

Article

Optimal H_2 Moment Matching-Based Model Reduction for Linear Systems Through (Non)convex Optimization

Ion Necoara^{1,2,†} and Tudor-Corneliu Ionescu^{1,2,*,†} 

¹ Department of Automatic Control and Systems Engineering, Politehnica University of Bucharest, Splaiul Independentei 313, 060042 Bucharest, Romania; ion.necoara@upb.ro

² Gheorghe Mihoc—Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, 050711 Bucharest, Romania

* Correspondence: tudor.ionescu@upb.ro

† These authors contributed equally to this work.

Abstract: In this paper, we compute a (local) optimal reduced order model that matches a prescribed set of moments of a stable linear time-invariant system of high dimension. We fix the interpolation points and parametrize the models achieving moment-matching in a set of free parameters. Based on the parametrization and using the H_2 -norm of the approximation error as the objective function, we derive a nonconvex optimization problem, i.e., we search for the optimal free parameters to determine the model yielding the minimal H_2 norm of the approximation error. Furthermore, we provide the necessary first-order optimality conditions in terms of the controllability and the observability Gramians of a minimal realization of the error system. We then propose two gradient-type algorithms to compute the (local) optimal models, with mathematical guarantees on the convergence. We also derive convex semidefinite programming relaxations for the nonconvex Problem, under the assumption that the error system admits block-diagonal Gramians, and derive sufficient conditions to guarantee the block diagonalization. The solutions resulting at each step of the proposed algorithms guarantee the achievement of the imposed moment matching conditions. The second gradient-based algorithm exhibits the additional property that, when stopped, yields a stable approximation with a reduced H_2 error norm. We illustrate the theory on a CD-player and on a discretized heat equation.

Keywords: model order reduction; moment matching; optimal H_2 -norm; (non)convex optimization; gradient method

MSC: 93C05



Citation: Necoara, I.; Ionescu, T.-C. Optimal H_2 Moment Matching-Based Model Reduction for Linear Systems Through (Non)convex Optimization. *Mathematics* **2022**, *10*, 1765. <https://doi.org/10.3390/math10101765>

Academic Editors: Jan Awrejcewicz, Hamed Farokhi, Roman Starosta, José M. Vega, Hari Mohan Srivastava and Ying-Cheng Lai

Received: 21 April 2022

Accepted: 19 May 2022

Published: 22 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To reduce the complexity and the high dimension of LTI systems yielded by, e.g., partial-differential equations or networks of interconnected systems, for analysis and control, model reduction is called for, see e.g., [1]. The approximation error must be small and the most important physical and/or structural properties, such as the stability, must be preserved.

State-of-the-art: Model reduction of linear systems consists of two main categories of methods: SVD-based and moment matching-based. Balanced truncation (BT) is a well-known SVD-based method, first introduced in [2], see also [3]. The method is based on bringing the given stable system into the so-called balanced realization, where the states are ordered according to the Hankel singular values (HSVs)—positive similarity invariants measuring the degree of controllability and observability of the states. For model reduction, the balanced realization is truncated, neglecting the dynamics corresponding to the smaller HSVs. The resulting model is stable and balanced and yields an aprioric upper bound of the H_∞ norm of the approximation error. However, for large-scale, dense systems, BT may become unscalable even if parallelized, due to the requirement to compute the balancing

coordinate transformation, see, e.g., [4,5] and the references therein. Moreover, balanced truncation also exhibits an upper bound on the H_2 norm of the approximation error, usually larger than the bound on the H_∞ norm see, e.g., ([1], Section 7.2.2.) Moment matching techniques represent an alternative tool for model order reduction, see, e.g., [1,6,7], where a low degree rational transfer function is constructed to match the coefficients of the series expansion of the original transfer function at various interpolation points in the complex plane, called moments. The moments are yielded by the Krylov projection matrix [6,8,9], computed as the (unique) solution of a Sylvester equation [10,11] constructed from a state-space realization of the system in the time domain. Then, families of parametrized reduced order models that match say that ν moments are computed [10,12].

To find more accurate models in the family, two-sided projections have been employed, see, e.g., [1,6,13], yielding a ν order model matching 2ν moments at two sets of ν interpolation points. Two-sided moment matching has also been achieved in [14] as well as in [15], reducing the H_2 approximation error. Furthermore, in [16,17], using two-sided projection-based interpolatory methods, the model minimizing the H_2 -norm of the approximation error has been computed. Based on [18], in [17], an Iterative Rational Krylov Algorithm (IRKA) for H_2 model reduction has been proposed. The resulting model interpolates the given transfer function and its derivatives at the mirror images of the poles of the approximant and is numerically efficient, involving only matrix-vector multiplications (such as the Lanczos iterations [19]). However, as pointed out in [20], IRKA does not guarantee the convergence to a (local) minimum, except for the special case discussed in [21]. Moreover, the resulting model may be unstable even if the original system is asymptotically stable. Note that, in moment matching, the selection of the interpolation points and the preservation of stability are still open issues. For instance, in [22], a two-sided projection method is proposed, where one projection contains SVD features such as stability preservation and an error bound, whereas the other projection contains iterative Krylov features. The resulting model is stable, solves a restricted H_2 -error norm minimization problem and matches some set of moments. In [20,23], using factorization of the error system and applying a greedy algorithm and adaptive procedure to find suitable interpolation points to minimize the H_2 error is given and, additionally, stability preservation is guaranteed. In [4], a stability preserving, projection-based method is developed, based on Lyapunov stability arguments. Results of the optimization-based model reduction that are similar to those in this paper, but for linear network systems, have been recently given in [24], with the preservation of the graph topology (however, the results in this paper have appeared on arxiv before [24]). In [25], the H_2 model reduction problem of discrete systems has been studied in the Stiefel manifold framework, yielding stable models at each step of the algorithm and converging to a (local) optimal solution. In [26], a Riemannian optimal model reduction method for stable systems is proposed, yielding (local) optimal, stable models.

Motivation: Note that the mentioned model reduction methods, focused on the search of interpolation points to preserve stability or achieve minimal approximation error in some norm, usually lead to unsatisfying choices of interpolation points such as, e.g., the mirror images of the *unknown* poles of the approximations. Hence, interpolation at zero, preserving the DC-gain of the given system to match the step response, may be lost causing disruptions to potential control designs. In the Sylvester equation-based moment matching framework, finding the set of parameters yielding the model, in the family of ν order parametrized models that match ν moments at a *fixed* set of interpolation points, that achieves the *minimal* H_2 -norm of the error is an open question. In this paper, we give a solution to this problem. In detail, given an n -th order stable linear system, and a set of *fixed* interpolation points, we find a set of parameters yielding a model that simultaneously minimizes the H_2 -error norm, is stable and matches ν moments of the given system at the fixed interpolation points. Hence, the proposed method is shifting the search of optimal interpolation points to the search of optimal free parameters, leaving the given frequencies unaltered.

Contributions: In this paper, we consider the family of reduced order models that match a prescribed set of ν moments of a given stable linear time-invariant (LTI) system.

Fixing the set of interpolation points, the set of prescribed moments is obtained calculating a single Krylov projection and the resulting models, matching the prescribed moments, are parametrized in a set of ν free parameters. Then, with the H_2 -norm of the approximation error as the objective function, we formulate a nonconvex optimization problem to find the (local) optimal free parameters yielding the model that achieves the minimal H_2 error norm and is stable. Based on a realization of the error system, we provide the necessary first-order Gramian-based optimality conditions. Then, we propose two gradient-type methods, with mathematical guarantees on the convergence to find the (local) optimal free parameters. We also provide convex SDP relaxations of the nonconvex problem through the assumption that the error system admits block-diagonal Gramians and then derive sufficient conditions to guarantee the feasibility of the SDPs. The main contributions are:

- (i) We first formulate the model reduction problem in the family of models matching ν moments, parameterized in a set of ν free parameters. Then, a suitable nonconvex optimization framework is derived, where the objective function is the H_2 approximation error norm, written in terms of the controllability and observability Gramians of a minimal realization of the error system. We also write the necessary first-order Gramian-based optimality conditions, i.e., the KKT system;
- (ii) For the formulated model reduction problem, we propose two numerical optimization algorithms. The first method is using a gradient update for solving the KKT system, leading to a simple iteration involving matrix multiplications. However, with this update, the stability of the approximation is achieved asymptotically. The second method is based on a partial minimization approach. We show that, for the evaluation of the gradient of the objective function, we need to solve two Lyapunov equations yielding the Gramians. We also prove that the gradient of the objective function is Lipschitz continuous. Therefore, a gradient-based algorithm is developed, ensuring the convergence to a local optimal solution due to the smoothness of the objective function. Although the gradient evaluation is expensive, each iteration provides a *stable* reduced model;
- (iii) Finally, we propose a convex SDP relaxation of the original nonconvex optimization problem by assuming that the error system admits a block-diagonal observability Gramian. We also derive sufficient conditions to guarantee block diagonalization.

We now emphasize the main properties of the proposed approach. First, each step of the proposed optimization algorithms yields a model that guarantees the achievement of the prescribed moment matching conditions. Moreover, when the second algorithm, developed for the partial minimization, is stopped, it always yields a stable approximation with a small H_2 error norm. Since the proposed methods do not perform a search over the optimal interpolation points, but over some free parameters, the resulting H_2 performance may be different from other methods, such as IRKA. Hence, one practical advantage of the proposed approach is that the application oriented properties encompassed in the matched moments can be preserved entirely. Other advantages of our methods include the simplicity of the iterations and the mathematical guarantees on the convergence.

Content: In Section 2, we briefly review the Sylvester equation-based moment matching framework. In Section 3, we formulate an optimal H_2 -norm model reduction problem, recast as an optimization problem with a Gramian-based cost function, and derive the corresponding first-order optimality conditions. In Section 4, we analyze several numerical optimization methods to solve it. In Section 5, we illustrate the theory with a CD-player and a discretized heat equation.

2. Preliminaries

In this section, we briefly review the Sylvester equation-based moment matching model reduction for LTI systems, see [10,12]. We also review the notions of controllability and observability Gramians, and the computation of the associated H_2 -norm.

2.1. Linear Systems

Consider a linear, time-invariant (LTI) dynamical system:

$$\Sigma: \quad \dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad t \geq 0, \quad (1)$$

with the state $x(t) \in \mathbb{R}^n$, the input $u(t) \in \mathbb{R}^m$ and the output $y(t) \in \mathbb{R}^p$. For the sake of readability, the argument t is usually dropped. Throughout the rest of the paper, we assume that the system (1) is stable (spectrum of A , denoted by $\sigma(A)$, satisfies $\sigma(A) \subset \mathbb{C}^-$, where \mathbb{C}^- is the set of complex numbers with negative real parts), and minimal (the pair (A, B) is controllable and the pair (A, C) is observable). The transfer function of (1) is:

$$K(s) = C(sI - A)^{-1}B, \quad K: \mathbb{C} \setminus \sigma(A) \rightarrow \mathbb{C}^{p \times m}. \quad (2)$$

Since the system is assumed to be minimal, $\sigma(A)$ is the set of poles of K .

2.2. Sylvester Equation-Based Moment Matching

Assume that (1) is a minimal realization of the transfer function $K(s)$. The moments of (2) are defined as follows:

Definition 1 ([1,10]). The k -moment of (1) at s_1 , along the direction $\ell \in \mathbb{R}^m$, is

$$\eta_k(s_1) = (-1)^k / k! \left[\frac{d^k K(s)}{ds^k} \ell \right]_{s=s_1} \in \mathbb{C}^p, k \geq 0.$$

Consider (1) and let the matrices $S \in \mathbb{R}^{v \times v}$, $L = [\ell_1 \ell_2 \dots \ell_l] \in \mathbb{R}^{m \times v}$, $\ell_i \in \mathbb{R}^m$, be such that the pair (S, L) is observable. Let $\Pi \in \mathbb{R}^{n \times v}$ be the solution of Sylvester equation:

$$A\Pi + BL = \Pi S. \quad (3)$$

Since (1) is minimal, assuming that $\sigma(A) \cap \sigma(S) = \emptyset$ and that the pair (S, L) is observable, then Π is the unique solution of the equation (3) and $\text{rank } \Pi = v$, see, e.g., [27].

Proposition 1 ([10,28]). The moments of (1) at the interpolation points $\{s_1, s_2, \dots, s_v\} = \sigma(S)$ are in a one-to-one relation with the elements of the matrix $C\Pi$ (by one-to-one relation between set of moments and elements of a matrix, we mean that moments are uniquely determined by the elements of the matrix.)

The next result gives necessary and sufficient conditions for a system to achieve moment matching:

Proposition 2 ([10,28]). Consider the system: $\hat{\Sigma}: \dot{\xi} = F\xi + Gu, \quad \psi = H\xi$, with $F \in \mathbb{R}^{v \times v}$, $G \in \mathbb{R}^{v \times m}$, $H \in \mathbb{R}^{p \times v}$ and the transfer function: $\hat{K}(s) = H(sI - F)^{-1}G$. Let $S \in \mathbb{R}^{v \times v}$ and $L \in \mathbb{R}^{m \times v}$ be such that the pair (S, L) is observable. Moreover, assume that $\sigma(S) \cap \sigma(A) = \emptyset$ and $\sigma(F) \cap \sigma(S) = \emptyset$. Then, the moments HP of the system $\hat{\Sigma}$ match the moments $C\Pi$ of the original system (1), at $\sigma(S)$, if and only if $HP = C\Pi$, where the invertible matrix $P \in \mathbb{R}^{v \times v}$ is the unique solution of the Sylvester equation $FP + GL = PS$.

We present a family of v order models that match v moments of the system (1). Fixing $S \in \mathbb{R}^{v \times v}$ and $L \in \mathbb{R}^{m \times v}$ such that the pair (S, L) is observable and $\sigma(S) \cap \sigma(A) = \emptyset$, for $P = I_v$, the system:

$$\hat{\Sigma}_G: \quad \dot{\xi} = (S - GL)\xi + Gu, \quad \psi = C\Pi\xi, \quad (4)$$

with the transfer function

$$\hat{K}(s) = C\Pi(sI - S + GL)^{-1}G, \quad (5)$$

describes a family of ν order models that match the moments of (1) at $\sigma(S)$, parameterized in the free parameters $G \in \mathbb{R}^{\nu \times m}$ such that $\sigma(S - GL) \cap \sigma(S) = \emptyset$.

2.3. The Computation of the Moments

In practice, the moments CII are not computed solving the Sylvester Equation (3), but using Krylov projections. In this section, we recall the notion of moments based on the Krylov projection matrix, see, e.g., [6,19,29,30]. For simplicity, given a set of distinct points $\{s_1, s_2, \dots, s_\nu\}$, closed under complex conjugation with $\{s_1, s_2, \dots, s_\nu\} \cap \sigma(A) = \emptyset$ and $\ell_i \in \mathbb{R}^p \setminus \{0\}, i = 1 : \nu$, a Krylov projection matrix is built as $V \in \mathbb{R}^{n \times \nu}$, with [1,12]:

$$V = [(s_1 I - A)^{-1} B \ell_1 \quad \dots \quad (s_\nu I - A)^{-1} B \ell_\nu]. \quad (6)$$

Then, the moments of (1) are given by the elements of the matrix CV , see, e.g., [10,12]. Furthermore, there is a direct relation between CV and CII from Proposition 1, see [12,28]:

Proposition 3 ([28]). *Let Π be the unique solution of (3) and consider the projection matrix V . Then, there exists a square, non-singular, matrix T such that $\Pi = VT$. For $T = I_\nu$, the Krylov projection matrix V is the unique solution of equation (3) for $S = \text{diag}(s_1, s_2, \dots, s_\nu)$ and $L = [\ell_1 \ \ell_2 \ \dots \ \ell_\nu] \in \mathbb{R}^{m \times \nu}$. Moreover, the moments of system (1) at $\sigma(S)$ are given by $\text{CII} = CVT$.*

We make the following working assumption: matrix Π , the unique solution of (3) is formed using the Krylov projection matrix V . Furthermore, the moments of system (1) at $\sigma(S)$ are computed using Proposition 3, i.e., without solving (3).

2.4. H_2 -Norm Based on the Gramians of Linear Systems

We briefly recall the Definition of the H_2 -norm of a stable and minimal LTI system (1) and its computation based on the controllability and the observability Gramians. Consider the system (1) with the transfer function $K \in \mathcal{H}_2$, the Hilbert space of complex functions analytic in the open right-half plane and square integrable. By [31], the H_2 -norm is defined as:

$$\|K\|_{H_2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{Trace}(K^*(j\omega)K(j\omega)) d\omega,$$

where $K^*(j\omega) = \overline{K^T(j\omega)}$ and can be computed as:

$$\|K\|_{H_2}^2 = \text{Trace}(CWC^T) = \text{Trace}(B^T MB), \quad (7)$$

where $W \succ 0$ and $M \succ 0$ are the controllability and the observability Gramians of (1), respectively. The Gramians are the unique solutions of the Lyapunov equations $AW + WA^T + BB^T = 0$ and $A^T M + MA + C^T C = 0$, respectively, see [1].

3. H_2 Model Reduction by Moment Matching and Optimization

In this section, based on the parametrization (4) and using the H_2 -norm of the approximation error as the objective function, we formulate a nonconvex optimization problem to calculate the approximation that achieves moment matching and minimizes the H_2 -norm. We derive the optimality (KKT) conditions in terms of the controllability and the observability Gramians of the error system and then propose two gradient-based methods, with mathematical guarantees on the convergence to a (local) optimum. Let us assume that the observable pair (S, L) is fixed a priori, such that $\sigma(S) \cap \sigma(A) = \emptyset$. We seek the reduced order model $\hat{\Sigma}_G$ in the family (4) parameterized in G , that yields the minimal H_2 -norm of the approximation error. To this end, we formulate the optimal moment matching-based H_2 -norm model reduction problem as:

Problem 1. Fix the matrices $S \in \mathbb{R}^{\nu \times \nu}$ and $L \in \mathbb{R}^{m \times \nu}$ such that the pair (S, L) is observable and $\sigma(S) \cap \sigma(A) = \emptyset$. Given the n -th order system (1), with the transfer function (2) and the family of ν order models $\hat{\Sigma}_G$ as in (4), with the transfer function \hat{K} as in (5), that match the ν fixed moments of (1) at $\sigma(S)$, find the free parameters G such that the following conditions are satisfied:

- (i) the H_2 norm of the error system $\|K - \hat{K}\|_2$ is minimal;
- (ii) the model $\hat{\Sigma}_G$ is stable, i.e., $\sigma(S - GL) \subset \mathbb{C}^-$;
- (iii) $\sigma(S) \cap \sigma(S - GL) = \emptyset$.

Since the Problem 1 is *nonconvex*, it is usually difficult to find the global optimum, but an optimization algorithm for such a problem may yield stationary points. However, the proposed algorithms are descent methods usually leading to a (local) minimum, depending on the initialization. Note that in most applications, a reduced order model is necessary to capture desired input-output behaviors/dynamics. In particular, the system responses to a set of prescribed harmonic signals (e.g., step, low frequency signals, etc.) are imposed through the selection of the interpolation point matrix S , see [10]. Note that Problem 1 is relevant from a systemic point of view, since by fixing S and L the desired input-output behaviors can be captured. Hence, from the family of reduced order models that achieve moment matching at $\sigma(S)$, we choose the desired inputs whose output behaviors are to be preserved through model reduction, see also [10,12,32] for similar settings. Problem 1 can be recast in terms of the computation of the \mathcal{H}_2 norm of the Gramians of the realization of the error system $\mathcal{K}_e = K - \hat{K}$, with \hat{K} from (5), parameterized in G . Let $(\mathcal{A}_e, \mathcal{B}_e, \mathcal{C}_e)$ be a state-space realization of the error transfer function $\mathcal{K}_e(s) = \mathcal{C}_e(sI - \mathcal{A}_e)^{-1}\mathcal{B}_e$, where

$$\mathcal{A}_e = \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}, \mathcal{B}_e = \begin{bmatrix} B \\ G \end{bmatrix}, \mathcal{C}_e = C \begin{bmatrix} I & -\Pi \end{bmatrix}. \quad (8)$$

Let \mathcal{W} and \mathcal{M} be the controllability and the observability Gramians of (8), solutions of the Lyapunov equations:

$$\mathcal{A}_e \mathcal{W} + \mathcal{W} \mathcal{A}_e^T + \mathcal{B}_e \mathcal{B}_e^T = 0, \quad (9a)$$

$$\mathcal{A}_e^T \mathcal{M} + \mathcal{M} \mathcal{A}_e + \mathcal{C}_e^T \mathcal{C}_e = 0. \quad (9b)$$

respectively. Let us recall a standard result for Lyapunov equations:

Lemma 1. Let \mathcal{A}_e be given stable matrix, i.e., $\sigma(\mathcal{A}_e) \subset \mathbb{C}^-$. Then, there exist the unique solutions \mathcal{W} and \mathcal{M} , positive semidefinite matrices of (9a) and (9b). Furthermore, if the error system is minimal (i.e., $(\mathcal{A}_e, \mathcal{B}_e)$ is controllable and $(\mathcal{A}_e, \mathcal{C}_e)$ is observable), then \mathcal{W} and \mathcal{M} are positive definite.

We partition \mathcal{W} and \mathcal{M} following the block structure of \mathcal{A}_e :

$$\mathcal{W} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12}^T & W_{22} \end{bmatrix}, \mathcal{M} = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix}. \quad (10)$$

Hence, Equation (9a) can be explicitly rewritten as:

$$AW_{11} + W_{11}A^T + BB^T = 0, \quad (11a)$$

$$AW_{12} + W_{12}(S - GL)^T + BG^T = 0, \quad (11b)$$

$$(S - GL)W_{22} + W_{22}(S - GL)^T + GG^T = 0, \quad (11c)$$

and Equation (9b) explicitly rewritten as:

$$A^T M_{11} + M_{11} A + C^T C = 0, \quad (12a)$$

$$A^T M_{12} + M_{12}(S - GL) - C^T C \Pi = 0, \quad (12b)$$

$$(S - GL)^T M_{22} + M_{22}(S - GL) + \Pi^T C^T C \Pi = 0, \quad (12c)$$

respectively.

3.1. Optimization Formulation of the proposed H_2 model reduction problem

Consider Problem 1 with the pair (S, L) fixed a priori. Note that we can find CII based on Proposition 3, i.e., $\Pi = VT$, where V as in (6) and T some fixed non-singular matrix. Moreover, if we choose S unstable, that is $\sigma(S) \subseteq \mathbb{C}^+$, then any (local) optimal solution of the Problem 1 automatically satisfies $\sigma(S) \cap \sigma(S - GL) = \emptyset$. Then, from (7), it follows that Problem 1 can be written equivalently as an optimization problem:

$$\begin{aligned} \min_{G \text{ s.t. } \sigma(S - GL) \subset \mathbb{C}^-} \|\mathcal{K}_e\|_2^2 &= \min_{(G, \mathcal{W}) \text{ s.t. } \sigma(S - GL) \subset \mathbb{C}^-, (9a)} \text{Trace} \left(C \begin{bmatrix} I & -\Pi \end{bmatrix} \begin{bmatrix} W_{11} & W_{12} \\ W_{12}^T & W_{22} \end{bmatrix} \begin{bmatrix} I \\ -\Pi^T \end{bmatrix} C^T \right) \\ &= \min_{(G, \mathcal{M}) \text{ s.t. } \sigma(S - GL) \subset \mathbb{C}^-, (9b)} \text{Trace} \left(\begin{bmatrix} B^T & G^T \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \begin{bmatrix} B \\ G \end{bmatrix} \right). \end{aligned}$$

In the sequel, we consider only the formulation in terms of the observability Gramian \mathcal{M} :

$$\begin{aligned} \min_{(G, \mathcal{M})} \text{Trace}(\mathcal{B}_e^T \mathcal{M} \mathcal{B}_e) \\ \text{s.t. : } \sigma(S - GL) \subset \mathbb{C}^- \quad \text{and} \quad \mathcal{A}_e^T \mathcal{M} + \mathcal{M} \mathcal{A}_e + \mathcal{C}_e^T \mathcal{C}_e = 0. \end{aligned} \quad (13)$$

We clearly observe that the reduced order model is parametrized in the matrix G . Let:

$$\mathcal{A}(G) = \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}, \quad \mathcal{B}(G) = \begin{bmatrix} B \\ G \end{bmatrix} \begin{bmatrix} B \\ G \end{bmatrix}^T, \quad \mathcal{C} = \mathcal{C}_e^T \mathcal{C}_e = \begin{bmatrix} I \\ -T^T V^T \end{bmatrix} C^T C \begin{bmatrix} I \\ -T^T V^T \end{bmatrix}^T.$$

In the next sections, we present several (equivalent) reformulations of the nonconvex problem (13), accompanied by their first-order optimality conditions.

3.2. KKT Approach

We determine the corresponding KKT system for the optimization problem (13). We first define the open set $\mathcal{D}_{(SL)} = \{G : \sigma(S - GL) \subset \mathbb{C}^-\}$. The Lagrangian function associated to problem (13) is given by:

$$\Gamma(\mathcal{W}, \mathcal{M}, G) = \text{Trace}(\mathcal{M} \mathcal{B}(G)) + \text{Trace}(\mathcal{W}(\mathcal{A}^T(G) \mathcal{M} + \mathcal{M} \mathcal{A}(G) + \mathcal{C})),$$

where the multiplier \mathcal{W} is associated to the equality constraint in (13). Then, we write (13) into the max-min form:

$$\max_{\mathcal{W}} \min_{\mathcal{M}, G \in \mathcal{D}_{(SL)}} \Gamma(\mathcal{W}, \mathcal{M}, G). \quad (14)$$

From standard optimization arguments, we know that any solution $(\mathcal{W}, \mathcal{M}, G)$ of problem (14) satisfies the KKT system:

$$\nabla \Gamma(\mathcal{W}, \mathcal{M}, G) = 0 \quad \Longleftrightarrow \quad \begin{cases} \nabla_{\mathcal{W}} \Gamma(\mathcal{W}, \mathcal{M}, G) = 0 \\ \nabla_{(\mathcal{M}, G)} \Gamma(\mathcal{W}, \mathcal{M}, G) = 0. \end{cases}$$

Using the block structure (10) for \mathcal{M} and \mathcal{W} , the next Theorem explicitly derives the corresponding KKT system:

Theorem 1. The KKT system of the optimization problem (13) is given by the following explicit expressions:

$$\nabla\Gamma(\mathcal{W}, \mathcal{M}, G) = 0 \iff \begin{cases} \mathcal{A}^T(G)\mathcal{M} + \mathcal{M}\mathcal{A}(G) + \mathcal{C} = 0 \\ \mathcal{A}(G)\mathcal{W} + \mathcal{W}\mathcal{A}^T(G) + \mathcal{B}(G) = 0 \\ M_{12}^T B + M_{22}G - M_{12}^T W_{12}L^T - M_{22}W_{22}L^T = 0. \end{cases} \quad (15)$$

Proof. Note that the KKT system has the explicit form:

$$\nabla\Gamma(\mathcal{W}, \mathcal{M}, G) = \begin{bmatrix} \nabla_{\mathcal{W}}\Gamma(\mathcal{W}, \mathcal{M}, G) \\ \nabla_{\mathcal{M}}\Gamma(\mathcal{W}, \mathcal{M}, G) \\ \nabla_G\Gamma(\mathcal{W}, \mathcal{M}, G) \end{bmatrix} = \begin{bmatrix} \mathcal{A}^T(G)\mathcal{M} + \mathcal{M}\mathcal{A}(G) + \mathcal{C} \\ \mathcal{A}(G)\mathcal{W} + \mathcal{W}\mathcal{A}^T(G) + \mathcal{B}(G) \\ \nabla_G\Gamma(\mathcal{W}, \mathcal{M}, G) \end{bmatrix}.$$

It remains to explicitly compute $\nabla_G\Gamma(\mathcal{W}, \mathcal{M}, G)$. However, to write the gradient expression, we introduce a gradient of Γ w.r.t. G using the trace as $\Gamma'(\mathcal{W}, \mathcal{M}, G) dG = \text{Trace}(\nabla_G^T\Gamma(\mathcal{W}, \mathcal{M}, G) dG)$, with $dG \in \mathbb{R}^{v \times m}$. Then:

$$\begin{aligned} \text{Trace}(\nabla_G^T\Gamma(\mathcal{W}, \mathcal{M}, G) dG) &= \text{Trace}(\mathcal{M}\mathcal{B}'(G)) + \text{Trace}(\mathcal{W}((\mathcal{A}'(G))^T\mathcal{M} + \mathcal{M}\mathcal{A}'(G))) \\ &= \text{Trace}\left(\mathcal{M} \begin{bmatrix} 0 & B dG^T \\ dGB^T & dGG^T + G dG^T \end{bmatrix} + \mathcal{W} \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix} \mathcal{W}\right) \\ &= 2\text{Trace}(B^T M_{12} dG + G^T M_{22} dG - L W_{12}^T M_{12} dG - L W_{22} M_{22} dG), \end{aligned}$$

with \mathcal{W} and \mathcal{M} as in (10). Then:

$$\nabla_G\Gamma(\mathcal{W}, \mathcal{M}, G) = 2(M_{12}^T B + M_{22}G - M_{12}^T W_{12}L^T - M_{22}W_{22}L^T).$$

Hence, we get the closed form for the KKT system (15). \square

Theorem 1 also yields the necessary optimality condition for the optimization problem (13) associated to the Problem 1:

Lemma 2. If \mathcal{M} and $G \in \mathcal{D}_{(SL)}$ solves the optimization problem (13) associated to the Problem 1, then there exists \mathcal{W} such that the triplet $(\mathcal{W}, \mathcal{M}, G)$ solves the KKT system (15).

3.3. Partial Minimization Approach

Consider $\mathcal{A} = \mathcal{A}(G)$. Then, the following partial minimization holds for (13):

$$(13) = \min_{G: \sigma(S-GL) \subset \mathbb{C}^-} \left(\min_{\mathcal{M}: \mathcal{A}_e^T \mathcal{M} + \mathcal{M}\mathcal{A}_e + \mathcal{C}_e^T \mathcal{C}_e = 0} \text{Trace}(\mathcal{B}_e^T \mathcal{M} \mathcal{B}_e) \right).$$

However, if $S - GL$ and A are stable, by Lemma 1, there exists $\mathcal{M} = \mathcal{M}(G) \succeq 0$, the unique solution of the Lyapunov equation $\mathcal{A}_e^T \mathcal{M} + \mathcal{M}\mathcal{A}_e + \mathcal{C}_e^T \mathcal{C}_e = 0$. Hence, for any stabilizing G , the partial minimization in \mathcal{M} leads to an optimal value:

$$f(G) = \min_{\mathcal{M}: \mathcal{A}_e^T \mathcal{M} + \mathcal{M}\mathcal{A}_e + \mathcal{C}_e^T \mathcal{C}_e = 0} \text{Trace}(\mathcal{B}_e^T \mathcal{M} \mathcal{B}_e) = \text{Trace}\left(\begin{bmatrix} B \\ G \end{bmatrix}^T \mathcal{M}(G) \begin{bmatrix} B \\ G \end{bmatrix}\right),$$

where the matrix $\mathcal{M}(G)$ is the unique, positive definite solution of the Lyapunov equation:

$$\begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix} + \begin{bmatrix} C^T C & -C^T C_V \\ -C_V^T C & C_V^T C_V \end{bmatrix} = 0, \quad (16)$$

with $C_V = C\Pi = CVT$. Explicitly, in terms of G , we have:

$$\min_G f(G) \left(= \text{Trace} \left(\begin{bmatrix} B \\ G \end{bmatrix}^T \mathcal{M}(G) \begin{bmatrix} B \\ G \end{bmatrix} \right) \right) \quad (17)$$

$$\text{s.t. : } \sigma(S - GL) \subset \mathbb{C}^- \quad \text{and} \quad (16),$$

where $\mathcal{M}(G)$ is the solution of (16). We now compute a closed form expression for the gradient of the objective function of (17):

$$f(G) = \text{Trace} \left(\begin{bmatrix} B \\ G \end{bmatrix}^T \mathcal{M}(G) \begin{bmatrix} B \\ G \end{bmatrix} \right) = \text{Trace} \left(\mathcal{M}(G) \begin{bmatrix} B \\ G \end{bmatrix} \begin{bmatrix} B \\ G \end{bmatrix}^T \right) = \text{Trace}(\mathcal{M}(G)\mathcal{B}(G)).$$

Theorem 2. The objective function f in (17) is differentiable on the set of stable matrices $\mathcal{D}_{(SL)}$ and the gradient of f at $G \in \mathcal{D}_{(SL)}$ is given explicitly by:

$$\nabla f(G) = 2 \left[-M_{12}^T(G)W_{12}(G)L^T - M_{22}(G)W_{22}(G)L^T + M_{12}^T(G)B + M_{22}(G)G \right], \quad (18)$$

where $\mathcal{M}(G)$ solves (16) and $\mathcal{W}(G)$ solves the equation:

$$\begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix} \mathcal{W}(G) + \mathcal{W}(G) \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T + \mathcal{B}(G) = 0. \quad (19)$$

Proof. To compute the gradient $\nabla f(G) \in \mathbb{R}^{n \times m}$, we write the derivative $f'(G) dG$ for some $dG \in \mathbb{R}^{n \times m}$ in gradient form using the trace inner product. We introduce the gradient as: $f'(G) dG = \text{Trace}(\nabla f(G)^T dG)$. From the expression of $f(G)$ we have:

$$f'(G) dG = \text{Trace}(\mathcal{M}'(G)\mathcal{B}(G) + \mathcal{M}(G)\mathcal{B}'(G)).$$

We now separately compute both terms in the above expression. Let

$$\Phi(G, \mathcal{M}) = \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}.$$

Since $G \in \mathcal{D}_{(SL)}$ and $\mathcal{D}_{(SL)}$ is an open set, then, by Lemma 1, we have that $\Phi_{\mathcal{M}}(G, \mathcal{M}) d\mathcal{M}$ given by:

$$\Phi_{\mathcal{M}}(G, \mathcal{M}) d\mathcal{M} = \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T d\mathcal{M} + d\mathcal{M} \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}$$

is surjective and furthermore:

$$\Phi_G(G, \mathcal{M}) dG = \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix}.$$

Since $\Phi(G, \mathcal{M}) + \mathcal{C} = 0$, the Implicit Function Theorem yields the differentiability of $\mathcal{M}(G)$ and the following relation:

$$\begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T \mathcal{M}'(G) + \mathcal{M}'(G) \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix}^T \mathcal{M}(G) + \mathcal{M}(G) \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix} = 0. \quad (20)$$

Consider the Lyapunov Equation (9a) with the unique solution $\mathcal{W}(G)$, provided that G is stabilizing:

$$\begin{bmatrix} A & 0 \\ 0 & s - GL \end{bmatrix} \mathcal{W}(G) + \mathcal{W}(G) \begin{bmatrix} A & 0 \\ 0 & s - GL \end{bmatrix}^T + \mathcal{B}(G) = 0. \quad (21)$$

Subtracting (20) multiplied by $\mathcal{W}(G)$ to the left from (21) multiplied by $\mathcal{M}'(G)$ to the right, taking the trace, and reducing the appropriate terms, yields:

$$\begin{aligned} \text{Trace}(\mathcal{M}'(G)\mathcal{B}(G)) &= \text{Trace}\left(\mathcal{W}(G) \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix}^T \mathcal{M}(G) + \mathcal{M}(G) \begin{bmatrix} 0 & 0 \\ 0 & -dGL \end{bmatrix} \mathcal{W}(G)\right) \\ &= 2\text{Trace}\left(-LW_{12}^T(G)M_{12}(G)dG - LW_{22}(G)M_{22}(G)dG\right). \end{aligned} \quad (22)$$

Similarly, for the second term using the block structure of \mathcal{M} and the Definition of trace, we get:

$$\begin{aligned} \text{Trace}(M(G)\mathcal{B}'(G)) &= \text{Trace}\left(\mathcal{M}(G) \begin{bmatrix} 0 & B dG^T \\ dGB^T & dGG^T + G dG^T \end{bmatrix}\right) \\ &= 2\text{Trace}\left(B^T M_{12}(G) dG + G^T M_{22}(G) dG\right). \end{aligned} \quad (23)$$

Hence, from (22) and (23) we get the closed form expression of the gradient in (18). \square

Note that the expression of the gradient ∇f in (18) is the same as the partial gradient of the Lagrangian $\nabla_G \Gamma$ in (15). Theorem 2 also yields the necessary optimality conditions for the model reduction Problem 1 expressed in terms of (17):

Lemma 3. *If $G \in \mathcal{D}_{(SL)}$ solves the optimization problem (17) corresponding to the model reduction Problem 1, then the following relation holds:*

$$M_{12}^T(G)W_{12}(G)L^T + M_{22}(G)W_{22}(G)L^T = M_{12}^T(G)B + M_{22}(G)G,$$

where M_{12}, M_{22} are blocks in $\mathcal{M}(G)$ and $\mathcal{W}(G)$, as in (10).

We can replace the open set $\mathcal{D}_{(SL)}$ with any sublevel set: $\mathcal{N}_{(SL)}^{G_0} = \{G \in \mathcal{D}_{(SL)} : f(G) \leq f(G_0)\}$, where $G_0 \in \mathcal{D}_{(SL)}$ is any initial stable parameter matrix. Using similar arguments, as in [33], we can show that $\mathcal{N}_{(SL)}^{G_0}$ is a compact set. Then, the Weierstrass Theorem implies that for any given matrix $G_0 \in \mathcal{D}_{(SL)}$, the model reduction Problem 1, given by optimization formulation (13), has a global minimum in the sublevel set $\mathcal{N}_{(SL)}^{G_0}$. We can also show that the gradient $\nabla f(G)$ is Lipschitz continuous on the compact sublevel set $\mathcal{N}_{(SL)}^{G_0}$. Let us briefly sketch the proof of this statement. First, note that $\mathcal{M}(G)$ and $\mathcal{W}(G)$ are continuous functions and, moreover, since they are solutions of algebraic linear systems, there exists a finite $\ell_M > 0$ such that:

$$\|\mathcal{M}(G) - \mathcal{M}(G')\| \leq \ell_M \|G - G'\| \quad \forall G, G' \in \mathcal{N}_{(SL)}^{G_0}.$$

Then, using the expression of $\nabla f(G)$, the compactness of $\mathcal{N}_{(SL)}^{G_0}$, the continuity of $\mathcal{M}(G)$ and $\mathcal{W}(G)$, and the previous relation, we conclude that there exists $\ell_f > 0$ such that:

$$\|\nabla f(G) - \nabla f(G')\| \leq \ell_f \|G - G'\| \quad \forall G, G' \in \mathcal{N}_{(SL)}^{G_0}.$$

This property of the gradient is useful later when analyzing the convergence behavior of the proposed algorithm for the optimization problem (17).

3.4. SDP Approach

Alternatively, problem (13) can be written equivalently in terms of matrix inequalities (semidefinite programming):

$$\begin{aligned} \min_{(G, \mathcal{M})} \quad & \text{Trace} \left(\begin{bmatrix} B \\ G \end{bmatrix}^T \mathcal{M} \begin{bmatrix} B \\ G \end{bmatrix} \right) \\ \text{s.t. : } \quad & \mathcal{M} \succeq 0 \quad \text{and} \\ & \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix} + \begin{bmatrix} C^T C & -C^T C_V \\ -C_V^T C & C_V^T C_V \end{bmatrix} \preceq 0. \end{aligned} \quad (24)$$

Note that the SDP problem (24) is not convex since it contains bilinear matrix inequalities (BMIs). However, using (10), the next result proves that a (local) optimal solution can be obtained through a convex relaxation.

Theorem 3. *If the following convex SDP relaxation:*

$$\begin{aligned} \min_{(G, X_{22}, Y_{22}, Z_{22}), M_{11} \succeq 0, M_{22} \succeq 0} \quad & \text{Trace} (B^T M_{11} B + X_{22}) \\ \text{s.t. : } \quad & S^T M_{22} - L^T Z_{22}^T + M_{22} S - Z_{22} L + C_V^T C_V \preceq Y_{22} \\ & \begin{bmatrix} X_{22} & Z_{22} \\ Z_{22}^T & M_{22} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} A^T M_{11} + M_{11} A + C^T C & -C^T C_V \\ -C_V^T C & Y_{22} \end{bmatrix} \preceq 0 \end{aligned} \quad (25)$$

has a solution, then we can recover a (local) optimal solution of the Problem 1 expressed in terms of (13), through the relations $G = M_{22}^{-1} Z_{22}$ and $\mathcal{M} = \text{diag}(M_{11}, M_{22})$.

Proof. With \mathcal{M} as in (10), we get the equivalent problem:

$$\begin{aligned} \min_{G, \mathcal{M} \succeq 0} \quad & \text{Trace} (B^T M_{11} B + B^T M_{12} G + G^T M_{12}^T B + G^T M_{22} G) \\ \text{s.t. : } \quad & \begin{bmatrix} A^T M_{11} + M_{11} A + C^T C & A^T M_{12} + M_{12} (S - GL) - C^T C_V \\ (S - GL)^T M_{12}^T + M_{12}^T A - C_V^T C & (S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V \end{bmatrix} \preceq 0. \end{aligned} \quad (26)$$

Introducing additional variables, we can reformulate (26) as an SDP subject to BMI constraints. Indeed, we have the equivalent formulation:

$$\begin{aligned} \min_{(G, X_{22}, Y_{22}), \mathcal{M} \succeq 0} \quad & \text{Trace} (B^T M_{11} B + X_{22}) \\ \text{s.t. : } \quad & X_{22} \succeq B^T M_{12} G + G^T M_{12}^T B + G^T M_{22} G, \\ & (S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V \preceq Y_{22}, \\ & \begin{bmatrix} A^T M_{11} + M_{11} A + C^T C & A^T M_{12} + M_{12} (S - GL) - C^T C_V \\ (S - GL)^T M_{12}^T + M_{12}^T A - C_V^T C & Y_{22} \end{bmatrix} \preceq 0. \end{aligned} \quad (27)$$

The Schur complement yields an SDP with a convex objective function but with nonconvex BMI constraints:

$$\begin{aligned} \min_{(G, X_{22}, Y_{22}), \mathcal{M} \succeq 0} \quad & \text{Trace} (B^T M_{11} B + X_{22}) \\ \text{s.t. : } \quad & (S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V \preceq Y_{22}, \\ & \begin{bmatrix} X_{22} - B^T M_{12} G - G^T M_{12}^T B & G^T M_{22} \\ M_{22} G & M_{22} \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} A^T M_{11} + M_{11} A + C^T C & A^T M_{12} + M_{12} (S - GL) - C^T C_V \\ (S - GL)^T M_{12}^T + M_{12}^T A - C_V^T C & Y_{22} \end{bmatrix} \preceq 0. \end{aligned} \quad (28)$$

Problem (28) is not convex since it contains bilinear matrix terms. However, if we assume that $M_{12} = 0$, then (28) can be recast as a convex SDP. That is, for $M_{12} = 0$ in (28), we get:

$$\begin{aligned} & \min_{(G, X_{22}, Y_{22}), M_{11} \succeq 0, M_{22} \succeq 0} \text{Trace}(B^T M_{11} B + X_{22}) \\ & \text{s.t. : } X_{22} \succeq G^T M_{22} G, \\ & (S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V \preceq Y_{22}, \\ & \begin{bmatrix} A^T M_{11} + M_{11} A + C^T C & -C^T C_V \\ -C_V^T C & Y_{22} \end{bmatrix} \preceq 0. \end{aligned} \quad (29)$$

Letting $Z_{22} = M_{22} G$ and using the Schur complement, (29) becomes the convex SDP (25). Moreover, if (25) has a solution, then we can easily recover $G = M_{22}^{-1} Z_{22}$ and $\mathcal{M} = \text{diag}(M_{11}, M_{22})$. Note also that the solution (G, \mathcal{M}) of the convex SDP problem is a (local) optimal solution of the original problem (13), since we impose $M_{12} = 0$. \square

Note that the SDP inequalities in (24) and the corresponding ones in Theorem 3 must be strict to infer asymptotic stability. This can be easily done, replacing 0 with $-\epsilon I$, where I is the identity matrix of appropriate dimension and ϵ is sufficiently small.

Remark 1. (Sufficient conditions for block diagonal Gramian). Theorem 3 shows that we can obtain a (local) optimal solution for the Problem 1 through a convex SDP under the assumption that the error system admits a block diagonal observability Gramian. While diagonal Gramians have recently been exploited in the balanced truncation of positive systems [34,35], the application of block diagonal Gramians in the moment matching-based reduction of general LTI systems is discussed now. In Appendix A we derive sufficient conditions for block diagonal Gramian.

4. Numerical Optimization Algorithms for Problem 1

In this section, we present several optimization algorithms to solve Problem 1 recast as the optimization Problems (13), (17) or (25). Note that the proposed optimization framework is general and flexible, allowing us to easily incorporate physical system constraints in the optimization formulations, such as stability, or structural constraints (positivity, network structure, see also Section 4.3). For example, in this paper, to solve the KKT system (15) of the nonconvex problem (13) or the nonconvex (partial) optimization problem (17), we propose first-order methods, since they are adequate for large-scale optimization Problems, i.e., when the dimension n is very large, see e.g., [36]. Note that second-order methods (e.g., Newton or quasi-Newton [37]) can be applied, but they typically require more expensive computations at each iteration (e.g., evaluation of the Hessians and finding solutions of linear systems), making them intractable when the dimension n is large. Efficient implementations of the (quasi-)Newton algorithms will be considered in our future work.

4.1. Gradient Type Method for the KKT System

One optimization algorithm that can be used for solving the KKT system (15) of problem (13) is the gradient method. Starting from an initial triplet $(\mathcal{W}_0, \mathcal{M}_0, G_0)$, update:

$$\begin{aligned} \mathcal{W}_{k+1} &= \mathcal{W}_k + \alpha_k \nabla_{\mathcal{W}} \Gamma(\mathcal{W}_k, \mathcal{M}_k, G_k), \\ \begin{bmatrix} \mathcal{M}_{k+1} \\ G_{k+1} \end{bmatrix} &= \begin{bmatrix} \mathcal{M}_k \\ G_k \end{bmatrix} - \alpha_k \nabla_{(\mathcal{M}, G)} \Gamma(\mathcal{W}_k, \mathcal{M}_k, G_k), \end{aligned} \quad (30)$$

where α_k is a stepsize selected to minimize an appropriate merit function in the search direction at each step. Under some mild assumptions, it is possible to prove that the iterative process (30) converges locally to a KKT point, see, e.g., [37] (Chapter 14). Moreover, if we start sufficiently close to a KKT point, we can even choose α_k constant and the sequence will converge linearly to a KKT point, with a speed of convergence depending on the starting point, see [37] (Chapter 14) for details.

Initialization of the Algorithm

If the convex SDP relaxation (25) admits a solution, then we can consider the solution provided by this relaxation as a starting point in the update (30), i.e., G_0 with $G_0 = M_{22}^{-1}Z_{22}$ and the block diagonal matrix $M_0 = \text{diag}(M_{11}, M_{22})$. Moreover, we can take \mathcal{W}_0 as the solution of (9a) with $G = G_0$. Sufficient conditions to guarantee the feasibility of the SDP relaxation, i.e., to guarantee that the error system admits a block diagonal observability Gramian, are given in the Appendix A. Otherwise, we can fix S and L such that the pair (S, L) is observable, and select a set $\{\lambda_1, \dots, \lambda_v\} \subset \mathbb{C}^-$. Then, from basic control theory, it is known that there exists (a stabilizing) G_0 , computed by standard control algorithms, such that the spectrum $\sigma(S - G_0L) = \{\lambda_1, \dots, \lambda_v\}$ (e.g., pole placement procedures).

The algorithm (30) has cheap iteration costs since it requires only matrix multiplications, making it adequate for large-scale systems. However, the update in (30) has the disadvantage that only the limit points of the sequence G_k lead to a stable model, whereas the intermediate iterates may lead to unstable models. Note that this disadvantage is also encountered in other model reduction algorithms, such as IRKA [17].

4.2. Gradient Method for the Partial Minimization Problem

We have proven that the objective function of the nonconvex optimization problem (17) is differentiable, with the gradient given in (18). Moreover, the gradient is Lipschitz continuous on any compact set. Then, we can apply a gradient method to solve (17). Starting from the initial stable matrix $G_0 \in \mathcal{D}_{(S,L)}$, we consider the following update:

$$G_{k+1} = G_k - \alpha_k \nabla f(G_k), \quad (31)$$

where the stepsize α_k can be chosen through a backtracking procedure or constant in the interval $(0, 2/\ell_f)$ (recall that ℓ_f denotes the Lipschitz constant of the gradient). With such choices for the stepsize and using the Lipschitz gradient property for the objective function, the sequence of value functions $f(G_k)$ is nonincreasing [36]:

$$f(G_{k+1}) \leq f(G_k) - \Delta \cdot \|\nabla f(G_k)\|^2 \quad \forall k \geq 0,$$

for some constant $\Delta > 0$, where $\|\cdot\|$ denotes the Frobenius norm. Hence, all the iterates remain in the compact sublevel set $\mathcal{N}_{(SL)}^{G_0}$. Moreover, since f is bounded from below by zero, then for any positive integer k , a global convergence rate follows from the previous descent inequality:

$$\min_{i=0:k-1} \|\nabla f(G_i)\|^2 \leq \frac{f(G_0) - f^*}{\Delta \cdot k} \quad \forall k \geq 0,$$

where f^* is the optimal value of problem (17). Under some mild assumptions, such as the Hessian of f at a local minimum being positive definite and bounded, then, starting sufficiently close to this local optimum, the gradient iteration converges linearly to this solution [36]. Therefore, the speed of convergence of this iterative process depends on the starting point. As a starting point, we can consider the initialization procedures described in Section 4.1. Note that, although the iterative process (31) requires solving two Lyapunov equations depending on the dimension of the original system, n , since these equations contain a large fixed part corresponding to the original system that can be solved before the start of the iterative process, it is reasonably cheap. Moreover, the variation of the matrix G can be viewed as a low-rank update of the error Lyapunov equation. Then, one only has to solve a small Lyapunov equation and a sparse Sylvester equation at each iteration, not very expensive, see, e.g., for more details [38]. Finally, the update in (31) has the advantage that any iterate G_k leads to a stable reduced order model, whereas for the iteration (30), only the limit points of the generated sequence G_k lead to a stable model.

4.3. Convex SDP Relaxation

There are several methods available to solve SDP Problems with convex objective functions subject to BMI constraints, see, e.g., [39]. However, there are more efficient solvers for the convex SDPs (e.g., problem (25)) that can scale to large instances, such as first order methods or interior point methods [36]. Note that, in the general case of LTI systems, the convex SDP relaxation (25) is not exact, since imposing the condition $M_{12} = 0$ renders the solution of the original SDP problem (24) (sub)optimal. If the convex SDP relaxation (25) admits a solution, then we can initialize the proposed gradient-based methods with the solution yielded by the relaxation. Sufficient conditions to guarantee the feasibility of the SDP relaxation, i.e., guarantee that the error system admits a block diagonal observability Gramian, are given in the Appendix A.

5. Illustrative Examples

We illustrate the numerical efficiency of our theory and algorithms with a CD-player [17] and on the discretized heat equation modeled as a positive system [40]. In particular, we compute and compare reduced order models achieving (possibly) the minimum H_2 -norm.

5.1. CD Player

We consider the model of the dynamics between a lens actuator and the radial arm position in a portable CD player, i.e., a single input single output system with $n = 120$, see, e.g., [1,17] for values of the matrices $A \in \mathbb{R}^{120 \times 120}$, $B \in \mathbb{R}^{120 \times 1}$, and $C \in \mathbb{R}^{1 \times 120}$. We compute the optimal models through the gradient-based solution to optimization problem (17) at orders $\nu = 1:10$. In Figure 1, we plot the H_2 -norm of the approximation error versus the reduced order index. The solution yields a reduced order model, with small H_2 -norm for the approximation error, in the family of reduced order models that match ν moments at ν fixed interpolation points. The interpolation points have been chosen at dense frequencies (black), e.g., 0, 0.2, 0.4, 0.6, ..., as well as rare (blue), e.g., 0, 2, 4, 6, etc. The latter choice of the interpolation points yields a lower H_2 error norm. Note that matching the moment at zero ensures the preservation of the DC-gain of the step response of the system.

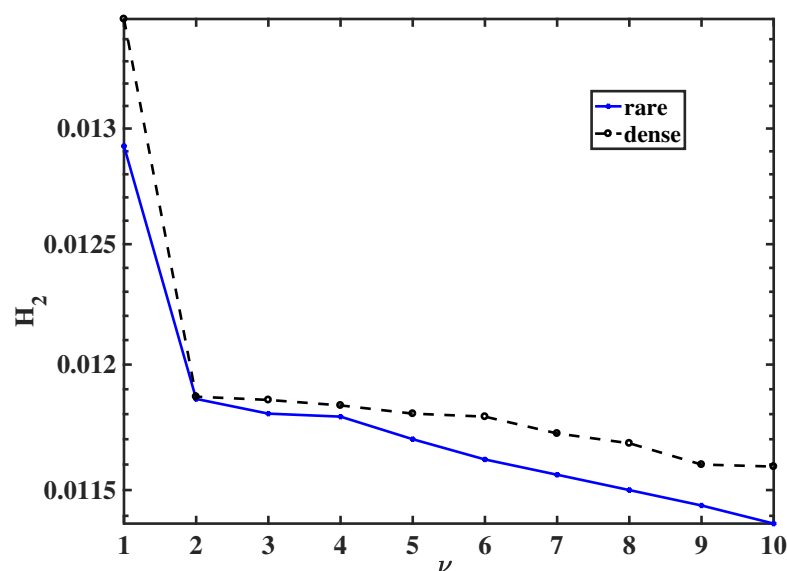


Figure 1. Relative H_2 -norm of the approximation error versus the approximation order ν for rare/-dense interpolation points for the CD-player.

Furthermore, Table 1 shows the results of the proposed gradient method-based solution of Problem 1 versus the results of applying the IRKA [17] and balanced truncation [5], for orders $\nu = 2$ and $\nu = 6$. Note that IRKA produces better H_2 error norms than the

gradient method. This is due to the fact that IRKA is searching for a solution over the whole space of ν order models parametrized in (F, G, H) , while the proposed method is searching for a solution only in the subspace of *stable* ν order models parameterized in G . Consequently, IRKA is usually producing a stable reduced model only at the solution, while we can stop our algorithm (31) at any iteration and still yield a stable reduced model. This argument is further emphasized by the placement of the poles of the (optimal) approximations yielded by these two methods, in the complex left half plane. Note that IRKA gives the maximum real part of the poles much closer to the imaginary axis (very close to instability or unstable even) than the proposed method, e.g., $-2.25 \cdot 10^{-5}$ versus $-3.9 \cdot 10^{-1}$ for $\nu = 2$ and $2.27 \cdot 10^{-5}$ versus $-4.2 \cdot 10^{-1}$ for $\nu = 6$, respectively. On the other hand, balanced truncation yields a stable model with similar H_2 error norms as the proposed method. However, even if the approximant is stable, the poles are very close to the imaginary axis. One remedy for improving the quality of the solutions of Problem 1 is to design algorithms that make use of the curvature information. For example, since we derived explicit expressions for the gradient of the objective function, we can consider the quasi-Newton methods instead of the gradient methods. Moreover, in general, the systematic selection of interpolation points is an open issue, and is typically left to the user, depending on the application goal. However, these issues are beyond the scope of this paper and they will be investigated in the future work.

Table 1. Comparison of the gradient method-based solution of Problem 1 with algorithm (31) vs. IRKA and balanced truncation, with an arbitrary set of interpolation points.

ν	Gradient Method for Problem 1			IRKA [17]			Balanced Truncation	
	H_2 Norm	Gradient Norm	max Re Pole	H_2 Norm	Gradient Norm	max Re Pole	H_2 Norm	max Re Pole
2	$1.1 \cdot 10^{-2}$	$4.3 \cdot 10^{-5}$	$-3.9 \cdot 10^{-1}$	$4.4 \cdot 10^{-4}$	$1.5 \cdot 10^{-6}$	$-2.2 \cdot 10^{-5}$	$7.5 \cdot 10^{-2}$	$-2.2 \cdot 10^{-5}$
6	$5.6 \cdot 10^{-3}$	$3.5 \cdot 10^{-5}$	$-4.2 \cdot 10^{-1}$	$6.4 \cdot 10^{-5}$	$1.8 \cdot 10^{-5}$	$-2.3 \cdot 10^{-5}$	$4.8 \cdot 10^{-3}$	$-2.2 \cdot 10^{-5}$

We further investigate several structured choices of the interpolation points. For order $\nu = 6$, we choose the interpolation points as the mirror images of the poles of the IRKA model and of the balanced truncation model of the CD-player system, respectively.

We compare the results of the proposed method to IRKA and the balanced truncation in Table 2. The structured choices of interpolation points yield similar results to the arbitrary choice of interpolation points, e.g., $\nu = 6$. Note that, when the interpolation points are generated by IRKA or by balanced truncation, the interpolation at zero is lost, hence the DC-gain is not preserved. Furthermore, in this case, the poles of the resulting solution are closer to the imaginary axis.

Table 2. Comparison of the gradient method-based solution of Problem 1 with algorithm (31) vs. IRKA and balanced truncation, with the interpolation points chosen as the mirror images of the poles of the IRKA and BT models, respectively.

ν	Interp. Points by:	Gradient Method for Problem 1			H_2 Norm $\cdot 10^{-5}$	IRKA [17]		Balanced Truncation	
		H_2 Norm	Gradient Norm	max Re Pole		Gradient Norm $\cdot 10^{-5}$	max Re Pole $\cdot 10^{-5}$	H_2 Norm	max Re Pole
6	IRKA	$2.2 \cdot 10^{-2}$	$1.3 \cdot 10^{-5}$	$-2.5 \cdot 10^{-3}$	6.4	1.8	-2.3	$4.8 \cdot 10^{-3}$	$-2.2 \cdot 10^{-5}$
	BT	$4.6 \cdot 10^{-3}$	$1.9 \cdot 10^{-5}$	$-2.3 \cdot 10^{-5}$					

5.2. Heat Equation

We consider the 2-dimensional heat equation on a square, see [34] for details:

$$\dot{T} = \partial^2 T / \partial x^2 + \partial^2 T / \partial y^2,$$

with the control of the Dirichlet boundary conditions of the four edges.

Using the finite difference discretization on a uniform grid with N points on each axes leads to the linear system

$$\dot{T} = AT + Bu, \quad y = CT,$$

where $n = N^2, m = 4$ and $p = 1$. Hence, $A \in \mathbb{R}^{N^2 \times N^2}$ is the Poisson matrix satisfying the Metzler condition, while $B \in \mathbb{R}^{N^2 \times 4}$ has entries in the set $\{0, 1\}$, both matrices being very sparse. Moreover, we take as the output the global average temperature yielding $C \in \mathbb{R}^{1 \times N^2}$. Hence, the system is positive, see [40]. We consider $N = 100$, leading to a mesh with $n = 10,000$ grid points and, for simplicity, we take $u_2 = u_3 = u_4 = 0$. We consider the model reduction Problem 1 and apply the SDP relaxation (25). For $N = 100$, we get a system of order 10,000, for the SDP approach needs hours to calculate a reduced model, due to the high complexity of conventional SDP-solvers. In the case of a large-scale system, we apply the gradient iteration to the KKT system to decrease the system to the order 100, followed by the solution of SDP relaxation (25), reducing the order to $\nu = 10$. The interpolation points are chosen as $s_i = i, i = 1 : 10$, without a specific constraint. The Bode magnitudes plots of K and \hat{K} , denoted by K_{reduced} , for Problem 1, are shown in Figure 2. Note that the low frequency responses of \hat{K} and $K(s)$ are very close, see Figure 2b. The H_2 -norm error obtained with the 10th order system is 7.41×10^{-2} .

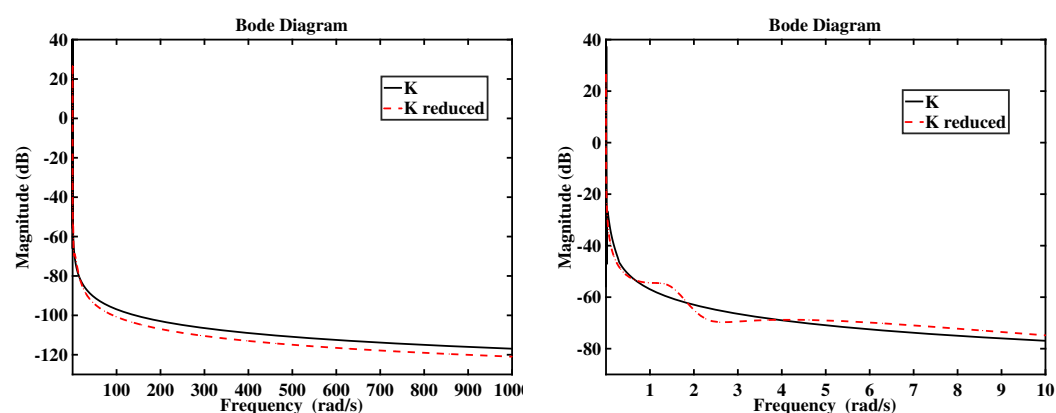


Figure 2. Bode magnitudes of the given system (solid black) and the approximation (dashed red) yielded by Problem 1. (left) Magnitude, (right) Low frequency zoom.

6. Conclusions

In this paper, we have formulated an optimization problem with respect to the H_2 -norm minimal approximation error in a family of reduced order models that match a prescribed set of fixed moments. We have derived first-order optimality conditions and numerical solutions have been proposed in terms of the gradient method or SDP. Using test examples such as the CD-player and the heat equation, we have also illustrated the numerical efficiency of the results.

Author Contributions: Conceptualization, I.N.; methodology, I.N. and T.-C.I.; software, T.-C.I.; validation, I.N. and T.-C.I.; formal analysis, I.N.; investigation, I.N. and T.-C.I.; resources, T.-C.I.; data curation, I.N. and T.-C.I.; writing—original draft preparation, T.-C.I.; writing—review and editing, I.N.; supervision, I.N.; project administration, I.N.; funding acquisition, I.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has received funding from the NO Grants 2014–2021, under project ELO-Hyp, contract no. 24/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Block Diagonal Gramians for General LTI Systems

In this section, we derive sufficient conditions for a general LTI system to admit block diagonal Gramians, i.e., derive the sufficient conditions for the feasibility of the SDP relaxation (25) in the case of general LTI systems. In other words, we write the sufficient conditions to guarantee that the error system admits a block diagonal observability Gramian. For this, we need the following result valid for any two vectors u and v :

$$u^T v + v^T u \preceq u^T P^{-1} u + v^T P v, \quad \forall P \succ 0. \quad (\text{A1})$$

The inequality follows from the relation $(u - Pv)^T P^{-1} (u - Pv) \succeq 0$, for $P \succ 0$. We are interested in deriving sufficient conditions to guarantee that the SDP (24) admits a feasible (G, \mathcal{M}) with \mathcal{M} of block diagonal form and, consequently, that the SDP relaxation (25) is well-defined.

Theorem A1. *Given the stable minimal system (1), there exists a stable reduced order model (4) such that the error system admits a block diagonal observability Gramian if the following conditions hold:*

$$\begin{aligned} A^T M_{11} + M_{11} A + C^T C + C^T C_V P_{x\xi}^{-1} C_V^T C &\preceq 0, \\ (S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V + P_{x\xi} &\preceq 0, \end{aligned} \quad (\text{A2})$$

for some matrix $\mathcal{M} = \text{diag}(M_{11}, M_{22}) \succeq 0$ and $P_{x\xi} \succeq 0$.

Proof. Note that the feasible set of (24) is nonempty if $\mathcal{M} \succeq 0$ and the following inequality holds:

$$\begin{bmatrix} x \\ \xi \end{bmatrix}^T \left(\begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix}^T \mathcal{M} + \mathcal{M} \begin{bmatrix} A & 0 \\ 0 & S - GL \end{bmatrix} + \begin{bmatrix} C^T C & -C^T C_V \\ -C_V^T C & C_V^T C_V \end{bmatrix} \right) \begin{bmatrix} x \\ \xi \end{bmatrix} \leq 0, \quad (\text{A3})$$

$\forall x, \xi$. Since $\mathcal{M} = \text{diag}(M_{11}, M_{22})$, (A3) is equivalent to

$$\begin{aligned} x^T (A^T M_{11} + M_{11} A + C^T C) x + \xi^T ((S - GL)^T M_{22} + M_{22} (S - GL) + C_V^T C_V) \xi \\ - x^T C^T C_V \xi - \xi^T C_V^T C x \leq 0, \end{aligned}$$

$\forall x, \xi$. Using now (A1) in the last term we get that

$$-x^T C^T C_V \xi - \xi^T C_V^T C x \leq x^T C^T C_V P_{x\xi}^{-1} C_V^T C x + \xi^T P_{x\xi} \xi \quad \forall x, \xi.$$

Hence, if the inequality

$$\begin{aligned} x^T (A^T M_{11} + M_{11} A + C^T C + C^T C_V P_{x\xi}^{-1} C_V^T C) x + \xi^T ((S - GL)^T M_{22} + M_{22} (S - GL) \\ + C_V^T C_V + P_{x\xi}) \xi \leq 0, \end{aligned}$$

$\forall x, \xi$ holds, then also (A3) is valid. Then (A2) follows. \square

Theorem A1 provides sufficient conditions and also a procedure to construct a reduced order model leading to an error system that admits a block diagonal observability Gramian.

Indeed, let us, for example, fix $L, M_{22} = I_\nu$ and some matrix $P_{x\xi} \succ 0$. Then, the existence of a solution (G, Π) of the system:

$$(S - GL)^T + (S - GL) = -\left((C\Pi)^T(C\Pi) + P_{x\xi}\right),$$

$$A\Pi + BL = \Pi S,$$

together with the existence of an $M_{11} \succ 0$ satisfying $A^T M_{11} + M_{11} A + C^T C + C^T C_V P_{x\xi}^{-1} C_V^T C \preceq 0$ guarantee that we have a reduced order model yielding a block diagonal observability Gramian of the error system.

References

1. Antoulas, A.C. *Approximation of Large-Scale Dynamical Systems*; SIAM: Philadelphia, PA, USA, 2005.
2. Moore, B.C. Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Autom. Control* **1981**, *26*, 17–32. [\[CrossRef\]](#)
3. Enns, D.F. Model reduction with balanced realizations: an error bound and a frequency weighted generalization. *Conf. Decis. Control* **1984**, 127–132. [\[CrossRef\]](#)
4. Selga, R.C.; Lohann, B.; Eid, E. Stability Preservation in Projection-based Model Order Reduction of Large Scale Systems. *Eur. J. Control* **2012**, *2*, 122–132. [\[CrossRef\]](#)
5. Benner, P.; Quintana-Orti, E.S.; Quintana-Orti, G. Balanced Truncation Model Reduction of Large-Scale Dense Systems on Parallel Computers. *Math. Comput. Model. Dyn. Syst.* **2000**, *6*, 383–405. [\[CrossRef\]](#)
6. de Villemagne, C.; Skelton, R.E. Model reductions using a projection formulation. *Int. J. Control* **1987**, *46*, 2141–2169. [\[CrossRef\]](#)
7. Antoulas, A.C.; Ball, J.A.; Kang, J.; Willems, J.C. On the solution of the minimal rational interpolation problem. *Lin. Alg. Its Appl.* **1990**, *137/138*, 511–573. [\[CrossRef\]](#)
8. Gugercin, S.; Willcox, K. Krylov projection framework for Fourier model reduction. *Automatica* **2008**, *44*, 209–215. [\[CrossRef\]](#)
9. van Dooren, P. The Lanczos algorithm and Padé approximation. In *Benelux Meeting on Systems and Control*; Minicourse: Dept. Mathematical Engineering, Univ. Catholique de Louvain: Louvain-la-Neuve, Belgium, 1995.
10. Astolfi, A. Model reduction by moment matching for linear and nonlinear systems. *IEEE Trans. Autom. Control* **2010**, *50*, 2321–2336. [\[CrossRef\]](#)
11. Gallivan, K.; Vandendorpe, A.; Dooren, P.V. Sylvester equations and projection based model reduction. *J. Comp. Appl. Math.* **2004**, *162*, 213–229. [\[CrossRef\]](#)
12. Ionescu, T.C.; Astolfi, A.; Colaneri, P. Families of moment matching based, low order approximations for linear systems. *Syst. Control Lett.* **2014**, *64*, 47–56. [\[CrossRef\]](#)
13. Grimme, E.; Sorensen, D.; Dooren, P.V. Model reduction of state space systems via an implicitly restarted Lanczos method. *Numer. Algorithms* **1995**, *12*, 1–31. [\[CrossRef\]](#)
14. Ionescu, T.C. Two-sided time-domain moment matching for linear systems. *IEEE Trans. Autom. Control* **2016**, *61*, 2632–2637. [\[CrossRef\]](#)
15. Xu, Y.; Zeng, T. Optimal H_2 model reduction for large scale MIMO systems via tangential interpolation. *Int. J. Numer. Model.* **2011**, *8*, 174–188.
16. Anic, B.; Beattie, C.A.; Gugercin, S.; Antoulas, A.C. Interpolatory weighted- H_2 model reduction. *Automatica* **2013**, *49*, 1275–1280. [\[CrossRef\]](#)
17. Gugercin, S.; Antoulas, A.C.; Beattie, C.A. H_2 model reduction for large-scale dynamical systems. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 609–638. [\[CrossRef\]](#)
18. Meier, L.; Luenberger, D.G. Approximation of linear constant systems. *IEEE Trans. Autom. Control* **1967**, *12*, 585–588. [\[CrossRef\]](#)
19. Gallivan, K.; Grimme, E.; Dooren, P.V. A rational Lanczos algorithm for model reduction. *Numer. Algorithms* **1996**, *12*, 33–63. [\[CrossRef\]](#)
20. Panzer, H.K.F. Model Order Reduction by Krylov Subspace Methods with Global Error Bounds and Automatic Choice of Parameters. Ph.D. Thesis, Technische Universität München, München, Germany, 2014.
21. Flagg, G.; Beattie, C.; Gugercin, S. Convergence of the iterative rational Krylov algorithm. *Syst. Control. Lett.* **2012**, *61*, 688–691. [\[CrossRef\]](#)
22. Gugercin, S. An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems. *Linear Algebra Its Appl.* **2008**, *428*, 1964–1986. [\[CrossRef\]](#)
23. Panzer, H.K.F.; Jaensch, S.; Wolf, T.; Lohmann, B. A Greedy Rational Krylov Method for H_2 -Pseudooptimal Model Order Reduction with Preservation of Stability. *Am. Control Conf.* **2013**, 5512–5519. [\[CrossRef\]](#)
24. Necoara, I.; Ionescu, T.C. H_2 model reduction of linear network systems by moment matching and optimization. *IEEE Trans. Autom. Control* **2020**, *65*, 5328–5334. [\[CrossRef\]](#)
25. Wang, W.-G.; Jiang, Y.-L. H_2 optimal model order reduction on the Stiefel manifold for the MIMO discrete system by the cross-Gramian. *Math. Comput. Model. Dyn. Syst.* **2018**, *24*, 610–625. [\[CrossRef\]](#)
26. Sato, K. Riemannian Optimal Model Reduction of Stable Linear Systems. *IEEE Access* **2019**, *7*, 14689–14698. [\[CrossRef\]](#)

27. De Souza, E.; Bhattacharyya, S.P. Controllability, observability and the solution of $AX - XB = C$. *Linear Algebra Its Appl.* **1981**, *39*, 167–188. [[CrossRef](#)]
28. Astolfi, A. Model reduction by moment matching, steady-state response and projections. In Proceedings of the 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, 15–17 December 2010; pp. 5344–5349.
29. Grimme, E.J. Krylov Projection Methods for Model Reduction. Ph.D. Thesis, ECE Department University of Illinois, Urbana-Champaign, IL, USA, 1997.
30. Jaimoukha, I.M.; Kasenally, E.M. Implicitly restarted Krylov subspace methods for stable partial realizations. *SIAM J. Matrix Anal. Appl.* **1997**, *18*, 633–652. [[CrossRef](#)]
31. Dooren, P.V.; Gallivan, K.A.; Absil, P.-A. \mathcal{H}_2 -optimal model reduction of MIMO systems. *Appl. Math. Lett.* **2008**, *21*, 1267–1273. [[CrossRef](#)]
32. Sato, K. Optimal model reduction for sparse linear systems. *arXiv* **2020**, arXiv:2009.14498.
33. Toivonen, H. A globally convergent algorithm for the optimal constant output feedback problem. *Int. J. Control* **1985**, *41*, 1589–1599. [[CrossRef](#)]
34. Grussler, C.; Damm, T. A symmetry approach for balanced truncation of positive linear systems. In Proceedings of the IEEE Conference on Decision and Control, Maui, HI, USA, 10–13 December 2012; pp. 4308–4313.
35. Reis, T.; Virnik, E. Positivity preserving balanced truncation for descriptor systems. *SIAM J. Control Optim.* **2009**, *48*, 2600–2619. [[CrossRef](#)]
36. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*; Springer: Berlin/Heidelberg, Germany, 2004; ISBN 1-4020-7553-7.
37. Luenberger, D. *Linear and Nonlinear Programming*; Springer: Berlin/Heidelberg, Germany, 2003; ISBN: 978-3-030-85450-8.
38. Benner, P.; Köhler, M.; Saak, J. *Sparse-Dense Sylvester Equations in H_2 -Model Order Reduction*; Max Planck Institute Magdeburg Preprints MPIMD/11-11: Magdeburg, Germany 2011.
39. Kocvara, M.; Stingl, M. Pennon: Software for linear and nonlinear matrix inequalities. In *Handbook on Semidefinite, Conic and Polynomial Optimization*; Anjos, M., Lasserre, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 755–794.
40. Rantzer, A. Scalable control of positive systems. *Eur. J. Control* **2015**, *24*, 72–80. [[CrossRef](#)]