Research Article

Cross-Gramian-Based Combined State and Parameter Reduction for Large-Scale Control Systems

Christian Himpe and Mario Ohlberger

Institute for Computational and Applied Mathematics, University of Münster, Einsteinstraße 62, 48149 Münster, Germany

Correspondence should be addressed to Christian Himpe; christian.himpe@uni-muenster.de

Received 14 April 2014; Accepted 1 June 2014; Published 2 July 2014

Academic Editor: Victor Sreeram

Copyright © 2014 C. Himpe and M. Ohlberger. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work introduces the empirical cross-gramian for multiple-input-multiple-output systems. The cross-gramian is a tool for reducing the state space of control systems, by conjoining controllability and observability information into a single matrix and does not require balancing. Its empirical gramian variant extends the applicability of the cross-gramian to nonlinear systems. Furthermore, for parametrized systems, the empirical gramians can also be utilized for sensitivity analysis or parameter identification and thus for parameter reduction. This work also introduces the empirical joint gramian, which is derived from the empirical cross-gramian. The joint gramian allows not only a reduction of the parameter space but also the combined state and parameter space reduction, which is tested on a linear and a nonlinear control system. Controllability- and observability-based combined reduction methods are also presented, which are benchmarked against the joint gramian.

1. Introduction

The evaluation of large-scale dynamical systems, which arise, for example, from complex networks or discretized partial differential equations, may require model reduction due to limitations in computing power or memory. A reduction of the state space generates a surrogate model resembling the same dynamics up to a small error. For parametrized systems, the model order reduction has to take into account the associated parameter space to ensure the validity of the reduced order model. If the parameter space is of high dimension, a repeated evaluation at various locations of the parameter space, for example, during optimization of inverse problems, may also necessitate a model reduction, yet for the parameter space. This contribution is concerned with combined state and parameter reduction, targeting models with high-dimensional state and parameter spaces.

The efficient reduction of large-scale nonlinear control systems is a challenging task, especially in the case of parametrized systems, with high-dimensional state and parameter spaces, where a combined reduction of parameters and states may be required to allow repeated evaluation. For instance, an inverse problem on a neural network with many nodes and unknown connectivity, modeled as a parametrized nonlinear control system, requires long runtimes during parameter estimation due to system size and parameter count. Large-scale neural networks have widespread use, such as forward control problems on artificial neural networks or inverse problems on biological neural networks. A real-life example is the reconstruction of connectivity between brain regions from activity measurements like EEG or fMRI (see, e.g., [1]).

To lower computational complexity, the parameter and state spaces are to be confined to low-dimensional subspaces without affecting the systems dynamics significantly. Projection-based model order reduction techniques are concerned with determining projections to such subspaces, mapping the high-dimensional model to a reduced order low-dimensional surrogate model.

The methods presented in this work are rooted in balanced truncation [2] and proper orthogonal decomposition (POD) [3]. As an alternative to the here presented method using empirical gramians, another class of balancing-related approaches focuses on solving Lyapunov and Sylvester equations (see, e.g., [4]). For parametrized systems, the reduced-basis also method [5] should be noted here.
Since the number of a systems inputs and outputs usually remains fixed, the maps to and from the intermediary states characterize the reducibility of a system [6]. The balanced truncation approach, introduced in [2], balances a system in terms of controllability and observability, where controllability quantifies how well a state is driven by the input and observability quantifies how well changes in a state are reflected in the output. Excluding the least controllable and observable states by truncating the balanced system, a reduced order mapping from inputs to outputs is approximated.

A linear time-invariant control system is composed of a linear dynamic system and a linear output transformation:

\[ \dot{x}(t) = Ax(t) + Bu(t), \]
\[ y(t) = Cx(t), \]

with states \( x(t) \in \mathbb{R}^n \), input or control \( u(t) \in \mathbb{R}^m \), and outputs \( y(t) \in \mathbb{R}^r \). The system matrix \( A \in \mathbb{R}^{n \times n} \) transforms the states, the input matrix \( B \in \mathbb{R}^{n \times m} \) introduces external input or control, and the output matrix \( C \in \mathbb{R}^{r \times n} \) transforms the states to the outputs.

Controllability and observability can be assessed through the associated controllability gramian \( W_C := \int_0^\infty e^{At} BB^T e^{A^T t} dt \) and observability gramian \( W_O := \int_0^\infty e^{A^T t} C^T Ce^{At} dt \). Classically, \( W_C \) and \( W_O \) are computed as the smallest positive semi-definite solutions of the Lyapunov equations \( AW_C + W_C A^T = -BB^T \) and \( A^T W_O + W_O A = -C^T C \), respectively. To make a compound statement about controllability and observability, \( W_C \) and \( W_O \) have to be balanced [7].

The singular values of the resulting balanced gramian correspond to the Hankel singular values of the system, with their magnitude describing how controllable and observable the associated state is.

This work focuses on cross-gramian-based methods for model reduction, which combine controllability and observability information into one gramian and is elaborately described in [8]. The cross-gramian \( W_X := \int_0^\infty e^{At} BC e^{A^T t} dt \) was introduced in [9] and corresponds to a solution of the Sylvester equation \( AW_X + W_X A = -BC \).

An alternative to solving the Lyapunov or Sylvester matrix equations, apart from the analytic approaches, for example, in [10], is the method of empirical gramians, which was introduced in the works [11, 12] and enables the computation of gramian matrices also for nonlinear systems by mere basic vector and matrix operations. This concept was extended among others in [13] providing more general input signals. Particularly noted should be [14, 15] for developing the empirical cross-gramian for single-input-single-output (SISO) systems in the context of sensitivity analysis.

In this paper the empirical cross-gramian is generalized to be applicable to multiple-input-multiple-output (MIMO) systems. For the gramian-based parameter reduction, the groundwork has been laid by [16] from the observability and by [17] from the controllability point of view. From the cross-gramian perspective of parameter reduction, a new gramian, namely, the joint gramian, is introduced in this work. Furthermore, the concept of gramian-based combined state and parameter reduction is established. Using empirical gramians, it is shown that combined reduction allows efficient model order reduction of linear and nonlinear control systems.

To begin, the cross-gramian and its properties are reviewed in Section 2. Next, the empirical cross-gramian for MIMO systems is developed in Section 3. Section 4 introduces combined state and parameter reduction in two variants: first, an observability- and, second, a controllability-based approach; the former is enhanced to a cross-gramian-based combined reduction, which is presented in Section 5. Finally, numerical experiments are conducted in Section 6 comparing the newly presented methods for a linear and nonlinear neural network as well as a nonlinear benchmark problem.

2. Review of the Cross-Gramian

A brief review of the cross-gramian along with its application to model reduction of linear time-invariant control systems is given next. The cross-gramian \( W_X \) (also known by the symbol \( W_{CO} \)) was introduced in a sequence of works [9, 18–23] and encodes controllability and observability into a single gramian matrix, defined as the product of controllability and observability operator, it can only be computed for square (a system with the same number of inputs and outputs) and asymptotically stable systems:

\[ W_X := \int_0^\infty e^{At} BC e^{A^T t} dt. \]

Equivalently, the cross-gramian is given as a solution to the Sylvester equation \( AW_X + W_X A = -BC \). Approximate solutions for the Sylvester equation were discussed in [24–26]. If the system is also symmetric, the following relation between the cross, the controllability, and observability gramian holds [9]:

\[ W_X^2 = W_C W_O \implies |\lambda(W_X)| = \sqrt{|\lambda(W_C W_O)|}. \]

While a SISO system is always symmetric [9], a linear MIMO system not only requires the same number of inputs and outputs, but also the system gain \( G = -CA^{-1}B \) has to be symmetric [24]; then a symmetric transformation \( J \), with \( AJ = JA^T \) and \( B = C^T J \) exist. Trivially, for \( J = I \) the system would be restricted by \( A = A^T \) and \( B = C^T \); such a system is called state space symmetric.

As presented in [9], the trace of the cross-gramian equals half the gain of a SISO system \((A, b, c)\) where now \( b \in \mathbb{R}^{mc \times 1} \) and \( c \in \mathbb{R}^{1 \times n} \):

\[ \text{tr}(W_X) = -\frac{1}{2} c A^{-1} b. \]

Because the trace equals the sum of eigenvalues, the cross-gramians eigenvalues are associated with the system gain (4). This result was used in [14, 15] for parameter identification purposes, using the system gain as a sensitivity measure. An extension of (4) from [9, Theorem 3] for MIMO systems is developed (see also [27]) next.
Corollary 1. Given a linear, square, asymptotically stable MIMO system, then the trace of the cross-gramian relates to the system gain as follows:

\[
\text{tr} (W_X) = -\frac{1}{2} \text{tr} (CA^{-1}B). \quad (5)
\]

Proof. For an asymptotically stable system, the trace of the cross-gramian, in the form of (2), is given by

\[
\text{tr} (W_X) = \text{tr} \left( \int_0^\infty e^{tA}BCe^{tA} \, dt \right)
\]

\[
= \int_0^\infty \text{tr} \left( e^{tA}BCe^{tA} \right) \, dt
\]

\[
= \int_0^\infty \text{tr} \left( Ce^{tA}B \right) \, dt
\]

\[
= \text{tr} \left( C \int_0^\infty e^{tA}B \, dt \right)
\]

\[
= \text{tr} \left( C \left( \frac{1}{2}A^{-1} \right) B \right)
\]

\[
= -\frac{1}{2} \text{tr} (CA^{-1}B).
\]

Employing the cross-gramian instead of controllability and observability gramian means only a single gramian has to be computed. And since no balancing is required, the truncation procedure can be simplified to a direct truncation ([28], [8, Ch. 12.3]). A balancing transformation can be truncated and observability gramian means only a single gramian has to be computed. This leads to the trace of the cross-gramian, in the form of (2), is given by

\[
\text{tr} (W_X) = \int_0^\infty \text{tr} \left( e^{tA}BCe^{tA} \right) \, dt
\]

\[
= \int_0^\infty \text{tr} \left( Ce^{tA}B \right) \, dt
\]

\[
= \text{tr} \left( C \int_0^\infty e^{tA}B \, dt \right)
\]

\[
= \text{tr} \left( C \left( \frac{1}{2}A^{-1} \right) B \right)
\]

\[
= -\frac{1}{2} \text{tr} (CA^{-1}B).
\]

Apart from truncation-based model reduction, the cross-gramian has applications, for example, in system identification [10] and decentralized control [29–31] by computing a participation matrix (see [32]) based on the cross-gramian. Lastly, the cross-gramian also has the benefit of conveying more information than controllability and observability gramian, since the system's Cauchy index is given by the cross-gramian's signature [19].

3. Empirical Cross-Gramian

In this section the empirical cross-gramian for MIMO systems is introduced. For general, possibly nonlinear, control systems of the form

\[
\dot{x}(t) = f(x(t), u(t)),
\]

\[
y(t) = g(x(t), u(t)),
\]

with states \(x(t) \in \mathbb{R}^n\), input or control \(u(t) \in \mathbb{R}^m\), outputs \(y(t) \in \mathbb{R}^o\), a vector field \(f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n\), and an output function \(g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^o\), the procedure from Section 2 is not viable. In [11, 12, 33] the concept of empirical (controllability and observability) gramians was introduced. This is a POD method based solely on state space simulations of the system [34]. These empirical gramians correspond to the classic gramians for linear systems as shown in [11]. Subsequently this approach and its field of application were advanced by [32, 35–37]. Because the empirical gramians can be aligned to the operating region of the underlying system in terms of initial states and input or control, the empirical gramians carry more detailed information on the system [38] than the gramians computed as solutions of matrix equations.

Empirical gramians are based on averaging the response of a system that is perturbed in inputs and initial states. Initially, the perturbed input was restricted to a delta impulse \(u(t) = \delta(t)\), which was broadened to more general input configurations in [13] under the name of empirical controllability covariance matrix and empirical observability covariance matrix.

The necessary perturbation sets are systematically defined next; these should reflect the operating range of the underlying system. \(E_u\) and \(E_x\) are sets of standard directions for the inputs and initial states. Sets \(R_u\) and \(R_x\) are orthogonal transformations (rotations) to these standard directions of inputs and initial states, respectively, while \(Q_u\) and \(Q_x\) hold scales to these directions:

\[
E_u = \left\{ e_i \in \mathbb{R}^j; \| e_i \| = 1; e_i e_{j \neq i} = 0; i = 1, \ldots, m \right\},
\]

\[
E_x = \left\{ f_i \in \mathbb{R}^n; \| f_i \| = 1; f_i f_{j \neq i} = 0; i = 1, \ldots, n \right\},
\]

\[
R_u = \left\{ S_i \in \mathbb{R}^{j \times j}; S_i^T S_i = 1; i = 1, \ldots, s \right\},
\]

\[
R_x = \left\{ T_i \in \mathbb{R}^{n \times n}; T_i^T T_i = 1; i = 1, \ldots, t \right\},
\]

\[
Q_u = \left\{ c_i \in \mathbb{R}; c_i > 0; i = 1, \ldots, q \right\},
\]

\[
Q_x = \left\{ d_i \in \mathbb{R}; d_i > 0; i = 1, \ldots, r \right\}.
\]

Along the lines of the empirical controllability gramian and empirical observability gramian [11], the empirical cross-gramian for SISO systems was introduced in [14]. In this
work, as a new contribution, the empirical cross-gramian is
generalized to square MIMO systems. Hence, the scope of the
cross-gramian is extended to nonlinear control systems
and provides an alternative nonlinear cross-gramian to [10]. For
a general (possibly nonlinear) MIMO system with \( \dim(u) = \dim(x) \) the empirical cross-gramian is defined as follows.

**Definition 2** (empirical cross-gramian). For sets \( E_u, E_x, R_u, R_x, Q_u, \) and \( Q_x \), input \( \overline{u} \) during steady state \( \overline{x} \) with output \( \overline{y} \), the empirical cross-gramian \( \overline{W}_X \) relating the states \( x^{hij} \) of inputs \( u^{hij}(t) = q_i S e_j \delta(t) + \overline{u} \) to output \( y^{kib} \)
for \( x_0^{kib} = d_k T f_b + \overline{x} \), is given by

\[
\overline{W}_X = \left[ \begin{array}{c|c|c} |Q_u| & |R_u| & m |Q_x| & |R_x| \\
\end{array} \right]^{1/2} \cdot \sum_{h=1}^{n-1} \sum_{i=1}^{n} \sum_{j=1}^{n-k} \sum_{k=1}^{k} \delta_k \int_0^\infty \psi^{hijkl}(t) dt T_i^T 
\]

\[
\psi^{hijkl}_{ab}(t) = f_a^T T_i^T \Delta x^{hij}(t) e_j^T \Delta y^{kib}(t) \in \mathbb{R},
\]

\[
\Delta x^{hij}(t) = \left( x^{hij}(t) - \overline{x} \right),
\]

\[
\Delta y^{kib}(t) = \left( y^{kib}(t) - \overline{y} \right).
\]

Essentially, the empirical cross-gramian is an averaged
cross-gramian over snapshots with the specified perturbations
in input and initial states around steady state input \( x(\overline{u}) \)
and steady state output \( y(X) \).

Next, similar to [11, 14], the equality of the cross-gramian
and the empirical cross-gramian for linear MIMO control
systems are shown next.

**Lemma 3** (empirical cross-gramian). For any nonempty sets
\( R_u, R_x, Q_u, \) and \( Q_x \) the empirical cross-gramian \( \overline{W}_X \) of an
asymptotically stable linear control system is equal to the cross-
gramian.

**Proof.** For an asymptotically stable linear control system, the
input-to-state and state-to-output maps are given by

\[
\Delta x(t) = x(t) = e^{At} Bu(t),
\]

\[
\Delta y(t) = y(t) = Ce^{At} x_0;
\]

thus,

\[
\psi^{hijkl}_{ab} = f_a^T T_i^T \left( e^{At} B_k e_j^T \right) e_j^T \Delta x^{hij} \Delta y^{kib} = q_i \delta_k \Delta x^{hij} \Delta y^{kib},
\]

\[
\Rightarrow \psi^{hijkl} = q_i \delta_k \Delta x^{hij} \Delta y^{kib},
\]

\[
\Rightarrow \overline{W}_X = \left[ \begin{array}{c|c|c} |Q_u| & |R_u| & m |Q_x| & |R_x| \\
\end{array} \right]^{1/2} \cdot \sum_{h=1}^{n-1} \sum_{i=1}^{n} \sum_{j=1}^{n-k} \sum_{k=1}^{k} \delta_k \int_0^\infty e^{At} B e^{At} dt
\]

\[
= \int_0^\infty e^{At} B e^{At} dt = W_X.
\]

As for the other empirical gramians, this proof is only
valid for impulse input, yet a similar approach to [13] can
be used to extend the empirical cross-gramian, yielding
an empirical cross-covariance matrix by allowing general
discrete input signals. The snapshots \( x^{hij} \) and \( y^{kib} \) can not
only be computed as simulations on demand but also be
included from observed experimental data. In case the data
is collected at discrete times \( t \) in regular intervals \( \Delta t \), following
[39], a discrete representation of the empirical cross-gramian
is given here too.

**Definition 4** (discrete empirical cross-gramian). For sets \( E_u, E_x, R_u, R_x, Q_u, \) and \( Q_x \), input \( \overline{u} \) during steady state \( \overline{x} \) with
output \( \overline{y} \), the discrete empirical cross-gramian \( \overline{W}_X \) relating the states \( x^{hij} \) of input \( u^{hij}(t) = q_i S e_j \delta(t) + \overline{u} \) to output \( y^{kib} \)
of \( x_0^{kib} = d_k T f_b + \overline{x} \), is given by

\[
\overline{W}_X = \left[ \begin{array}{c|c|c} |Q_u| & |R_u| & m |Q_x| & |R_x| \\
\end{array} \right]^{1/2} \cdot \sum_{h=1}^{n-1} \sum_{i=1}^{n} \sum_{j=1}^{n-k} \sum_{k=1}^{k} \delta_k \sum_{i=0}^{\infty} T_i^T \psi^{hijkl}_{ab} T_i,
\]

\[
\psi^{hijkl}_{ab}(t) = f_a^T T_i^T \Delta x^{hij}(t) e_j^T \Delta y^{kib}(t) \in \mathbb{R},
\]

\[
\Delta x^{hij}(t) = \left( x^{hij}(t) - \overline{x} \right),
\]

\[
\Delta y^{kib}(t) = \left( y^{kib}(t) - \overline{y} \right).
\]

Computational complexity depends largely on the number of
scales and rotations of perturbations as well as the order of
integration used to generate the snapshots \( x^{hij} \) and \( y^{kib} \).

The empirical cross-gramian enables state reduction for
square nonlinear control systems without an additional bal-
ancing procedure using direct truncation, where the approx-
imately balancing projection \( U \) is computed by an SVD and
truncated to \( U_1 \), analogous to (7):

\[
\dot{x}(t) = U_1^T f(U_1 x(t), u(t)),
\]

\[
y(t) = g(U_1 x(t), u(t)).
\]

Like POD, to quantify how close the subspace obtained
by reduction is approximating the state space, a measure of
total preserved energy [11, 40] can also be employed here:

\[
\overline{E} = \frac{\sum_{i=1}^{k} \sigma_i}{\sum_{i=1}^{n} \sigma_i}
\]

for \( k \) retained states of an \( n \)-dimensional model with \( n - k \)
truncated states related to the \( n - k \) lowest singular values of
the empirical cross-gramian \( \sigma(W_X) \).
4. Combined State and Parameter Reduction

Two methods for combined reduction, allowing simultaneous reduction of state and parameter spaces, are proposed: an observability-based and a controllability-based ansatz. For parametrized general control systems with parameters \( \theta \in \mathbb{R}^p \):

\[
\begin{align*}
\dot{x}(t) &= f(x(t), u(t), \theta), \\
y(t) &= g(x(t), u(t), \theta),
\end{align*}
\]

(17)

in [16], the identifiability gramian was introduced (in [41] a similar concept is presented), which extends the concept of observability from states to parameters. Augmenting the parameters is obtained through the observability of these parameter-states by the augmented observability gramian:

\[
\tilde{W}_O = \begin{pmatrix} W_O^T \\ W_M \\ W_P \end{pmatrix} \in \mathbb{R}^{(n+p)\times(n+p)},
\]

(19)

with the state observability gramian \( W_O \in \mathbb{R}^{n\times n} \), the parameter observability gramian \( W_P \in \mathbb{R}^{p\times p} \), and a mixture matrix \( W_M \in \mathbb{R}^{n\times p} \). From the observability gramian of this augmented system, the identifiability gramian \( W_I \in \mathbb{R}^{p\times p} \) can be extracted via the Schur complement:

\[
W_I = W_P - W_M^T W_O^{-1} W_M.
\]

(20)

For an approximation of the identifiability gramian \( W_I \), the parameter observability gramian \( W_P \approx W_I \) itself is often sufficient. The identifiability is then given as the observability of the parameters, through the singular values of \( W_I \) or approximately by \( W_P \), respectively. Instead of using the identifiability information for parameter identification as in [16, 42], a projection to the dominant parameter subspace is computed from \( W_I \). Similar to the cross-gramian approach, a singular value decomposition of the approximate identifiability gramian yields the reduced parameters \( \bar{\theta} \):

\[
W_I = \Pi \Sigma \Pi^T = \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} \Pi_1^T \\ \Pi_2^T \end{pmatrix},
\]

(21)

\[
\Rightarrow \bar{\theta} = \Pi_1^T \theta, \quad \Pi_1 \bar{\theta} = \theta.
\]

The partitioning depends on the singular values in \( \Delta \). A truncation of the projection \( \Pi \) results in the reduced parameters \( \bar{\theta} \) and the associated parameter reduced order model:

\[
\begin{align*}
\dot{x}(t) &= f(x(t), u(t), \Pi_1 \bar{\theta}), \\
\tilde{y}(t) &= g(x(t), u(t), \Pi_1 \bar{\theta}),
\end{align*}
\]

(22)

Next, a combined reduction of state and parameter space is introduced. With the identifiability gramian-based parameter reduction, first, the parameter space of the system is reduced. The observability of the states is encoded in the augmented observability gramian too; it can be extracted as the upper-left \( n \times n \) matrix from \( \tilde{W}_O \). Then, after computation of a controllability gramian \( W_C \), the state space is reduced by balanced truncation (balanced truncation provides two-sided truncated projection matrices \( U_1 \) and \( V_1 \); see [43]). This results in an observability-based combined state and parameter reduction:

\[
\begin{align*}
\dot{x}(t) &= V_1 f(U_1 x(t), u(t), \Pi_1 \bar{\theta}), \\
\tilde{y}(t) &= g(U_1 x(t), u(t), \Pi_1 \bar{\theta}).
\end{align*}
\]

(23)

Similarly, a controllability-based combined reduction can be achieved by a parameter reduction using the sensitivity gramian from [44] for additive partitionable systems:

\[
\begin{align*}
\dot{x}(t) &= \sum_{k=1}^{p} f(x(t), \theta_k), \\
\Rightarrow W_C &= W_{C0} + \sum_{k=1}^{p} W_{C,k},
\end{align*}
\]

(24)

\[
W_S = \begin{pmatrix} tr(W_{C1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & tr(W_{C,p}) \end{pmatrix} \in \mathbb{R}^{p \times p},
\]

which is based on [17, 45] and treats the parameters as additional inputs. By the sensitivity gramian, controllability information on the parameters is provided, which also allows a parameter reduction, again by a singular value decomposition of \( W_C \). Since an approximate controllability gramian \( W_C \) is also computed in process (24), after computation of an observability gramian \( W_O \), the parameter reduced system is reduced in states by balanced truncation. This results in a controllability-based combined state and parameter reduction. The controllability-based combined reduced order model has the same form as the observability-based reduced model (23).

5. Joint Gramian and Combined Reduction

In addition to controllability- and observability-based combined reduction, a cross-gramian-based combined reduction is proposed next, which, for symmetric control systems, is enabled by the empirical cross-gramian for MIMO systems from Section 3. Aggregating the computation of controllability and observability, not only for states like the cross-gramian but also for identifiability of parameters, leads to a reduction and identification method requiring a new single gramian. Here, the same augmented system (18) is used. The systems symmetry is not affected by the augmentation with the constant (parameter-) states, since in terms of a linear system the system components \( \{ A, B, C \} \) are expanded with zeros. This leads to the following new gramian matrix, which
utilizes the cross-gramian and thus unifies controllability and observability of states and parameters.

Definition 5 (joint gramian). The joint gramian $W_J$ is the cross-gramian of a square augmented system; see (18).

The joint gramian also has a $2 \times 2$ block structure:

$$W_J = \begin{pmatrix} W_X & W_M \\ W_M & W_P \end{pmatrix} \in \mathbb{R}^{(m,p) \times (m,p)},$$  \hspace{1cm} (25)

with the state cross-gramian $W_X \in \mathbb{R}^{n \times n}$, the parameter cross-gramian $W_P \in \mathbb{R}^{p \times p}$, and the mixture matrices $W_M \in \mathbb{R}^{n \times p}$, $W_M \in \mathbb{R}^{p \times n}$. The parameter-states are uncontrollable, yielding $W_P = 0$ and $W_m = 0$, since no inputs affect the augmented states. Thus a Schur complement of the joint gramian to extract the parameter associated lower right block matrix will always be zero:

$$W_I = W_P - W_M \bar{W}_X^{-1}W_M = 0 - 0 \ W_M^{-1}W_M = 0.$$  \hspace{1cm} (26)

Yet, the identifiability information on the parameters is encoded in the nonzero mixture matrix $W_M$. By taking the symmetric part of the joint gramian $\bar{W}_J = 0.5(W_J + W_J^T)$, one obtains the cross-identifiability gramian $W_I:  

$$W_I = 0 - \frac{1}{4} W_M W_M^{-1} W_M.$$  \hspace{1cm} (27)

Taking the inverse of the symmetric part of the cross-gramian is too costly in a large-scale setting. But the Schur complement can be approximated by using

$$D := \text{diag}(\bar{W}_X),$$  \hspace{1cm} (28)

as a coarse approximation (this approximation of the inverse is of complexity $n^2$) to the inverse from [46]. Thus, a more efficient cross-identifiability gramian is given by

$$W_I = -\frac{1}{4} W_M W_M^{-1} W_M.$$  \hspace{1cm} (29)

A reduced set of parameters $\bar{\theta}$ is again computed by a truncated projection obtained from the singular value decomposition of the cross-identifiability gramian:

$$W_I \approx W_P^{\text{SVD}} = \Pi \Lambda \Pi^T = \begin{pmatrix} \Pi_1 & \Pi_2 \end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} \Pi_1^T & \Pi_2^T \end{pmatrix},$$  \hspace{1cm} (30)

$$\Rightarrow \bar{\theta} = \Pi_1 \bar{\theta}, \quad \Pi_1 \bar{\theta} = \bar{\theta}.$$  

After a parameter reduction,

$$\dot{\bar{x}}(t) = f(\bar{x}(t),u(t),\Pi_1 \bar{\theta}),$$  \hspace{1cm} (31)

$$\dot{\bar{y}}(t) = g(\bar{x}(t),u(t),\Pi_1 \bar{\theta}),$$

the states can be reduced with a state reduction by direct truncation, employing the usual cross-gramian $W_X$, a byproduct of the joint gramian $W_J$, which is the upper-left $n \times n$ block matrix of $W_J$:

$$\dot{\bar{x}}(t) = V_1 f(\bar{U}_1 \bar{x}(t),\bar{u}(t),\Pi_1 \bar{\theta}),$$

$$\dot{\bar{y}}(t) = V_1 g(\bar{U}_1 \bar{x}(t),\bar{u}(t),\Pi_1 \bar{\theta}).$$  \hspace{1cm} (32)

For this combined reduction of states and parameters no further gramians have to be computed and no balancing transformation is required.

6. Implementation and Numerical Results

For an efficient implementation, the structure of the gramian computation is exploited. First, the empirical gramians allow extensive parallelization. Each combination of direction, transformation, and scale can be processed separately yielding a sub-gramian. Second, the assembly of each sub-gramian can be comprehensively vectorized, since it consists of vector additions, inner products, and outer products. In the special case of the empirical cross-gramian and, thus, the empirical joint gramian which is an encapsulation of the empirical cross-gramian, organizing the observability snapshots into a third-order-tensor and exploiting generalized transpositions result in a very efficient gramian assembly (see emgr.m). The final resulting gramian is the normalized accumulation over all sub-gramians. For further details about the implementation see [44].

All gramians for the numerical results are computed by the empirical gramian framework introduced in [44]. The empirical gramian framework emgr [47] can be found at http://gramian.de and is compatible with Octave [48] and Matlab [49]. The source code, used for the following experiments, is provided as supplementary material.

The error measure employed in the following experiments is the relative $L^2$-error for a vector valued time series $y$ and $\bar{y}$ is given by

$$e = \frac{\|y - \bar{y}\|_2}{\|y\|_2},$$  \hspace{1cm} (33)

with the $L^2$-norm of a (discrete) time series

$$\|y\|_2 = \sqrt{\sum_t \|y(t)\|_2^2}. $$  \hspace{1cm} (34)

Numerical results for three models are presented next. First, state reduction is applied to a nonlinear benchmark problem to validate the applicability of the empirical cross-gramian. Then, a linear and nonlinear parametrized control system is considered for the state, parameter, and combined reduction.

6.1. Nonlinear Benchmark. Introduced in [51], this nonlinear benchmark (this benchmark is also listed in the MORwiki
[52]) has been used in [37, 40, 53] as a test problem for the assessment of the empirical controllability gramian and empirical observability gramian in a balanced truncation model order reduction setting. This benchmark system models a circuit consisting of capacitors and nonlinear resistors (a resistor parallel-connected to a diode). Its mathematical model is given by the following nonlinear SISO control system:

\[
\dot{x}(t) = \begin{pmatrix}
-g(x_1(t)) - g(x_1(t) - x_2(t)) \\
g(x_1(t) - x_2(t)) - g(x_2(t) - x_3(t)) \\
\vdots \\
g(x_{k-1}(t) - x_k(t)) - g(x_k(t) - x_{x+1}(t)) \\
\vdots \\
g(x_{N-1}(t) - x_N(t)) \\
\end{pmatrix} + \begin{pmatrix}
u(t) \\
0 \\
\vdots \\
0 \\
0
\end{pmatrix},
\]

\[y(t) = x_1(t),\]

with the nonlinear function \(g: \mathbb{R} \to \mathbb{R}:
\]

\[g(x) = \exp(x) + x - 1.\]  

In this setting with \(\text{dim}(x) = 100\), a zero initial state \(x_0 = 0\) and a decaying exponential input \(u(t) = e^{-t}\) are employed. Figure 1 shows the relative \(L^2\)-error in the reduced order models outputs reduced by balanced truncation of empirical controllability and observability gramian and direct truncation of the empirical cross-gramian.

After a steep initial drop in the error, the relative output error remains near the machine precision. Both methods, balanced truncation and direct truncation, perform very similarly and require less than 10\% of the full order model's state dimension to reach the error plateau. The empirical cross-gramian-based direct truncation exhibits a slightly lower output error.

6.2. State Reduction. Next, a parametrized linear control system and a parametrized nonlinear control system are reduced in states and parameters and combined in states and parameters. The parametrized linear control system model [17] is given by

\[
\dot{x} = Ax + Bu + \theta,
\]

\[y = Cx,\]  

with an additional parametrized source term \(\theta \in \mathbb{R}^N\).

The system is chosen to be state space symmetric: \(A = A^T\) and \(C = B^T\). A state space symmetric system provides that all system gramians, the controllability gramian, the observability gramian, and the cross-gramian are equal [54]. This allows the verification of the empirical cross-gramian.

The parametrized nonlinear control system model is given by

\[
\dot{x} = A \tanh\left(\frac{1}{4}x\right) + Bu + \theta,
\]

\[y = Cx,\]  

with a hyperbolic tangent nonlinearity using the same system matrices \(\{A, B, C\}\) as in (37). This model is related to the hyperbolic network model from [55].

The linear and nonlinear model are subject to zero initial states \(x_0 = 0\) and impulse input \(u(t) = \delta(t)\). For the following experiments a state space dimension \(n = \text{dim}(x(t)) = 256\); thus \(p = \text{dim}(\theta) = 256\) is assumed as well as an input and output dimension of \(m = \text{dim}(u(t)) = o = \text{dim}(y(t)) = 16\). The parameters are drawn from a uniform distribution \(\theta_i = U(0, 0.1)\); the system matrix \(A\) is generated as a sparse uniformly random matrix with ensured stability, and the input matrix \(B\) is a dense uniformly random matrix, yielding the output matrix \(C = B^T\).

The linear model in (37) and the nonlinear model in (38) are first reduced in states using the following methods: balanced truncation utilizing the empirical controllability gramian and empirical observability gramian; direct truncation of the empirical cross-gramian presented in Section 3. Furthermore, a method closely related to balanced POD [56,
Figure 2: Relative $L^2$ output error in the reduced order linear and nonlinear models after the state reduction using balanced POD, balanced truncation, and direct truncation of the empirical cross-gramian.

57] is tested too. For the linear model an approximate cross-gramian $W_Y$ is computed using the approach from [22, 58], which obtains the cross-gramian by computing the controllability gramian of the system augmented with its adjoint system (this approach is also implemented in the empirical gramian framework as empirical linear cross gramian). For the nonlinear model an approximate cross-gramian is computed following [56] by the product $W_Y = W CW_O$. In Figure 2 the relative $L^2$-error in the outputs $y$ is plotted for reduced orders $\dim(\tilde{x}) = 1, \ldots, n - 1$.

The state reduced models in the linear setting, generated by balanced POD, balanced truncation, and direct truncation, are of similar quality; yet balanced POD and direct truncation perform slightly better than balanced truncation. In the nonlinear setting balanced truncation outperforms balanced POD and direct truncation for which the output error flattens above a reduced order of about 40% of the original model order.

The better performance of balanced truncation for the nonlinear model is due to use of two-sided projections as opposed to the one-sided projections used for balanced POD and direct truncation here.

6.3. Parameter Reduction. The linear model in (37) and the nonlinear model in (38) are reduced in parameters using the empirical sensitivity gramian, the empirical identifiability gramian, and empirical cross identifiability gramian from the empirical joint gramian. Figure 3 depicts the relative $L^2$-error in the outputs $y$ for the reduced orders $\dim(\tilde{\theta}) = 1, \ldots, n - 1$.

For the linear and the nonlinear model, the controllability-based sensitivity gramian performs worst with the slowest decline in output error. The observability-based identifiability gramian and the cross-gramian-based cross-identifiability gramian show a sharper descent of the output error of similar quality in the linear and nonlinear setting, yet the observability-based parameter reduction exhibits a steeper drop of the output error for reduced orders up to 40% of the original parameter space.

Since the sensitivity gramian is a diagonal matrix, the associated projections reorder the parameters and thus exclude all effects of the truncated parameters, while the identifiability and cross-identifiability gramian use linear combinations of parameters to be truncated.

6.4. Combined Reduction. The linear model in (37) and the nonlinear model in (38) are next reduced in states and parameters employing the methods presented in Section 4 and Section 5. Controllability-based combined reduction uses the empirical sensitivity gramian for parameter reduction and balanced truncation for the state reduction. Observability-based combined reduction uses the empirical identifiability gramian for parameter reduction and also balanced truncation for the state reduction. The cross-gramian-based combined reduction utilizes the empirical joint gramian; the cross-identifiability gramian is used for the parameter reduction and the direct truncation of the cross-gramian is used for the state reduction. In Figures 4 and 5 the relative $L^2$-error in the outputs $y$ is plotted for reduced orders $\dim(\tilde{x}) = 1, \ldots, n - 1$ and $\dim(\tilde{\theta}) = 1, \ldots, n - 1$. 
For a better comparison, a cross-section of the surfaces in Figures 4 and 5 along the diagonals are plotted in Figure 6 showing the reduced order model’s output error for the same reduced order in states and parameters: $\dim(\tilde{x}) = \dim(\tilde{\theta})$.

For all reduced order models obtained by combined state and parameter reduction, the parameter reduction error dominates the output error. As for the parameter reduction, the controllability-based combined reduction by sensitivity gramian and balanced truncation performs worst with a slow descent in the output error. In the linear setting the cross-gramian-based joint gramian performs significantly better than the combination of identifiability gramian and balanced truncation. In the nonlinear setting the reduced order models of these methods exhibit similar behavior, which is due to...
the higher state reduction error (Figure 2) in the cross-gramian-based approach.

To assess the efficiency of the presented methods, the offline times (the time required to assemble the necessary empirical gramian matrices) are compared in Figure 7.

The state reduction of the linear model is accomplished fastest by the balanced-POD-related method, since the computational effort required for computation is equivalent to a single empirical controllability gramian. For the nonlinear model this advantage does not exist, since no adjoint system
for the nonlinear model is provided. Notably, the cross-gramian, for the direct truncation, is computed slightly faster than controllability, observability gramian, and balancing transformation for the balanced truncation.

Among the empirical gramians for parameter reduction the sensitivity gramian is computed fastest, yet due to the high error this is the least applicable. Between the identifiability and cross-identifiability gramian, which exhibit a comparable output error, the cross-gramian-based approach is considerably faster.

For the combined reduction the offline times are similar to the offline times of the parameter reduction, with the exception of the controllability-based combined reduction, which now takes longer than the cross-gramian-based joint gramian because of the additional observability gramian required for the balanced truncation state reduction. Thus the empirical joint gramian is the fastest of the tested methods for combined reduction and provides a competitive output error.

Further research has to be conducted on two-sided projections and error bounds for the reduced order models and on applying the empirical cross-gramian to nonlinear or nonsymmetric systems. Existing extensions for nonsymmetric systems are generalizing the symmetry constraint to orthogonal symmetry [59] or embedding into a symmetric system [24].

The empirical cross-gramian for MIMO systems completes the set of empirical gramians for state reduction, while the joint gramian completes the body of parameter identification gramians and enables combined reduction without balancing.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

This work was supported by the Deutsche Forschungsgemeinschaft, the Open Access Publication Fund of the University of Münster, DFG EXC 1003 Cells in Motion Cluster of Excellence, Münster, Germany, as well as the Center for Developing Mathematics in Interaction, DEMAIN, Münster, Germany.

**References**


