beta}{\text{tr}(E)} \hat{H} 1, \frac{1}{s} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \hat{F} + \Sigma_e(s),$$

where $\Sigma_e(s) = C_e (s I_{m-r} - \hat{A}_e)^{-1} \hat{B}_e$,

$$\hat{A}_e = \begin{bmatrix} -S_n^L E^{-1} S_n & 0 \\ 0 & -S_r^L E^{-1} S_r \end{bmatrix}, \quad \hat{B}_e = \begin{bmatrix} S_n^L F \\ \beta S_r^L \hat{F} \end{bmatrix}, \quad \hat{C}_e = \begin{bmatrix} H E^{-1} S_n - \alpha \hat{H} E^{-1} S_r \end{bmatrix},$$

and $m = n + r - 2$. Thus, the applied similarity transformation isolates the zero eigenvalues of $A_e$, which means that $A_e$ is Hurwitz and $\Sigma_e$ is BIBO stable. From the properties of $\Pi$ and the definitions of $F$ and $H$, it follows that
Lemma 1 further guarantees that \(\|\Sigma - \hat{\Sigma}\|_{H_\infty}\) and \(\|\hat{\Sigma}\|_{H_\infty}\) are invariant under the specific choices of \(\alpha\) and \(\beta\) as long as \(\alpha \beta = \text{tr}(\hat{E})/\text{tr}(E)\). For the remainder of the paper, we make the convenient choice of \(\alpha = 1\) and \(\beta = \text{tr}(\hat{E})/\text{tr}(E)\).

### B. \(\|\Sigma - \hat{\Sigma}\|_{H_\infty}\) Optimization

To solve Problem (3), we first obtain a characterization of \(\|\Sigma_e\|_{H_\infty}\) in terms of \(\hat{W}\) and \(\hat{E}\).

**Theorem 1:** Consider the network system \(\Sigma\) defined in (1). There exists a reduced network system \(\hat{\Sigma}\) defined as in (2) such that \(\|\Sigma_e\|_{H_\infty} < \gamma\) if and only if there exist matrices \(\hat{X} \in \mathbb{S}_{++}^m, \hat{W} \in \mathbb{D}^{|\hat{E}|}_{++}\), and \(\hat{E} \in \mathbb{D}^{|E|}_{++}\), and a sufficiently small \(\theta \in \mathbb{R}_{++}\) such that the following inequality is satisfied:

\[
\Phi_\infty(\hat{X}, \hat{\gamma}, \hat{W}, \hat{E}) = \psi_\infty(\hat{X}, \hat{\gamma}) + \phi_\infty(\hat{W}, \hat{E}) < 0, \tag{6}
\]

where \(\hat{\gamma} = \hat{\theta}\gamma\), the matrix-valued mappings \(\psi_\infty : \mathbb{S}_{++}^m \times \mathbb{R}_{++} \to \mathbb{R}^{2m+p+q}\) and \(\phi_\infty : \mathbb{D}^{|\hat{E}|}_+ \times \mathbb{D}^{|E|}_+ \to \mathbb{R}^{2m+p+q}\) are given by

\[
\psi_\infty(\hat{X}, \hat{\gamma}) = \begin{bmatrix}
\hat{X}^T \hat{X} + \hat{X} \hat{A}_{en} \hat{X} \hat{B}_{en} \hat{\theta} \hat{C}_{en}^T \hat{X} \hat{J} \\
\hat{B}_{en}^T \hat{X} \\
\hat{\theta} \hat{C}_{en} \\
\hat{J}^T \hat{X}
\end{bmatrix},
\]

\[
\phi_\infty(\hat{W}, \hat{E}) = \begin{bmatrix}
-\hat{A}_{en}^T \hat{X} - \hat{A}_{en} \hat{B}_{en} \hat{\theta} \hat{C}_{en} \hat{A}_{en}^T \\
-\hat{B}_{en} \hat{B}_{en} \hat{\theta} \hat{C}_{en} \hat{B}_{en} \\
\hat{\theta} \hat{C}_{en} \\
\hat{A}_{en} \hat{B}_{en}
\end{bmatrix} - \begin{bmatrix}
0 \\
0 \\
0 \\
-\hat{I}_m
\end{bmatrix}, \tag{7}
\]

and the matrices \(\hat{A}_{en} \in \mathbb{R}^{m \times m}, \hat{A}_{er} \in \mathbb{R}^{m \times m}, \hat{B}_{en} \in \mathbb{R}^{m \times p}, \hat{B}_{er} \in \mathbb{R}^{m \times p}, \hat{C}_{en} \in \mathbb{R}^{m \times r}, \hat{C}_{er} \in \mathbb{R}^{m \times r}, \) and \(\hat{J} \in \mathbb{R}^{m \times m}\) are defined in equations (8) through (11).

**Proof:** Recall the terms introduced in the proof of Lemma 1. First, we define the matrix decompositions \(\hat{A}_e = \hat{A}_{en} + \hat{J} \hat{A}_{er}, \hat{B}_e = \hat{B}_{en} + \hat{J} \hat{B}_{er},\) and \(\hat{C}_e = \hat{C}_{en} + \hat{C}_{er},\) where

\[
\hat{A}_{en} = \begin{bmatrix}
-S_n^T \hat{L} E^{-1} S_n \\
0 \\
0 \\
0
\end{bmatrix}, \quad \hat{A}_{er} = \begin{bmatrix}
S_n^T \hat{F} \\
0 \\
0 \\
\beta \hat{S}_n^T \hat{F}
\end{bmatrix},
\]

\[
\hat{B}_{en} = \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}, \quad \hat{B}_{er} = \beta \hat{S}_n \hat{F}, \tag{9}
\]

\[
\hat{C}_{en} = \begin{bmatrix}
\hat{H} \hat{E}^{-1} S_n \\
0
\end{bmatrix}, \quad \hat{C}_{er} = \begin{bmatrix}
0 \\
0 \\
0 \\
-\alpha \hat{H} \hat{E}^{-1} S_r
\end{bmatrix},
\]

\[
\hat{J} = \begin{bmatrix}
0 \\
I_{r-1}
\end{bmatrix}. \tag{10}
\]

Next, from the Bounded Real Lemma [16], it follows that \(\|\Sigma_e\|_{H_\infty} = \|\Sigma_e\|_{H_\infty} < \gamma\) if and only if there exists \(X \in \mathbb{S}_{++}^m\) such that

\[
\begin{bmatrix}
\hat{A}_e^T \hat{X} + \hat{X} \hat{A}_e & \hat{X} \hat{B}_e & \hat{C}_e^T \\
\hat{B}_e^T \hat{X} & -\gamma I_p & 0 \\
\hat{C}_e & 0 & -\gamma I_q
\end{bmatrix} < 0. \tag{12}
\]

By the Schur complement formula and for a sufficiently large \(\theta > 0\), (12) is equivalent to the following inequality:

\[
\begin{bmatrix}
\hat{A}_e^T \hat{X} + \hat{X} \hat{A}_e & \hat{X} \hat{B}_e & \hat{C}_e^T & \hat{X} J \\
\hat{B}_e^T \hat{X} & -\gamma I_p & 0 & 0 \\
\hat{C}_e & 0 & -\gamma I_q & 0 \\
\hat{J}^T \hat{X} & 0 & 0 & -\theta I_m
\end{bmatrix} < 0. \tag{13}
\]

Based on the decompositions of \(\hat{A}_e, \hat{B}_e,\) and \(\hat{C}_e\) defined in (8) through (11), a nonsingular transformation matrix \(U\) is applied to (13) to decouple the variables \(X\) and \(\gamma\) from the variables \(\hat{W}\) and \(\hat{E}\). Namely, let

\[
U = \begin{bmatrix}
I_m & 0 & 0 & 0 \\
0 & I_p & 0 & 0 \\
0 & 0 & I_q & 0 \\
-\hat{A}_{er} & -\hat{B}_{er} & 0 & I_m
\end{bmatrix}.
\]

Then, pre- and post-multiplying both sides of (13) by \(U^T\) and \(U\), respectively, dividing both sides by \(\theta > 0\), and performing the substitutions \(\theta = 1/\theta\), \(\hat{X} = (1/\theta) X > 0\), and \(\hat{\gamma} = (1/\theta) \gamma\), yield the equivalent matrix inequality in (9).

**Theorem 1** gives a characterization of the \(H_\infty\)-norm of the error system \(\Sigma_e\) that will be used to formulate an optimization problem for the selection of the edge weights and nodal time-scales of \(\Sigma\). Specifically, we need to find \(\hat{X} \in \mathbb{S}_{++}^m, \hat{W} \in \mathbb{D}^{|\hat{E}|}_+\), and \(\hat{E} \in \mathbb{D}^{|E|}_+\) that minimize \(\hat{\gamma}\) such that (6) holds for a given \(\theta > 0\).

In [13], an analogous decoupling procedure to the one used in the proof of Theorem 1 is applied on a matrix inequality (counterpart of (13)) that characterizes the \(H_2\)-norm of the error system. The goal therein is to only decouple \(X\) from \(\hat{W}\) (since \(\hat{E}\) is fixed). In that case, the left-hand-side of the resulting inequality takes the form of the difference of two positive semidefinite-convex matrix-valued mappings, also known as a psd-convex-concave decomposition of the counterpart of \(\Phi_\infty\) therein. Thus, an optimization problem for the selection of the edge weights is formulated therein that can be solved by following the procedure in [15]. This procedure consists of iteratively solving a sequence of convex optimization problems and is guaranteed to converge to a locally optimal solution of the nonconvex problem of interest. Namely, starting from a feasible solution to the original problem, a convex inequality constraint is obtained by replacing the psd-concave term in the original inequality by its linearization about this initial solution. A convex optimization problem is then solved in which the original inequality constraint is replaced by its convex counterpart. In the second iteration, the psd-concave term in the original inequality is linearized about the solution of the convex optimization problem solved in the first iteration. The process repeats until convergence. As per Lemma 1, all \(\hat{W} \in \mathbb{D}^{|\hat{E}|}_+\) are feasible starting points for this iterative algorithm. As a 'good' starting point, the solution obtained from applying the PGP paradigm can be used to initialize the algorithm, which ensures that the obtained results are at least as good as the ones from this classic paradigm.
In our case, Lemma 1 ensures that all \( \hat{W} \in \mathbb{D}_{++}^{\hat{n}} \) and \( \hat{E} \in \mathbb{D}_{++}^{\hat{n}} \) are feasible starting points for any similar iterative algorithm to be proposed. However, although \( \psi_\infty \) is psd-convex in \((\hat{X}, \hat{\gamma})\) and \( \phi_\infty \) is psd-concave in \( \hat{W} \), \( \phi_\infty \) is not psd-concave in \( \hat{E} \). In what follows, we propose to solve Problem 3 by alternating between optimizing the edge weights for fixed nodal time-scales and optimizing the nodal time-scales for fixed edge weights. In particular, for fixed \( \hat{E} \), \( \phi_\infty(\hat{W}) \) is psd-concave in \( \hat{W} \), and so, to optimize the edge weights for fixed time-scales, we follow a procedure similar to the one in [13]. However, for fixed \( \hat{W} \), \( \psi_\infty(\hat{E}) \) is not psd-concave in \( \hat{E} \). Nonetheless, constraint 6 in this case can still be modified and rewritten in a form amenable to solution by the procedure in [15]. Namely, we treat \( \hat{E}^{-1} \) as the variable instead of \( \hat{E} \) and introduce an additional matrix-valued variable \( Z \) such that \( Z = \hat{E} \) to rewrite \( B_{\gamma} \) as a function of \( Z \) (not of \( \hat{E}^{-1} \)). By doing so, inequality 6 is transformed into the following set of constraints:

\[
\psi_\infty(\hat{X}, \hat{\gamma}) + \phi_\infty(\hat{E}^{-1}, Z) < 0, \tag{14}
\]

\[
\begin{bmatrix}
\hat{E}^{-1} & I_r \\
I_r & Z
\end{bmatrix} \succeq 0, \tag{15}
\]

\[
Z - (\hat{E}^{-1})^{-1} \preceq 0. \tag{16}
\]

In other words, the added equality constraint \( Z = \hat{E} \) is split into \( \hat{E} \preceq Z \) and \( Z \preceq \hat{E} \). By application of the Schur complement formula, the inequality \( \hat{E} \preceq Z \) is transformed into the linear matrix inequality in \((\hat{E}^{-1}, Z)\) in (15). The inequality \( Z \preceq \hat{E} \) is written as \( Z - (\hat{E}^{-1})^{-1} \preceq 0 \) as in (16), where the left-hand-side takes the form of the difference of two psd-convex matrix-valued mappings in \( \hat{E} \) and \( \hat{E}^{-1} \), namely, \( Z \) and \( (\hat{E}^{-1})^{-1} \). Thus, by introducing the additional variable \( Z \) and expressing the constraint (6) as the equivalent set of constraints (14) and (16), we obtain the following result.

**Lemma 2:** For any fixed \( \hat{W} \in \mathbb{D}_{++}^{\hat{n}} \), the matrix-valued mapping \( \phi_\infty \) in (14) is psd-concave in \((\hat{E}^{-1}, Z)\). To prove Lemma 2, we first need to prove the following proposition.

**Proposition 1:** Consider the full and reduced order systems \( \Sigma \) and \( \bar{\Sigma} \), and assume that the nodal time-scale matrices are partitioned as in \( \hat{E} = \begin{bmatrix} \hat{E} & 0 \\ 0 & e_n \end{bmatrix} \) and \( \hat{E} = \begin{bmatrix} \hat{E} & 0 \\ 0 & e_r \end{bmatrix} \), where \( \hat{E} \in \mathbb{D}_{++}^{\hat{n}} \) and \( \hat{E} \in \mathbb{D}_{++}^{\hat{r}} \), respectively. Define \( S_n, S_r, S_n^L, \) and \( S_r^L \) as in (9). Then,

\[
\begin{align*}
\text{tr}(\hat{E}) S_n^L &= \begin{bmatrix} \hat{E} & 0 \\ 0 & \hat{E} \end{bmatrix}, \tag{17} \\
S_n^L D &= \begin{bmatrix} I_{n-1} & 0 \\ 0 & I_r \end{bmatrix}, \tag{18} \\
\text{tr}(\hat{E}) S_r^L &= \begin{bmatrix} \hat{E} & 0 \\ 0 & \hat{E} \end{bmatrix}, \tag{19} \\
S_r^L D &= \begin{bmatrix} I_{r-1} & 0 \\ 0 & I_r \end{bmatrix}. \tag{20}
\end{align*}
\]

**Proof:** We prove (17) and (18), (19) and (20) are proved similarly. From the definitions of \( S_n \) and \( S_n^L \), we have

\[
S_n^L = (S_n^T \hat{E}^{-1} S_n)^{-1} S_n^T \hat{E}^{-1} = (\begin{bmatrix} I_{n-1} & 0 \\ 0 & I_r \end{bmatrix})^{-1} S_n^T \hat{E}^{-1} = \begin{bmatrix} \hat{E} & 0 \\ 0 & \hat{E} \end{bmatrix}.
\]

where the last equality follows from the matrix inversion lemma. Multiplying both sides by \( \text{tr}(\hat{E}) \) yields (17). Equation (17) is then used to prove equation (18). Namely,

\[
S_n^L D = \left( \frac{1}{\text{tr}(\hat{E})} \hat{E} - [I_{n-1} \ 0] \right) D = \frac{1}{\text{tr}(\hat{E})} \hat{E} (JD) - [I_{n-1} \ 0] D,
\]

where \( J \) is the matrix of ones of size \((n-1) \times n\). The structure of the incidence matrix \( D \) implies that \( JD = 0 \), which completes the proof.

**Proof of Lemma 2** By Proposition 1, \( \text{tr}(\hat{E}) S_n^L \) is linear in \( \hat{E} \) and \( S_n^L D \) is independent of \( \hat{E} \) for all \( \hat{E} \in \mathbb{D}_{++}^{\hat{n}} \). As such, with the assumed choice of \( \alpha = 1 \) and \( \beta = \text{tr}(\hat{E})/\text{tr}(E) \), \( A_{\gamma} \) is linear in \( \hat{E}^{-1} \), \( \beta = \text{tr}(\hat{E})/\text{tr}(E) \) is linear in \( \hat{E} \), and \( \beta = \text{tr}(\hat{E})/\text{tr}(E) \) is linear in \( \hat{E}^{-1} \). Thus, for fixed \( \hat{W} \), the structure of \( \phi_\infty \), shown in (17), can be leveraged by introducing the additional variable \( Z \) such that \( Z = \hat{E} \) and expressing \( B_{\gamma} \) in terms of \( Z \) to get \( \phi_\infty(\hat{E}^{-1}, Z) \) as in (17). Define the linear matrix-valued mapping \( \mathcal{L}(\hat{E}^{-1}, Z) = [A_{\gamma} \ B_{\gamma}] \). Since the off-diagonal block matrices of \( -\phi_\infty(\hat{E}^{-1}, Z) \) are linear in \( \hat{E}^{-1}, Z \), it is enough to check the psd-convexity of the upper-left block of \( -\phi_\infty(\hat{E}^{-1}, Z) \) given by \( \mathcal{L}(\hat{E}^{-1}, Z)^T \mathcal{L}(\hat{E}^{-1}, Z) \). This term is indeed psd-convex in \((\hat{E}^{-1}, Z) \) since it corresponds to the composition of the psd-convex function \( X^T X \) with the linear mapping \( \mathcal{L}(\hat{E}^{-1}, Z) \) (17).

Based on the above discussion, two optimization schemes that tune the weights and time-scales of the reduced system independently are proposed. Depending on whether \( \hat{E} \) is fixed and \( \hat{W} \) is solved for or vice versa, a different formulation of the problem of minimizing the \( \psi_\infty \)-norm of the error system is solved by following the iterative procedure in [15]. The iterative optimization algorithms for the \( \psi_\infty \)-tuning of edge weights for fixed nodal time-scales and the \( \psi_\infty \)-tuning of nodal time-scales for fixed edge weights are given in Algorithm 1 and Algorithm 2 respectively. Algorithm 1 outputs an \( \psi_\infty \)-suboptimal matrix of edge weights \( \hat{W} \) for a given matrix of time-scales \( \hat{E} \) by iteratively solving the optimization problem (P1) defined as follows:

\[
\begin{align*}
\min_{\hat{X}, \hat{W}, \hat{\gamma}} \quad & \hat{\gamma} \\
\text{s.t.} \quad & \psi_\infty(\hat{X}, \hat{\gamma}) + \phi_\infty(\hat{W}) < 0, \quad \text{(P1)}
\end{align*}
\]

where \( k \) denotes the iteration counter. Algorithm 2 outputs an \( \psi_\infty \)-suboptimal matrix of nodal time-scales \( \hat{E} \) for a given matrix of edge weights \( \hat{W} \) by iteratively solving the optimization problem (P2) defined as follows:

\[
\begin{align*}
\min_{\hat{X}, \hat{E}^{-1}, Z, \hat{\gamma}} \quad & \hat{\gamma} \\
\text{s.t.} \quad & \psi_\infty(\hat{X}, \hat{\gamma}) + \phi_\infty(\hat{W}) < 0, \quad \text{(P2)}
\end{align*}
\]

The linearized terms \( \hat{\gamma}(\hat{W}) \), \( \hat{\gamma}(\hat{E}^{-1}, Z) \), and \( f(k)(\hat{E}^{-1}) \)
in (P1) and (P2) at iteration \( k \) are given by

\[
\phi_0(k)(\hat{W}) = \phi_\infty(\hat{W}(k)) + D\phi_\infty(\hat{W}(k))[\hat{W} - \hat{W}(k)],
\]

\[
\phi_\infty(\hat{E}^{-1}, Z) = \phi_\infty(\hat{E}^{-1}(k), Z(k))
\]

\[
+ D\phi_\infty(\hat{E}^{-1}(k), Z(k))[(\hat{E}^{-1} - \hat{E}^{-1}(k), Z - Z(k))],
\]

\[
f^{(k)}(\hat{E}^{-1}) = f\left(\hat{E}^{-1}(k)\right) + Df\left(\hat{E}^{-1}(k)\right)[\hat{E}^{-1} - \hat{E}^{-1}(k)],
\]

and \( f(X) = X^{-1} \). Let \( \hat{w}_i(k) \) be a vector such that \( \hat{W}(k) = \text{diag}(\hat{w}_i(k)) \) and denote its \( i \)-th component by \( [\hat{w}_i(k)] \). Then, the derivative operator in (21) at iteration \( k \) satisfies

\[
D\phi_\infty(\hat{W}(k))[\hat{W} - \hat{W}(k)] = \sum_{i=1}^{\lvert \mathcal{E} \rvert} \left[ \hat{w}_i - [\hat{w}_i(k)] \right] \frac{\partial \phi_\infty}{\partial \hat{w}_i}(\hat{w}_i).
\]

The other derivative operators are defined similarly. In Problem (P2), \( \lambda \) should be ideally set to 1. In practice, however, \( \lambda \) is set to \( 1 + \nu \) for a small \( \nu > 0 \) to avoid numerical problems when solving Problem (P2).

Algorithm 1: \( H_\infty \) Iterative Edge Weight Tuning

Input : \( E, \hat{E}, \hat{W}_0, L, F, H, \Pi, \theta, \hat{\gamma}_0, i_{tot}, \varepsilon \)

Output: \( \hat{W}_* \)

Initialize \( \hat{W}_0 = \hat{W} \) and \( \hat{\gamma}_0 = \hat{\gamma}_0 \)

Set iteration step \( k \leftarrow 0 \)

while \( k \leq i_{tot} \) and \( \hat{\gamma}_k - \hat{\gamma}_{k+1} \geq \varepsilon \) do

\( \text{Solve (P1) to obtain } \hat{W}_*, \hat{X}_*, \text{ and } \hat{\gamma}_s \)

Update \( k \leftarrow k + 1 \)

Update \( \hat{W}_k \leftarrow \hat{W}_s \) and \( \hat{\gamma}_k \leftarrow \hat{\gamma}_s \)

end

Algorithm 2: \( H_\infty \) Iterative Nodal Time-scale Tuning

Input : \( E, \hat{E}_0, \hat{W}, L, F, H, \Pi, \theta, \hat{\gamma}_0, i_{tot}, \varepsilon \)

Output: \( \hat{E}_* \)

Initialize \( (\hat{E}_0^{-1}, Z_0) = (\hat{E}_0^{-1}, 0) \) and \( \hat{\gamma}_0 = \hat{\gamma}_0 \)

Set iteration step \( k \leftarrow 0 \)

while \( k \leq i_{tot} \) and \( \hat{\gamma}_k - \hat{\gamma}_{k+1} \geq \varepsilon \) do

\( \text{Solve (P2) to obtain } \hat{E}_k^{-1}, Z_k, \hat{X}_k, \text{ and } \hat{\gamma}_s \)

Update \( k \leftarrow k + 1 \)

Update \( (\hat{E}_k^{-1}, Z_k) \leftarrow (\hat{E}_k^{-1}, 0) \) and \( \hat{\gamma}_k \leftarrow \hat{\gamma}_s \)

end

C. \( \| \Sigma - \hat{\Sigma} \|_{H_\infty} \) Optimization

In this section, we extend the results in [13], in the case of undirected graphs, to account for parameterized nodal time-scales in addition to parameterized edge weights in the reduced order system \( \hat{\Sigma} \).

Theorem 2: Consider the network system \( \Sigma \) defined in (1). There exists a reduced network system \( \hat{\Sigma} \) defined as in (2) such that \( \| \Sigma - \hat{\Sigma} \|_{H_\infty} < \gamma \) if and only if there exists \( X \in \mathbb{S}^m_+, \hat{W} \in \mathbb{D}_{++}^{\lvert \mathcal{E} \rvert}, \hat{E} \in \mathbb{D}_{++}^m, \hat{R} \in \mathbb{S}^m_+ \) and a sufficiently small \( \hat{\theta} \in \mathbb{R}^m_+ \) such that the following inequalities are satisfied:

\[
\Phi_2(\hat{X}, \hat{W}, \hat{E}) = \psi(\hat{X}) + \phi_2(\hat{W}, \hat{E}) < 0,
\]

\[
\begin{bmatrix}
\hat{X} \\
\hat{\theta}C^T_e \\
\hat{R}
\end{bmatrix}
\]

\( > 0 \),

\[
\text{tr}(\hat{R}) < \hat{\gamma},
\]

where the linearized term \( \tilde{\phi}_2(k)(\hat{W}) \) at iteration \( k \) is defined similarly to \( \tilde{\phi}_2(k)(\hat{W}) \) in (21). Algorithm 4 returns an \( H_\infty \) suboptimal matrix of nodal time-scales \( \hat{E}_* \). For a given matrix of edge weights \( \hat{W} \) by iteratively solving the following optimization problem (P3), where \( \tilde{\phi}_2(k)(\hat{E}^{-1}(k), Z(k)) \) is defined
IV. ILLUSTRATIVE EXAMPLE

This section considers the example adopted in [12] and [14] to illustrate the effectiveness of the proposed model reduction method and the edge weights and nodal time-scales tuning algorithms. The consensus network system $\Sigma$ considered is characterized by a graph $\mathcal{G}(\mathcal{V},\mathcal{E},E,W)$ that has ten unscaled nodes and fifteen weighted interconnections. Figure 1 shows the underlying graphs $\mathcal{G}$ and $\hat{\mathcal{G}}$ of the original and reduced network systems $\Sigma$ and $\hat{\Sigma}$ respectively. A colour-coded clustering of $\mathcal{G}$ is given by $C_1 = \{1,2,3,4\}, C_2 = \{5,6\}, C_3 = \{7\}, C_4 = \{8\}, \text{ and } C_5 = \{9,10\}$. Nodes 6 and 7 receive external inputs and are highlighted with an additional circle. Finally, the difference between the state variables of nodes 6 and 10 is measured. The incidence matrix $D$ of $\Sigma$ is given by

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$ 

The system matrices and the clustering matrix are given by $W = \text{diag}(5,3,2,1,2,3,5,2,6,7,6,7,1,1,1)$. Similarly to $\hat{\phi}_n^{(k)}(\hat{E}^{-1},Z)$ in [22]:

$$\min_{\hat{X},\hat{E}^{-1},Z,R} \text{tr}(\hat{R})$$
$$\text{s.t. } \psi_2(\hat{X}) + \psi_2^{(k)}(\hat{E}^{-1},Z) \leq 0, Z - f^{(k)}(\hat{E}^{-1}) \leq 0, \begin{bmatrix} \hat{X} \\ \hat{E}^{-1} \\ \hat{I}_r \end{bmatrix} \succeq 0, \begin{bmatrix} \hat{X} \\ \hat{I}_r \end{bmatrix} \geq 0.$$ 

(P4)

Algorithm 3: $H_2$ Iterative Edge Weight Tuning

Input : $E, \hat{E}, W_0, L, F, H, \Pi, \theta, r, \varepsilon$

Output: $W_t$

Initialize $W_0 = W_0 \text{ and } \text{tr}(\hat{R}(0)) = r$

Set iteration step $k \leftarrow 0$

while $k \leq \varepsilon$ and $\text{tr}(\hat{R}(k)) - \text{tr}(\hat{R}(k+1)) \geq \varepsilon$ do

Solve (P3) to obtain $\hat{X}_e, \hat{X}_w, \text{ and } \hat{R}_e$

Update $k \leftarrow k + 1$

Update $\hat{R}(k) \leftarrow \hat{R}_e$

end

Algorithm 4: $H_2$ Iterative Nodal Time-scale Tuning

Input : $E, \hat{E}, W, L, F, H, \Pi, \theta, r, \varepsilon$

Output: $E_t$

Initialize $(\hat{E}_0, Z_0) = (\hat{E}_0, \hat{E}_0) \text{ and } \text{tr}(\hat{R}(0)) = r$

Set iteration step $k \leftarrow 0$

while $k \leq \varepsilon$ and $\text{tr}(\hat{R}(k)) - \text{tr}(\hat{R}(k+1)) \geq \varepsilon$ do

Solve (P4) to obtain $\hat{E}_e, \hat{Z}, \hat{X}_w, \text{ and } \hat{R}_e$

Update $k \leftarrow k + 1$

Update $(\hat{E}_e, Z_k) \leftarrow (\hat{E}_e, \hat{E}_w) \text{ and } \hat{R}_e \leftarrow \hat{R}_e$

end

Using the characteristic clustering matrix $\Pi$, the reduced incidence matrix $\hat{D}$ of $\hat{\Sigma}$ is obtained by removing the duplicate and zero columns from $\Pi^T D$:

$$\hat{D} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$ 

The remaining matrices of the parameterized reduced network system $\hat{\Sigma}$ are given by

$$\hat{L} = \begin{bmatrix} \hat{w}_1 & -\hat{w}_1 & 0 & 0 \\ -\hat{w}_2 & \hat{w}_1 + \hat{w}_3 & -\hat{w}_3 & -\hat{w}_1 \\ 0 & -\hat{w}_3 & \hat{w}_2 + \hat{w}_4 + \hat{w}_3 - \hat{w}_4 - \hat{w}_5 \\ 0 & 0 & -\hat{w}_2 & \hat{w}_4 + \hat{w}_5 \end{bmatrix}, \hat{F} = \begin{bmatrix} \hat{0} & \hat{0} \\ \hat{0} & \hat{0} \\ \hat{0} & \hat{0} \\ \hat{0} & \hat{0} \end{bmatrix},$$ 

$$\hat{E} = \text{diag}(\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4, \hat{e}_5), \hat{H} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$ 

$$\hat{W} = \text{diag}(\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \hat{w}_5),$$

where $\hat{E}$ and $\hat{W}$ are the parameter matrices to be determined by the algorithms proposed in the paper.

Algorithms [1] to [4] are summarized as follows:

- Algorithm [1] outputs an $H_\infty$-suboptimal matrix of weights $W$ for a given matrix of time-scales $E$.
- Algorithm [2] outputs an $H_\infty$-suboptimal matrix of time-scales $E$ for a given matrix of weights $W$.
- Algorithm [3] outputs an $H_2$-suboptimal matrix of weights $W$ for a given matrix of time-scales $E$.
- Algorithm [4] outputs an $H_2$-suboptimal matrix of time-scales $E$ for a given matrix of weights $W$.

The PGP paradigm [14] is used to determine the fixed matrix in the algorithms as well as the initial guess for the tuning. For example, $\hat{E} = \text{diag}(4,2,1,1,2)$ and $\hat{W}_0 = \text{diag}(20,12,14,1,2)$. This is done to illustrate the advantage of the proposed optimal parameter selection.
algorithms in guaranteeing smaller reduction errors than the classical paradigm. Note that in this case, Algorithm 2 solves the problem considered in [12] following the approach in [13]. The expression $AX \rightarrow AY$ indicates that we first perform a tuning of the edge weights using Algorithm X, and then we fix the resulting suboptimal edge weight matrix when tuning the nodal time-scales via Algorithm Y, or vice versa. At each iteration of the algorithms, the SDPs are parsed using YALMIP [19] and solved using MOSEK [20]. We set $\theta = 0.005$ for Algorithms 1 and 2, $\theta = 0.001$ for Algorithms 3 and 4, and $\lambda = 1.0001$ for Algorithms 2 and 4. While the algorithms are guaranteed to converge to a locally optimal point, in our implementation, the algorithms are stopped after a specified $i_{\text{tot}}$ number of iterations, and we report the achieved tolerance values $\varepsilon_\infty = \gamma(i_{\text{tot}} - 1) - \gamma(i_{\text{tot}})$ and $\varepsilon_2 = \text{tr}(\hat{R}(i_{\text{tot}} - 1)) - \text{tr}(\hat{R}(i_{\text{tot}}))$ at exit.

The results are reported in Tables I and II. Table I and Table II show the normalized $H_\infty$ and $H_2$-based reduction errors, which are computed using the built-in $\text{norm}()$ and $\text{norm}()$ functions in MATLAB, respectively. The built-in $\text{minreal}()$ function is used on the original system $\Sigma$ before computing $\|\Sigma\|_{\infty}$ and $\|\Sigma\|_2$. The suboptimal edge weights and nodal time-scales corresponding to the final output returned by each tuning algorithm are given in Table III where $\hat{W}_s = \text{diag}(\hat{w}_s)$ and $\hat{E}_s = \text{diag}(\hat{e}_s)$.

It can be seen from the results that a significant reduction in the normalized reduction errors was achieved when the nodal time-scales are parameterized and tuned in addition to the edge weights. Compared with the PGP paradigm, Algorithms $A^2 \rightarrow A^1$ and $A^1 \rightarrow A^2$ reduce the normalized $H_\infty$ reduction error by 73.3% and 72.6%, respectively. Similarly, Algorithms $A^3 \rightarrow A^4$ and $A^4 \rightarrow A^3$ achieve counterpart reductions in the normalized $H_2$ reduction error of 87.5% and 80.6%, respectively. Compared to Algorithm 3, Algorithm $A^3 \rightarrow A^4$ reduces the normalized $H_2$ reduction error by 75.1%. In the case of $H_\infty$-based model reduction, the tuning of the edge weights only, i.e., Algorithm II remains a novel contribution of this work. Finally, it seems that, for this example, tuning the edge weights after tuning the nodal time-scales shows a small improvement in the normalized reduction errors compared to tuning the nodal time-scales only. For example, when applied after Algorithm 2, Algorithm II reduces the normalized $H_\infty$ reduction error by 2.5% only (from 0.040 to 0.039); and when applied after Algorithm 4, Algorithm II reduces the normalized $H_2$ reduction error by 2.6% only (from 0.078 to 0.076). It is of interest to conduct extensive simulations to validate this observation in other examples.

V. CONCLUSION

In this paper, $H_\infty$- and $H_2$-based model reduction methods for consensus network systems are proposed, whereby a reduced network system with parameterized edge weights and nodal time-scales is tuned via iterative optimization algorithms. Compared to the PGP paradigm in [14] and the $H_2$-based methods for optimal edge weights selection in [12], [13], our $H_2$-based methods allow for additionally optimizing the nodal time-scales of the reduced order system, thereby further reducing the reduction error. Our $H_\infty$-based approaches are novel for optimizing the edge weights, nodal time-scales, or both sequentially. Further investigation into simultaneous edge weight and nodal time-scale optimization is of interest for future work.

REFERENCES

[1] Y. Yazıcıoğlu and A. Speranzon, “High dimensional robust consensus over networks with limited capacity,” IEEE Control Systems Letters, vol. 5, no. 2, pp. 2020-2025, 2020.
[2] F. Rossi, S. Bandyopadhyay, M. Wolf, and M. Pawone, “Review of multi-agent algorithms for collective behavior: A structural taxonomy,” IFAC-PapersOnLine, vol. 51, no. 12, pp. 112–117, 2018.
[3] Z. Li, Z. Duan, G. Chen, and L. Huang, “Consensus of multi-agent systems and synchronization of complex networks: A unified viewpoint,” IEEE Transactions on Circuits and Systems, vol. 57, no. 1, pp. 213–224, 2009.
[4] O. Farhat, D. Abou Jaude, and M. Hudoba de Badyn, “$H_\infty$ network optimization for edge consensus,” European Journal of Control, vol. 62, pp. 2–13, 2021, 2021 European Control Conference Special Issue.
[5] D. R. Foiht, M. Hudoba de Badyn, and M. Mesbahi, “Performance and design of consensus on matrix-weighted and time-scaled graphs,” IEEE Transactions on Control of Network Systems, vol. 7, no. 4, pp. 1812–1822, 2020.
[6] X. Cheng and J. Scherpen, “Balanced truncation approach to linear network system model order reduction,” IFAC-PapersOnLine, vol. 50, no. 1, pp. 2451–2456, 2017.
[7] D. Abou Jaude and M. Fahrood, “Balanced truncation model reduction of nonstationary systems interconnected over arbitrary graphs,” Automatica, vol. 85, pp. 405–411, 2017.
[8] ——, “Model reduction of distributed nonstationary LPV systems,” European Journal of Control, vol. 40, pp. 27–39, 2018.
[9] ——, “Coprime factors model reduction of spatially distributed LTV systems over arbitrary graphs,” IEEE Transactions on Automatic Control, vol. 62, no. 10, pp. 5254–5261, 2017.
[10] ——, “Coprime factors reduction of distributed nonstationary LPV systems,” International Journal of Control, vol. 92, no. 11, pp. 2571–2583, 2019.
[11] X. Cheng and J. Scherpen, “Model reduction methods for complex network systems,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 425–453, 2021.

[12] X. Cheng, L. Yu, and J. Scherpen, “Reduced order modeling of linear consensus networks using weight assignments,” in *18th European Control Conference (ECC)*, 2019, pp. 2005–2010.

[13] X. Cheng, L. Yu, D. Ren, and J. Scherpen, “Reduced order modeling of diffusively coupled network systems: An optimal edge weighting approach,” *arXiv preprint arXiv:2003.03559*, 2020.

[14] N. Monshizadeh, H. Trentelman, and M. K. Camlibel, “Projection-based model reduction of multi-agent systems using graph partitions,” *IEEE Transactions on Control of Network Systems*, vol. 1, no. 2, pp. 145–154, 2014.

[15] Q. T. Dinh, S. Gumsuy, W. Michiels, and M. Diehl, “Combining convex-concave decompositions and linearization approaches for solving BMIs, with application to static output feedback,” *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1377–1390, 2011.

[16] S. Boyd, L. Ghaoul, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, 1994.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] G. Pipeleers, B. Demeulenaere, J. Swevers, and L. Vandenberghe, “Extended LMI characterizations for stability and performance of linear systems,” *Systems & Control Letters*, vol. 58, no. 7, pp. 510–518, 2009.

[19] J. Lofberg, “YALMIP: A toolbox for modeling and optimization in MATLAB,” *IEEE International Conference on Robotics and Automation*, pp. 284–289, 2004.

[20] M. ApS, *The MOSEK optimization toolbox for MATLAB manual*. Version 9.3.6, 2021. [Online]. Available: [https://docs.mosek.com/9.3/toolbox/index.html](https://docs.mosek.com/9.3/toolbox/index.html)