Scalable Identification of Stable Positive Systems

Jack Umenberger and Ian R. Manchester

***Abstract*— Positive systems frequently appear in applications, and enjoy substantially simplified analysis and control design compared to the general LTI case. In this paper we construct a polytopic parameterization of all stable positive systems, and a convex upper bound for simulation error (a.k.a. output error) for which the resulting optimization is a linear program. Previous work on analogous methods for both the positive and general LTI case result in semidefinite programs. We exploit the decomposability of the constraints in these linear programs to develop distributed solutions applicable to identification of large-scale networked systems.**

I. INTRODUCTION

Traffic flow through urban centers, antiretroviral treatment of infectious disease and the smart electricity grid are but a few examples of the diverse array of large scale systems for which modeling and control is becoming increasingly important. In many of these applications, physical constraints imply that the quantities of interest - e.g., number of cars passing through a tunnel, concentrations of pathogens, or power through a transmission line - are nonnegative. In such cases, it is appropriate to model the situation as a so-called *positive system*, in which the set of nonnegative internal states remains invariant under the dynamics.

Over the past decade, positive systems have received increased attention from the control community, largely due to the fact that many performance and stability results in linear system theory are simplified when the dynamics are positive. For example, static state and output feedback controllers were designed using *linear programming* in [1] and [2], respectively. Stability and dissipativity theory for positive systems based on *linear* storage functions and supply rates was developed in [3], and employed for robust stability analysis in [4]. Similarly, the work of [5] provided a bounded real lemma for positive systems based on a *diagonal* quadratic storage function, which enabled the design of *structured* H_∞ controllers. More recently, [6] has presented novel versions of many of the above results, with an emphasis on scalable controller synthesis and verification.

Model-based design and analysis depends of course on the availability accurate system models. While in some applications these come from first-principles, when physical models are either unknown or too complex, some form of data-driven modeling, i.e. system identification, is appropriate [7].

Two major problems in system identification are model instability and non-convexity of the standard quality-of-fit cost functions [8]. Subspace methods have been very successful in identifying linear state-space models from input-output data [9], [10], and various strategies have been developed to guarantee stability of the identified model; c.f. [11], [12], [13], [14]. In a series of recent papers a new approach was developed for obtaining convex parameterizations of stable models and convex upper-bounds for simulation error [15], [16], [17]; these methods were extended to problems with noisy data in [18], [19], [20]. In the context of positive systems, the work of [21] formulated the subspace identification problem, subject to model stability constraints, as a semidefinite program (SDP) with a *diagonal* quadratic Lyapunov function, rather than a dense quadratic Lyapunov function as in the general LTI case [13]; see also [22, §18].

Despite this progress, none of the above methods are concerned with identification of large scale systems. Distributed optimization techniques, such as dual decomposition [23], the alternating direction method of multipliers [24] and game theoretic approaches [25], [26], can be difficult to apply to identification methods based on SDP, as the decomposition of the associated barrier functions can be problematic. In this paper we exploit the simplified control-theoretic results available for positive systems to develop convex optimization problems for system identification that admit scalable, distributed solutions.

II. PRELIMINARIES

A. Notation

For real matrices and vectors $A, B \in \mathbb{R}^{m\times n}$, $A < (\leq) B$ denotes element-wise inequality, whereas $A \prec (\preceq) B$ means $B - A$ is positive definite (semidefinite). The transpose of A is denoted A'. For $A \in \mathbb{R}^{m\times n}$, $A^{(i,j)}$ denotes the scalar entry in the i^{th} row and j^{th} column. We define the sets $\mathbb{R}^n_{++} \coloneqq \{a \in \mathbb{R}^n : a > 0\}$ and $\mathbb{R}^n_{+} \coloneqq \{a \in \mathbb{R}^n : a \geq 0\}$. For $a \in \mathbb{R}^n$, $|a| \in \mathbb{R}^n_+$ denotes element-wise absolute value, and $|a|_\sigma \coloneqq \left(\sum_{i=1}^n |a^{(i)}|^\sigma\right)^{1/\sigma}$ denotes the σ-norm. We define $\mathbf{1} \in \mathbb{R}^n$ as the vector with all elements equal to 1; the dimension n can be inferred from the context. We use $A^{(\omega_i,:)}$ to denote the nonzero elements of the i^{th} row of A, and $A^{(:,\omega_i)}$ for the nonzero elements of the i^{th} column.

B. Positive state space models

This paper concerns the identification discrete time *positive* LTI models of the form

$$x_{t+1} = Ax_t + Bu_t \tag{1a}$$

$$y_t = Cx_t + Du_t \tag{1b}$$

where $u \in \mathbb{R}^{n_u}$, $y \in \mathbb{R}^{n_y}$ and $x \in \mathbb{R}^{n_x}$ denote the input, output and state, respectively. For convenience, we group all model parameters into a single variable, $\Sigma = \{A, B, C, D\}$.

Definition 1. *A system of the form (1) is said to be* positive *if A, B, C and D are element-wise nonnegative.*

The notation $\Sigma \geq 0$ is used to denote positive systems of the form (1), and should be interpreted as shorthand for $A \geq 0, B \geq 0, C \geq 0$ and $D \geq 0$. Nonnegativity of the state

This work was supported in part by the Australian Research Council.

J. Umenberger and I.R. Manchester are with the Australian Centre for Field Robotics (ACFR), Department of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, NSW 2006, Australia ({j.umenberger, i.manchester}@acfr.usyd.edu.au).

variable x_t greatly simplifies Lyapunov stability analysis of positive systems:

Lemma 1 ([27, Lemma 6.2.1])**.** *For $A \geq 0$ the following statements are equivalent:*

1. *The matrix A is Schur stable.*
2. *There exists $p \in \mathbb{R}^n_{++}$ such that $p'A < p'$.*

The dynamical systems interpretation of this result is that $V(x) = p'x$ serves as a linear Lyapunov function for the system, Σ. We denote the set of all stable positive models, and corresponding Lyapunov functions p, by

$$\Theta \triangleq \{p, \Sigma : \ \Sigma \geq 0, \ p \in \mathbb{R}^{n_x}_{++}, \ p'A < p'\}. \tag{2}$$

Observe that this set is not jointly convex in A and p, due to the bilinear stability condition, $p'A < p'$. For convenience, we will often group a model Σ and vector p as $\theta = \{p, \Sigma\}$.

C. Problem data

We assume data of the form $\mathcal{Z}^T_{\mathrm{DT}} = \{\tilde{u}_t, \tilde{y}_t, \tilde{x}_t\}^T_{t=1}$ where $\tilde{u}$, $\tilde{y}$ and $\tilde{x}$ denote (possibly noisy) measurements of u, y and x, respectively. Notice that we require measurements of the state x. In this paper, we have in mind the identification of networked systems such that $x = \left[x^1, \ldots, x^n\right]$, where x^i denotes the measurable state at node i, e.g. transport networks where x^i denotes traffic density [28]. This is a rather restrictive assumption, necessitated by the fact that popular state estimation techniques, e.g. subspace methods, return state estimates subject to an arbitrary coordinate transformation [29, §2.2], which may not be consistent with a positive realization of the dynamics. Subspace methods for positive systems are an important subject for future research.

D. Simulation error

In system identification, a common measure of model quality, given a dataset $\mathcal{Z}^T_{\mathrm{DT}}$, is the *simulation error*, defined as $\bar{\mathcal{E}}_\sigma \triangleq \sum_{t=1}^T |\tilde{y}_t - y_t|^\sigma_\sigma$, where y_t represents the *simulated output*: $y_t = CA^{t-1}\tilde{x}_1 + \Sigma_{\tau=1}^{t-1} CA^{t-1-\tau}B\tilde{u}_\tau + D\tilde{u}_t$. This dependence on the simulated output renders $\bar{\mathcal{E}}_\sigma$ a highly nonlinear function of the model parameters.

E. Problem statement

In this paper, our goal is to minimize (globally) the simulation error via a distributed search over all stable positive systems; i.e., we seek a distributed solution to $\min_{\theta \in \Theta} \bar{\mathcal{E}}_\sigma$, for $\sigma = 1, 2$. Nonconvexity of the feasible set Θ, and nonlinearity of the objective $\bar{\mathcal{E}}_\sigma$, make this is a challenging optimization problem that we do not attempt to solve directly. Rather, we proceed by constructing a polytopic parametrization of all stable positive systems, and develop convex approximations to the simulation error.

III. MINIMIZATION OF EQUATION ERROR

Global minimization of the simulation error is a formidable task, due to the nonlinear dependence of $\bar{\mathcal{E}}_\sigma$ on the model parameters, Σ. A common approach for circumventing this difficulty is to minimize the *equation error*

$$\mathcal{E}_2 = \sum_{t=1}^T |\tilde{y}_t - C\tilde{x}_t - D\tilde{u}_t|_2^2 + \sum_{t=1}^T |\tilde{x}_{t+1} - A\tilde{x}_t - B\tilde{u}_t|_2^2 \tag{3}$$

in place of the simulation error; this is the approach adopted by, e.g., subspace methods. In this section, we show that minimization of a related cost function, the *weighted equation error*, subject to the model stability constraint (2) can be formulated as a convex optimization problem.

A. Constrained minimization for LTI positive systems

From Lemma 1, stability of a positive linear system is equivalent to the existence of $p \in \mathbb{R}^{n_x}_{++}$ such that $A'p - p < 0$. By introducing the change of variables $\mathcal{A} = PA$, where $P = \mathrm{diag}(p^{(1)}, \ldots, p^{(n_x)})$ is a diagonal matrix such that $\mathcal{A}^{(i,j)} = p^{(i)}A^{(i,j)}$, the stability condition can be written as $\mathcal{A}'\mathbf{1} - p < 0$, which is linear in p and $\mathcal{A}$.

This motivates the introduction of a new cost function, the *weighted equation error* [21], [13] defined

$$\mathcal{E}_2^p \triangleq \sum_{t=1}^T |\eta_t|_2^2 + \sum_{t=1}^T |P\epsilon_t|_2^2. \tag{4}$$

where $\epsilon_t = \tilde{x}_{t+1} - A\tilde{x}_t - B\tilde{u}_t$ and $\eta_t = \tilde{y}_t - C\tilde{x}_t - D\tilde{u}_t$. By defining a second change of variables, $\mathcal{B} = PB$, we have $P\epsilon_t = P\tilde{x}_{t+1} - \mathcal{A}\tilde{x}_t - \mathcal{B}\tilde{u}_t$, and thus the cost function $\mathcal{E}_2^p$ is a convex quadratic function of $P, \mathcal{A}, \mathcal{B}, C$ and D.

Remark 2. *Notice that $A'p - p < 0$ holds for any positive scalar multiple of p, and so $A'p - p + \delta\mathbf{1} \leq 0$, for some $\delta > 0$, defines an identical condition, with no conservatism.*

As in Section II-B, we group the model parameters under this change of variables into a single vector, $\bar{\Sigma} = \{\mathcal{A}, \mathcal{B}, C, D\}$. In light of Remark 2,

$$\bar{\Theta}_s \triangleq \{p, \bar{\Sigma} : \ \bar{\Sigma} \geq 0, \ p \geq \delta\mathbf{1}, \ \mathcal{A}'\mathbf{1} - p + \delta\mathbf{1} \leq 0\} \tag{5}$$

defines a convex parametrization of all stable positive models: $\bar{\theta} = \{p, \bar{\Sigma}\} \in \bar{\Theta}_s$ implies $\theta = \{p, \Sigma\} \in \Theta$ where $\Sigma = \mathcal{M}(\bar{\theta}) \triangleq \{P^{-1}\mathcal{A}, P^{-1}\mathcal{B}, C, D\}$.

Therefore, the optimization problem:

$$\min_{\bar{\theta} \in \bar{\Theta}_s} \ \mathcal{E}_2^p \tag{6}$$

represents minimization of a convex quadratic function subject to linear inequality constraints; i.e., a QP.

B. Stable minimization of ℓ_1-norm of equation error

Further reductions in computational complexity can be achieved by minimizing the ℓ_1-norm of the weighted equation error, defined

$$\mathcal{E}_1^p \triangleq \sum_{t=1}^T |\eta_t|_1 + \sum_{t=1}^T |P\epsilon_t|_1. \tag{7}$$

By employing the same change of variables introduced in §III-A, the optimization problem $\min_{\bar{\theta} \in \bar{\Theta}_s} \mathcal{E}_1^p$ represents a linearly constrained least deviations problem, which can be formulated as a LP by introducing slack variables for the absolute values in the cost function; see, e.g., [30, §6.1].

C. Relationship to existing methods

In the context of positive systems, this weighted equation error technique has been employed in the work of [21], in which the following optimization problem was proposed:

$$\min_{\bar{\theta}} \quad \mathcal{E}_2^p \tag{8a}$$

$$\text{s.t.} \quad \mathcal{A} \geq 0, \mathcal{B} \geq 0, C \geq 0, D \geq 0 \tag{8b}$$

$$\begin{bmatrix} P - \delta I_{n_x} & \mathcal{A}' \\ \mathcal{A} & P \end{bmatrix} \succeq 0. \tag{8c}$$

The stability condition (8c) is identical to that which appears in the stable identification of general LTI systems [13], except here use of diagonal P introduces no conservatism, as the model is constrained to be positive. While this does introduce some degree of simplification over the general LTI case (i.e. fewer decision variables), as (8c) represents a LMI, (8) must be solved as a SDP.

In the formulation of §III-A, we invoke the simpler stability condition in Lemma 1, which enables model stability to be enforced with the linear inequality (5), in place of the LMI (8c). The resulting QP (6) has superior scalability to the SDP (8), and, in particular, introduces the possibility for a distributed solution, as shall be shown in §V.

IV. BOUNDING SIMULATION ERROR

Generally speaking, minimization of equation error (3), is not guaranteed to identify models with good long term predictive power, as small equation error does not necessarily imply small simulation error. In this section, we derive a convex condition that ensures the ℓ_1-norm of weighted equation error, $\mathcal{E}_1^p$, upper bounds the simulation error.

A. ℓ_1-gain of incremental error dynamics

The goal of this section is to develop a convex upper bound for the simulation error. To this end, by expressing $\bar{\mathcal{E}}_1$ as

$$\sum_{t=1}^{T} |\eta_t + C\tilde{x}_t - Cx_t|_1 \leq \sum_{t=1}^{T} |\eta_t|_1 + \sum_{t=1}^{T} |C\tilde{x}_t - Cx_t|_1$$

where x_t denotes the *simulated state sequence* given by $x_{t+1} = Ax_t + B\tilde{u}_t$, it is evident that it is sufficient to upper bound $\sum_t |C\tilde{x}_t - Cx_t|_1$. To study this quantity, we rewrite the equation error, ϵ_t, as $\tilde{x}_{t+1} = A\tilde{x}_t + B\tilde{u}_t + \epsilon_t$ and remark that the simulated state sequence, x_t, and the estimated state sequence $\tilde{x}_t$, are in fact solutions to the same dynamical system. Specifically, by considering the system

$$x_{t+1} = Ax_t + B\tilde{u}_t + v_t \tag{9}$$

with input v_t and initial condition $x_1 = \tilde{x}_1$, we observe:

i. for $v_t = 0$ the solution to (9) is x_t.
ii. for $v_t = \epsilon_t$, the solution to (9) is $\tilde{x}_t$.

Defining $\Delta_t := \tilde{x}_t - x_t$, it is clear that Δ_t satisfies the *incremental error dynamics* given by the system

$$\Delta_{t+1} = A\Delta_t + \epsilon_t \tag{10a}$$

$$z = C\Delta_t. \tag{10b}$$

This analysis, much in the spirit of [15], has uncovered a relationship between the equation error ϵ_t, and the quantity of interest $C(\tilde{x}_t - x_t)$, which can be quantified as the ℓ_1-gain of the incremental error system in (10). Furthermore, observe that when the original system (1) is positive, so too are the incremental error dynamics in (10). The following result characterizes the ℓ_1-gain of positive systems:

Lemma 3 ([4, Lemma 1])**.** *Let (1) denote a positive system. The following statements are equivalent:*

1. *The matrix A is Schur and the ℓ_1-gain of $u \mapsto y$ is less than γ.*
2. *There exists $p \in \mathbb{R}^{n_x}_{++}$ such that*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}' \begin{bmatrix} p \\ \mathbf{1} \end{bmatrix} < \begin{bmatrix} p \\ \gamma\mathbf{1} \end{bmatrix}. \tag{11}$$

Lemma 3 can be used to quantify the contribution of each 'channel' of the equation error ϵ to the simulation error:

Lemma 4. *Given a stable, positive system of the form (1), the ℓ_1-gain from the j^{th} input channel $\epsilon^{(j)}$ to the output z of the incremental error system (10), is given by $p^{(j)}$, where*

$$p = \arg\min_{p \in \mathbb{R}^{n_x}_{++}} \sum_i p^{(i)} \quad \text{s.t.} \quad C'\mathbf{1} + (A' - I)p < 0. \tag{12}$$

Lemma 4 implies that for minimization of $\mathcal{E}_1^p$ subject to the constraint $C'\mathbf{1} + A'p - p < 0$, the vector p serves as both a stability certificate (linear Lyapunov function) and a meaningful weighting of equation error. Specifically, p penalizes most heavily the 'channels' of ϵ_t that contribute most significantly to the simulation error, as p encodes the ℓ_1-gain of $\epsilon \mapsto z$ in the incremental error system (10).

We define the set of all positive models $\bar{\Sigma}$ and vectors p that satisfy the constraint in (12) by

$$\bar{\Theta}_b \triangleq \{p, \bar{\Sigma} : \ \bar{\Sigma} \geq 0, \ p \geq \delta\mathbf{1}, \ C'\mathbf{1} + \mathcal{A}'\mathbf{1} - p + \delta\mathbf{1} \leq 0\} \tag{13}$$

and remark that this set in convex in $\bar{\Sigma}$ and p.

Remark 5. *The strict inequality in (12) has been replaced by a non-strict inequality in (13) to ensure that the constraints lead to a well posed optimization problem. The extent of the conservatism introduced by this approximation is characterized by δ, and can be made arbitrarily small.*

B. Upper bound on simulation error

We are now in a position to present the main theoretical contribution of this paper:

Theorem 6. *Given a dataset $\mathcal{Z}_{\text{DT}}^T$, consider the LP:*

$$\bar{\theta}^* = \arg\min_{\bar{\theta} \in \bar{\Theta}_b} \ \mathcal{E}_1^p. \tag{14}$$

where $\mathcal{E}_1^p$ is given in (7). The following statements hold:

1. *$\mathcal{E}_1^p \geq \bar{\mathcal{E}}_1$ for all $\bar{\theta} \in \bar{\Theta}_b$.*
2. *$\bar{\Theta}_b$, defined in (13), is a parametrization of all stable positive systems.*
3. *If $\mathcal{Z}_{\text{DT}}^T$ represents noiseless data signals generated by the true system, $\{\tilde{p}, \tilde{\Sigma}\} \in \bar{\Theta}_b$, then $\mathcal{E}_1^p = \bar{\mathcal{E}}_1 = 0$, i.e., the upper bound is tight.*

V. ALGORITHMS FOR DISTRIBUTED IDENTIFICATION

In §III and §IV, we derived convex approximations to the problem of minimizing simulation error over all stable positive systems. In this section we present distributed solutions to the convex optimization problems (6) and (14).

A. Distributed problem formulation

Consider a multiagent system consisting of a set of agents denoted $N = \{1, \dots, n_x + n_y\}$. Each agent $i \in N$, for $i \leq n_x$, is associated with an element of the state vector x, i.e., agent i is associated with $x^{(i)}$. Agent i is responsible for estimating the model parameters that directly affect the dynamics of the i^{th} state, i.e. $A^{(i,:)}$ and $B^{(i,:)}$, and is assumed to measure $\tilde{x}^{(i)}$ directly. Similarly, each agent $i \in N$, for $i > n_x$, is associated with an element of the output y. Specifically, agent $i > n_x$ measures $\tilde{y}^{(i-n_x)}$, and maintains the parameters that effect the $i-n_x^{\text{th}}$ output, i.e., $C^{(i-n_x,:)}$ and $D^{(i-n_x,:)}$. Working with the change of variables introduced in §III-A, we group these parameters into the vector $v_i \in \mathbb{R}^{d_i}$, defined for $i \in N$ by

$$v_i = \begin{bmatrix} \mathcal{A}^{(\omega_i,:)'} & \mathcal{B}^{(\omega_i,:)'} & p^{(i)} \end{bmatrix}', \ i \leq n_x \tag{15a}$$

$$v_i = \begin{bmatrix} C^{(\omega_{i-n_x},:)'} & D^{(\omega_{i-n_x},:)'} \end{bmatrix}', \ i > n_x \tag{15b}$$

where $\mathcal{A}^{(\omega_i,:)}$ denotes the nonzero elements of $\mathcal{A}^{(i,:)}$.

The cost function, weighted equation error, may then be decomposed into the sum of *local* cost functions, i.e. $\mathcal{E}_\sigma^p = \sum_{i \in N} \phi_i^\sigma(v_i)$, where

$$\phi_i^\sigma(v_i) = \begin{cases} \sum_{t=1}^{T-1} |p^{(i)} \epsilon_t^{(i)}|_\sigma^\sigma, & i \leq n_x \\ \sum_{t=1}^{T} |\eta_t^{(i-n_x)}|_\sigma^\sigma, & i > n_x, \end{cases} \tag{16}$$

with $\sigma = 1, 2$ for problems (14) and (6), respectively. Despite separability of $\mathcal{E}_\sigma^p$, communication between agents is necessitated by the constraints $\bar{\theta} \in \bar{\Theta}_s$, c.f. (5), and $\bar{\theta} \in \bar{\Theta}_b$, c.f. (13), which couple parameters maintained different agents. We define the *neighbors* of agent i as the set of agents with which i exchanges model parameters.

B. Alternating direction method of multipliers

In this section, we present a decentralized solution to problem (6), i.e. $\min_{\bar{\theta} \in \bar{\Theta}_s} \mathcal{E}_2^p$, based on the alternating direction method of multipliers (ADMM). We represent the quadratic cost function[1] as $\sum_t |P\epsilon_t|_2^2 = v'Qv = \sum_i v_i' Q_i v_i$ where $v = \begin{bmatrix} v_1' & \dots v_{n_x}' \end{bmatrix}'$ and $Q = \text{blkdiag}(Q_1, \dots, Q_{n_x})$. For convenience, we define $N^x = \{1, \dots, n_x\}$.

To obtain a formulation of ADMM amenable to a decentralized implementation, we introduce a *copy* of the decision variable v, which we denote s. Let $\mathcal{A}_s$, $\mathcal{B}_s$ and p_s denote copies of $\mathcal{A}$, $\mathcal{B}$ and p respectively, such that $s = \begin{bmatrix} s_1' \dots s_{n_x}' \end{bmatrix}'$ where $s_i = \begin{bmatrix} \mathcal{A}_s^{(\omega_i,:)'} & \mathcal{B}_s^{(\omega_i,:)'} & p_s^{(i)} \end{bmatrix}'$. In this way, $\{s_i\}_{i \in N^x}$ represents a partition of the model parameters w.r.t. the *rows* of $\mathcal{A}$, $\mathcal{B}$ and p. Let us introduce a second partition of s, $\{s_{ci}\}_{i \in N^x}$, w.r.t. the *columns* of $\mathcal{A}$, $\mathcal{B}$ and p', i.e. $s_{ci} = \begin{bmatrix} \mathcal{A}_s^{(:,\omega_i)'} & \mathcal{B}_s^{(:,\omega_i)'} & p^{(i)} \end{bmatrix}'$. For ease of exposition, we define the *row* and *column* neighbors of agent i by the sets $N_i^R = \{j : i \cup \mathcal{A}^{(j,i)} \neq 0\}$ and $N_i^C = \{j : i \cup \mathcal{A}^{(i,j)} \neq 0\}$.

We may now express the original problem (6) as

$$\min_{s,v} \ v'Qv + \sum_{i=1}^{n_x} g_i(s_{ci}), \quad \text{s.t.} \quad v - s = 0$$

where $\sum_{i \in N^x} g_i(s_{ci})$ encodes the nonnegativity constraints $\mathcal{A} \geq 0$, $\mathcal{B} \geq 0$ and stability condition $\mathbf{1}'\mathcal{A} - p' \leq -\delta \mathbf{1}'$, i.e.

$$g_i(s_{ci}) = \mathcal{I}\left(\mathcal{A}_s^{(:,\omega_i)}\right) + \mathcal{I}\left(\mathcal{B}_s^{(:,\omega_i)}\right) + \mathcal{I}\left(\mathbf{1}'\mathcal{A}_s^{(:,\omega_i)} - p^{(i)} + \delta\right).$$

Here $\mathcal{I}(x)$ denotes the element-wise indicator function, which is zero for all $x \geq 0$, and infinite elsewhere. The Lagrangian for the scaled formulation of ADMM is

$$L_\rho(v, s, \mu) = \sum_{i \in N^x} v_i' Q_i v_i + \sum_{i \in N^x} g_i(s_{ci}) + \frac{\rho}{2} \|v - s + \mu\|^2,$$

where ρ denotes the *penalty parameter*, and μ denotes the scaled dual variable (i.e. Lagrange multiplier). The k^{th} iteration of the ADMM involves the following computations:

$$v_i^{k+1} = \arg\min_{v_i} \ v_i' Q_i v_i + \frac{\rho}{2} \|v_i - s_i^k + \mu_i^k\|^2 \tag{17a}$$

$$= (2Q_i + \rho I)^{-1}(\rho(s_i^k - \mu_i^k)) \tag{17b}$$

$$s_{ci}^{k+1} = \arg\min_{s_{ci}} \ g_i(s_{ci}) + \frac{\rho}{2} \|v_{ci}^{k+1} - s_{ci} + \mu_{ci}^k\|^2 \tag{17c}$$

$$\mu_i^{k+1} = \mu_i^k + v_i^{k+1} - s_i^{k+1}. \tag{17d}$$

C. Gradient play on state-based potential games

The formulation of the ADMM in §V-B gave no consideration to robustness against imperfect inter-agent communication. In this section, we address this by adopting a recently developed approach to distributed optimization, in which the solution is obtained from the Nash equilibrium of a state-based potential game [25], [26]. Such an approach is known to be robust to delays in communication and heterogeneous clock rates [25]. We can represent $\min_{\bar{\theta} \in \bar{\Theta}_b} \mathcal{E}_1^p$ in a form compatible with these methods as follows:

$$\min_{v_i \in \mathcal{V}_i, i \in N} \ \sum_{i \in N} \phi_i^1(v_i), \quad \text{s.t.} \ \sum_{i \in N} M_i^l v_i + \delta \leq 0, \ l \in N^x, \tag{18}$$

where v_i is defined in (15), $\mathcal{V}_i = \mathbb{R}_+^{d_i}$ for $i \in N$ enforces nonnegativity of the model parameters, and $\sum_{i=1}^{n_x} M_i^k v_i + \delta \leq 0$ encodes the dissipation inequality from (13), i.e. $\sum_{i=1}^{n_x} \mathcal{A}^{(i,l)} - p^{(l)} + C^{(i,l)} \leq -\delta$, for $l \in N^x$.

To develop a decentralized solution to (18), we employ the method of [31] and introduce auxiliary variables $e_i = \{e_i^l\}_{l \in N^x}$ for $i \in N$, where e_i^l denotes agent i's estimate of the l^{th} constraint, i.e. $e_i^l \sim \sum_{j=1}^{n_x} M_j^l v_j + \delta$. We can now define a state-based potential game equivalent to (18) by introducing a state $\xi_i = (v_i, e_i)$ for each agent $i \in N$, as well as a state *action* $\hat{\xi}_i = (\hat{v}_i, \hat{e}_i)$. Here, $\hat{e}_i = \{\hat{e}_i^l\}_{l \in N^x}$ with $\hat{e}_i^l = \{\hat{e}_{i \to j}^l\}_{j \in N_i}$. The term $\hat{e}_{i \to j}^l$ denotes the change in the estimation of the l^{th} constraint that agent i communicates to agent $j \in N_i$. Agent states then evolve according to the dynamics $(\tilde{v}, \tilde{e}) = f(\xi, \hat{\xi})$ given explicitly by

$$\tilde{v}_i = v_i + \hat{v}_i \tag{19a}$$

$$\tilde{e}_i^l = e_i^l + M_i^l \hat{v}_i + \sum_{j \in N_i} \hat{e}_{j \to i}^l - \sum_{j \in N_i} \hat{e}_{i \to j}^l. \tag{19b}$$

The introduction of estimation variables e_i decouples the constraints in (18). Each agent may then be assigned its own

[1] Minimization of $\sum_t |\eta_t|_2^2$ can be trivially parallelized, and carried out independently to minimization of $\sum_t |P\epsilon_t|_2^2$ so we restrict our attention to $\sum_t |P\epsilon_t|_2^2$.

cost function, which depends only on the states (and actions) of neighboring agents (c.f. [31, §III-B] for details):

$$J_i(\xi,\hat{\xi}) = \phi_i^1(\tilde{v}_i) + \beta \sum_{j\in N_i}\sum_{l\in N^x} \left(\max(0,\tilde{e}_j^l)\right)^2. \qquad (20)$$

Here $\beta > 0$ denotes a trade-off parameter, which balances agent i's local cost function with the penalty on inconsistencies between estimation terms. To solve (18), each agent $i \in N$ minimizes its own individual cost function J_i. Whilst any almost any optimization policy can be employed, we use a gradient play algorithm (c.f., e.g., [26, §4]). At each iteration, agent i's state is updated in accordance with (19), with expressions for $\hat{v}_i$ and $\hat{e}_{i\to j}^l$ given in [31, §4].

Finally, it is clear that the gradient play policy outlined above requires agent i to communicate with neighboring agents $j \in N_i$. However, unlike the ADMM algorithm in §V-B, it is not clear how this neighbor set should be specified. In fact, any specification of the neighbor sets $\{N_i\}_{i\in N}$ that gives rise to an *connected*, *undirected* communication graph is sufficient; c.f. [31, Theorem 1]. The effect of different choices of $\{N_i\}_{i\in N}$ is a subject for future research.

VI. CASE STUDIES

A. Scalability of identification under stability constraint

In this section we illustrate the effect that replacing the LMI stability condition in (8c), with the simple linear inequality in (5), has on the scalability of minimization of $\mathcal{E}_2^p$ subject to model stability and positivity constraints. Specifically, we compare the QP (6) of §III-A to the SDP (8) of §III-C, proposed by [21]. The comparison comprised 10 experimental trials. In each trail, we randomly generated a stable positive system, in which A was banded with 4 diagonals (on each side of the main diagonal). For such a system, the number of model parameters increases linearly with n_x. The system was simulated for $T = 1\times 10^4$ time steps,[2] and $\tilde{Z}_{\mathrm{DT}}^T$ was obtained by adding Gaussian disturbances to the simulated $\{x_t\}_{t=1}^T$ and $\{y_t\}_{t=1}^T$. Both the QP and SDP were formulated with Yalmip [32], and solved by interior point methods (IPM) using Mosek 7.0.0.119. Average computation times for increasing model order n_x are given in Table I.

TABLE I: Computation time (in seconds, to 3 s.f.) for varying model order n_x, with $n_u = 5$ and $T = 1\times 10^4$, averaged over 10 trials. Computations were carried out on a desktop machine (Intel i7, 3.40GHz 8GB RAM).

n_x	250	500	1000	1500	2000	4000	8000
SDP	10.9	73.6	540	1670	-	-	-
QP	0.0642	0.128	0.238	0.388	3.66	-	-
ADMM	11.0	23.6	43.7	67.2	90.3	178	363

From Table I it is clear that, as expected, QP scales better than SDP, with regards to both computation time and memory requirements. Computation times increase (approx.) linearly and exponentially with increasing n_x, for the QP and SDP, respectively. Furthermore, the QP can handle $n_x < 2500$ before memory is exhausted, whereas the SDP is limited to $n_x < 1500$. Table I also demonstrates that ADMM is capable of identifying higher order systems, given the same available memory, compared to IPM.

[2] As $\sum_{t=1}^T |P\epsilon_t|_2^2$ can be expressed as $v'Qv$, c.f. §V-B, the size of the optimization problem, for each method, is independent of T.

B. Identification of structured systems

In this section, we illustrate the advantages of the bound on simulation error developed in §IV-B over the basic stability condition of §III, in the context of identification of structured systems. Incorporating known structural information is recognized as a central problem in the industrial application of system identification algorithms [8, Section 6]; e.g., it may be known that two components are identical, and connected in series. To enforce model stability under structural constraints of the form $A^{(i,j)} = A^{(k,l)}$, we require $\mathcal{A}^{(i,j)}/p^{(i)} = \mathcal{A}^{(k,l)}/p^{(k)}$, which is not jointly convex in $\mathcal{A}$ and p.

To circumvent this nonconvexity, one may consider a two-step approach, solving first (21a) and then (21b):

$$\bar{\theta} = \{p_d, \bar{\Sigma}\} \triangleq \arg\min_{\bar{\theta}\in\bar{\Theta}_b} \ \mathcal{E}_1^p \qquad (21a)$$

$$\bar{\theta}_d \triangleq \arg\min_{\bar{\theta}\in\bar{\Theta}_b} \ \mathcal{E}_1^p \quad \text{s.t. } p = p_d,\ A^{(i,j)} = A^{(k,l)} \qquad (21b)$$

for all i, j, k, l for which we wish to enforce such constraints, based on *a priori* structural knowledge. The idea is that (21a) furnishes us with an approximation of p, which is then fixed as $p = p_d$ in (21b), such that the structural constraints $p_f^{(i)}A^{(i,j)} = p_f^{(k)}A^{(k,l)}$ are convex in A.

To demonstrate the advantages of enforcing $\bar{\theta} \in \bar{\Theta}_b$ (c.f. §IV-B) over $\bar{\theta} \in \bar{\Theta}_s$ (c.f. §III) we compare the approach in (21) to the following similar two-step procedure:

$$\bar{\theta} = \{p_s, \bar{\Sigma}\} \triangleq \arg\min_{\bar{\theta}\in\bar{\Theta}_s} \ \mathcal{E}_1^p \qquad (22a)$$

$$\bar{\theta}_s \triangleq \arg\min_{\bar{\theta}\in\bar{\Theta}_s} \ \mathcal{E}_1^p \quad \text{s.t. } p = p_s,\ A^{(i,j)} = A^{(k,l)}. \qquad (22b)$$

Notice that in (21) we enforce the dissipation inequality $p'A - p' + 1'C \le -\delta 1'$, whereas in (22) we simply enforce the stability condition $p'A - p' < 0$. The difference between these two conditions is illustrated by identifying a model Σ with structure of the form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}, C = \begin{bmatrix} C_1 & 0 \end{bmatrix}$$

where $A_{11} = A_{22}$ and $B_1 = B_2$. The interpretation is that Σ represents two identical subsystems, that are coupled by A_{12} and A_{21}, however only the first subsystem contributes directly to the measured plant output.

We generate a dataset from a model $\tilde{\Sigma}$ with parameters:

$$A_{11} = [0.2, 0.7; 0.5, 0.4],\ A_{22} = [0.5, 0.2; 0.4, 0.7],$$
$$A_{12} = [0, 0.001; 0, 0.02],\ A_{21} = [0.01, 0.05; 0.03, 0.01],$$
$$B_1 = [0.1; 0],\ B_2 = [0.2, 0],\ C_1 = [1, 0].$$

To form $\mathcal{Z}_{\mathrm{DT}}^T$, $\tilde{y}$ and $\tilde{x}$ were taken to be the true simulated quantities, corrupted by additive Gaussian noise. Notice that $\tilde{\Sigma}$ is not in the model class, as there is some mismatch between the two subsystems that are assumed to be identical.

The normalized error of each identified model, defined as $\bar{\mathcal{E}}_1/\sum_t |\tilde{y}_t|$, for 1000 experimental trials is plotted in Fig. 1, from which it is evident that (21) outperforms (22). Greater insight can be gained by studying Fig. 2, which depicts the equation error of Σ_d (from (21)) and Σ_s (from (22)), for a

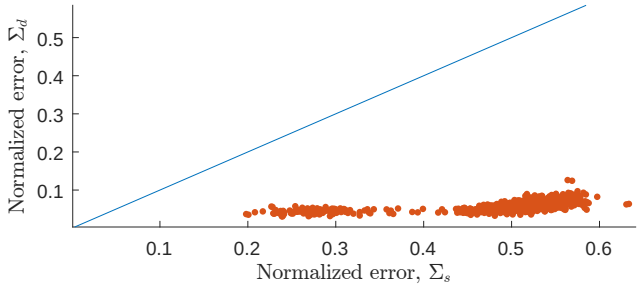

Fig. 1: Normalized simulation error for models Σ_d, identified by (21), and Σ_s, identified by (22). 1000 experimental trials were conducted.

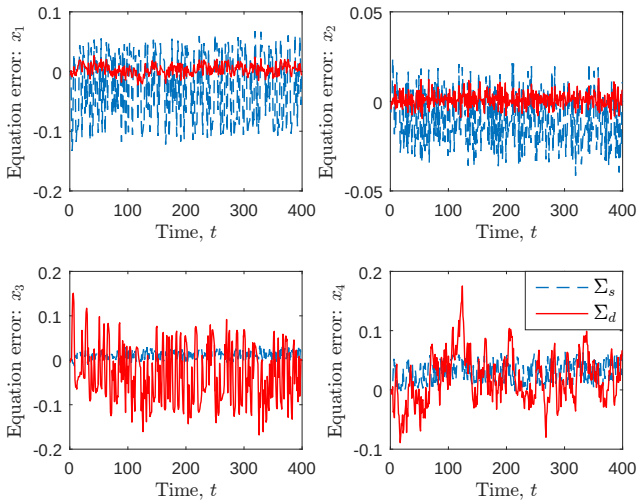

Fig. 2: Equation error for Σ_d from (21), and Σ_s from (22). The approach of (21) detects that equation error in x_1 and x_2 contributes most significantly to simulation error, and returns a model that achieves good fit in these states, at the expense of poor fit in x_3 and x_4, due to structural constraints.

typical experimental trial. The ℓ_1 gain from ϵ to $\bar{\mathcal{E}}_1$ for the true model $\tilde{\Sigma}$ is given by $[4.66, 5.44, 0.09, 0.11]$, for each state, respectively; i.e. equation error in x_1 and x_2 contribute most significantly to simulation error. In this trial, $p_d = [4.11, 4.51, 0.25, 0.30]$ and $p_s = [1.03, 1.34, 1.11, 1.09] \times 10^{-4}$. As p functions as a weight on equation error, (21) will prioritize minimization of equation error in x_1, x_2, at the expense of poor fit in x_3, x_4; c.f. Fig. 2. This accounts for the superior performance of Σ_d over Σ_s, in Fig. 1.

References

[1] P. De Leenheer and D. Aeyels, "Stabilization of positive linear systems," *Systems & Control Letters*, vol. 44, no. 4, pp. 259–271, 2001.

[2] M. A. Rami, "Solvability of static output-feedback stabilization for LTI positive systems," *Systems & control letters*, vol. 60, no. 9, pp. 704–708, 2011.

[3] W. M. Haddad and V. Chellaboina, "Stability and dissipativity theory for nonnegative dynamical systems: a unified analysis framework for biological and physiological systems," *Nonlinear Analysis: Real World Applications*, vol. 6, no. 1, pp. 35–65, 2005.

[4] C. Briat, "Robust stability and stabilization of uncertain linear positive systems via integral linear constraints: L_1- and L_∞-gains characterization," *International Journal of Robust and Nonlinear Control*, 2013.

[5] T. Tanaka and C. Langbort, "The bounded real lemma for internally positive systems and h-infinity structured static state feedback," *IEEE transactions on automatic control*, vol. 56, no. 9, pp. 2218–2223, 2011.

[6] A. Rantzer, "Scalable control of positive systems," *European Journal of Control*, vol. 24, pp. 72–80, 2015.

[7] L. Ljung, *System Identification: Theory for the User (2nd Edition)*, 2nd ed. Prentice Hall, Jan. 1999.

[8] ——, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, Apr. 2010.

[9] W. E. Larimore, "System identification, reduced-order filtering and modeling via canonical variate analysis," in *American Control Conference, 1983*. IEEE, 1983, pp. 445–451.

[10] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[11] J. M. Maciejowski, "Guaranteed stability with subspace methods," *Systems & Control Letters*, vol. 26, no. 2, pp. 153–156, 1995.

[12] T. Van Gestel, J. A. Suykens, P. Van Dooren, and B. De Moor, "Identification of stable models in subspace identification by using regularization," *Automatic Control, IEEE Transactions on*, vol. 46, no. 9, pp. 1416–1420, 2001.

[13] S. L. Lacy and D. S. Bernstein, "Subspace identification with guaranteed stability using constrained optimization," *Automatic Control, IEEE Transactions on*, vol. 48, no. 7, pp. 1259–1263, 2003.

[14] D. N. Miller and R. A. De Callafon, "Subspace identification with eigenvalue constraints," *Automatica*, vol. 49, no. 8, pp. 2468–2473, 2013.

[15] A. Megretski, "Convex optimization in robust identification of nonlinear feedback," *2008 47th IEEE Conference on Decision and Control*, pp. 1370–1374, 2008.

[16] M. M. Tobenkin, I. R. Manchester, J. Wang, A. Megretski, and R. Tedrake, "Convex optimization in identification of stable non-linear state space models," *49th IEEE Conference on Decision and Control (CDC)*, no. 4, pp. 7232–7237, Dec. 2010.

[17] B. N. B. Bond, Z. Mahmood, Y. Li, R. Sredojevic, A. Megretski, V. Stojanovi, Y. Avniel, and L. Daniel, "Compact modeling of nonlinear analog circuits using system identification via semidefinite programming and incremental stability certification," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 29, no. 8, pp. 1149–1162, 2010.

[18] I. Manchester, M. M. Tobenkin, and A. Megretski, "Stable nonlinear system identification: Convexity, model class, and consistency," in *Proceedings of the 16th IFAC Symposium on System Identification (SYSID)*, Brussels, Belgium, 2012.

[19] M. M. Tobenkin, I. R. Manchester, and A. Megretski, "Stable nonlinear identification from noisy repeated experiments via convex optimization," in *American Control Conference (ACC), 2013*. IEEE, 2013, pp. 3936–3941.

[20] J. Umenberger, J. Wagberg, I. Manchester, and T. Schön, "On identification via EM with latent disturbances and Lagrangian relaxation," in *Proceedings of the 17th IFAC Symposium on System Identification (SYSID)*, Beijing, China, 2015.

[21] Q. Hui and W. M. Haddad, "Subspace identification of stable nonnegative and compartmental dynamical systems via constrained optimization," in *American Control Conference, 2006*. IEEE, 2006, pp. 6–pp.

[22] W. M. Haddad, V. Chellaboina, and Q. Hui, *Nonnegative and compartmental dynamical systems*. Princeton University Press, 2010.

[23] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations research*, vol. 8, no. 1, pp. 101–111, 1960.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[25] J. R. Marden, "State based potential games," *Automatica*, vol. 48, no. 12, pp. 3075–3088, 2012.

[26] N. Li and J. R. Marden, "Designing games for distributed optimization," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 2, pp. 230–242, 2013.

[27] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994.

[28] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[29] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: TheoryImplementationApplications*. Springer Science & Business Media, 2012.

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

[31] N. Li and J. R. Marden, "Decoupling coupled constraints through utility design," *Automatic Control, IEEE Transactions on*, vol. 59, no. 8, pp. 2289–2294, 2014.

[32] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip