Advanced Model-Order Reduction Techniques for Large-Scale Dynamical Systems

by

Seyed-Behzad Nouri, B.Sc., M.A.Sc.

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Electronics
Carleton University
Ottawa, Ontario, Canada

© 2014

Seyed-Behzad Nouri

Abstract

Model Order Reduction (MOR) has proven to be a powerful and necessary tool for various applications such as circuit simulation. In the context of MOR, there are some unaddressed issues that prevent its efficient application, such as “reduction of multiport networks” and “optimal order estimation” for both linear and nonlinear circuits. This thesis presents the solutions for these obstacles to ensure successful model reduction of large-scale linear and nonlinear systems.

This thesis proposes a novel algorithm for creating efficient reduced-order macromodels from multiport linear systems (e.g. massively coupled interconnect structures). The new algorithm addresses the difficulties associated with the reduction of networks with large numbers of input/output terminals, that often result in large and dense reduced-order models. The application of the proposed reduction algorithm leads to reduced-order models that are sparse and block-diagonal in nature. It does not assume any correlation between the responses at ports; and thereby overcomes the accuracy degradation that is normally associated with the existing (Singular Value Decomposition based) terminal reduction techniques.

Estimating an optimal order for the reduced linear models is of crucial importance to ensure accurate and efficient transient behavior. Order determination is known to be a challenging task and is often based on heuristics. Guided by geometrical considerations, a novel and efficient algorithm is presented to determine the minimum sufficient order that ensures the

accuracy and efficiency of the reduced linear models.

The optimum order estimation for nonlinear MOR is extremely important. This is mainly due to the fact that, the nonlinear functions in circuit equations should be computed in the original size within the iterations of the transient analysis. As a result, ensuring both accuracy and efficiency becomes a cumbersome task. In response to this reality, an efficient algorithm for nonlinear order determination is presented. This is achieved by adopting the geometrical approach to nonlinear systems, to ensure the accuracy and efficiency in transient analysis.

Both linear and nonlinear optimal order estimation methods are not dependent on any specific order reduction algorithm and can work in conjunction with any intended reduced modeling technique.

Dedicated:

To the living memories of my father, who lived by example to inspire and motivate his students and children. I also dedicate this to my mother, my understanding wife, and my wonderful sons Ali and Ryan for their endless love, support, and encouragements.

Acknowledgments

First and foremost, I would sincerely like to express my gratitude to my supervisor, Professor Michel Nakhla. Without his guidance, this thesis would have been impossible. I appreciate his insight into numerous aspects of numerical simulation and circuit theory, as well as his enthusiasm, wisdom, care and attention. I have learned from him many aspects of science and life. Working with him was truly an invaluable experience.

I am also sincerely grateful to my co-supervisor, Professor Ram Achar, for his helpful suggestions and guidance, which was crucial in many stages of the research for this thesis. Most of all I wish to thank him for his motivation and encouragements.

I would like to thank my current and past fellow colleagues in our Computer-Aided Design group for keeping a spirit of collaboration and mutual respect. They were always readily available for some friendly deliberations that made my graduate life enjoyable. I will always fondly remember their support and friendship.

I am thankful towards the staff of the Department of Electronics at Carleton University for having been so helpful, supportive, and resourceful.

Last but not least, I give special thanks to my family for all their unconditional love, encouragement, and support. I am eternally indebted to my wife and both my sons for their unconditional, invaluable and relentless support, encouragement, patience and respect. I would like to thank Mrs Zandi for all her understandings and gracious friendship with my

family. My final thoughts are with my parents to whom I am forever grateful. I cherish the memories of my late father with great respect. Words cannot express my admiration for the endless kindness, dedication, and sacrifices that my parents have made for their children. I believe that I could not have achieved this without their unlimited sacrifice. This is for them.

Thank you all sincerely,

Table of Contents

Abstract	ii
Acknowledgments	v
Table of Contents	vii
List of Tables	xiii
List of Figures	xiv
List of Acronyms	xx
List of Symbols	xxii
Introduction	1
1 Background and Preliminaries	6
1.1 Dynamical Systems	6
1.2 Linear Systems	7
1.2.1 Important Property of Linear Systems	9
1.2.2 Mathematical Modeling of Linear Systems	10

1.3	Nonlinear Systems	13
1.3.1	Solutions of Nonlinear Systems	15
1.3.2	Linear versus Nonlinear	16
1.4	Mathematical Modeling of Electrical Networks	16
1.5	Overview of Formulation of Circuit Dynamics	18
1.5.1	MNA Formulation of Linear Circuits	19
1.5.2	MNA Formulation of Nonlinear Circuits	20
2	Model Order Reduction - Basic Concepts	25
2.1	Motivation	26
2.2	The General Idea of Model Order Reduction	26
2.3	Model Accuracy Measures	28
2.3.1	Error in Frequency Domain	31
2.4	Model Complexity Measures	32
2.5	Main Requirements for Model Reduction Algorithms	33
2.6	Essential Characteristic of Physical Systems	34
2.6.1	Stability of Dynamical Systems	34
2.6.2	Internal Stability	35
2.6.3	External Stability	38
2.6.4	Passivity of a Dynamical Model	38
2.7	The Need for MOR for Electrical Circuits	39
3	Model Order Reduction for Linear Dynamical Systems	40

3.1	Physical Properties of Linear Dynamical Systems	41
3.1.1	Stability of Linear Systems	41
3.1.2	Passivity of Linear Systems	46
3.2	Linear Order Reduction Algorithms	49
3.3	Polynomial Approximations of Transfer Functions	50
3.3.1	AWE Based on Explicit Moment Matching	52
3.4	Projection-Based Methods	53
3.4.1	General Krylov-Subspace Methods	56
3.4.2	Truncated Balance Realization (TBR)	58
3.4.3	Proper Orthogonal Decomposition (POD) Methods	64
3.5	Non-Projection Based MOR Methods	67
3.5.1	Hankel Optimal Model Reduction	67
3.5.2	Singular Perturbation	67
3.5.3	Transfer Function Fitting Method	68
3.6	Other Alternative Methods	76
4	Model Order Reduction for Nonlinear Dynamical Systems	77
4.1	Physical Properties of Nonlinear Dynamical Systems	78
4.1.1	Lipschitz Continuity	79
4.1.2	Existence and Uniqueness of Solutions	80
4.1.3	Stability of Nonlinear Systems	81
4.2	Nonlinear Order Reduction Algorithms	84
4.2.1	Projection framework for Nonlinear MOR - Challenges	84

4.2.2	Nonlinear Reduction Based on Taylor Series	86
4.2.3	Piecewise Trajectory based Model Order Reduction	91
4.2.4	Proper Orthogonal Decomposition (POD) Methods	95
4.2.5	Empirical Balanced Truncation	98
4.2.6	Summary	100

5 Reduced Macromodels of Massively Coupled Interconnect Structures via

Clustering		101
5.1	Introduction	101
5.2	Background and Preliminaries	104
5.2.1	Formulation of Circuit Equations	105
5.2.2	Model-Order Reduction via Projection	106
5.3	Development of the Proposed Algorithm	107
5.3.1	Formulation of Submodels Based on Clustering	108
5.3.2	Formulation of the Reduced Model Based on Submodels	110
5.4	Properties of the Proposed Algorithm	114
5.4.1	Preservation of Moments	114
5.4.2	Stability	115
5.4.3	Passivity	116
5.4.4	Guideline for Clustering to Improve Passivity	123
5.5	Numerical Examples	125
5.5.1	Example I	126
5.5.2	Example II	130

6	Optimum Order Estimation of Reduced Linear Macromodels	136
6.1	Introduction	136
6.2	Development of the Proposed Algorithm	137
6.2.1	Preliminaries	137
6.2.2	Geometrical Framework for the Projection	140
6.2.3	Neighborhood Preserving Property	142
6.2.4	Unfolding the Projected Trajectory	148
6.3	Computational Steps of the Proposed Algorithm	150
6.4	Numerical Examples	153
6.4.1	Example I	153
6.4.2	Example II	156
7	Optimum Order Determination for Reduced Nonlinear Macromodels	162
7.1	Introduction	162
7.2	Background	163
7.2.1	Formulation of Nonlinear Circuit Equations	163
7.2.2	Model Order Reduction of Nonlinear Systems	164
7.2.3	Projection Framework	164
7.3	Order Estimation for Nonlinear Circuit Reduction	166
7.3.1	Differential Geometric Concept of Nonlinear Circuits	166
7.3.2	Nearest Neighbors	172
7.3.3	Geometrical Framework for the Projection	173
7.3.4	Proposed Order Estimation for Nonlinear Reduced Models	175

7.4	Computational Steps of the Proposed Algorithm	180
7.5	Numerical Examples	185
7.5.1	Example I	185
7.5.2	Example II	188
8	Conclusions and Future Work	196
8.1	Conclusions	196
8.2	Future Research	198
	List of References	200
	Appendix A Properties of Nonlinear Systems in Compare to Linear	226
	Appendix B Model Order Reduction Related Concepts	228
B.1	Tools From Linear Algebra and Functional Analysis	228
B.1.1	Review of Vector Space and Normed Space	228
B.1.2	Review of the Different Norms	231
B.2	Mappings Concepts	232
	Appendix C Proof of Theorem-5.1 in Section 5.4	238
	Appendix D Proof of Theorem-5.2 in Section 5.4	244

List of Tables

1.1	Summary: general properties of linear and nonlinear systems	17
2.1	Measuring reduction accuracy in time domain	30
3.1	Time complexities of standard TBR.	61
4.1	Comparison of properties of the available nonlinear model order reduction algorithm	100
5.1	CPU-cost comparison between original system, PRIMA and proposed method.	129

List of Figures

1.1	Illustration of linear physical system $\mathcal{L}$.	8
1.2	Illustration of a subcircuit that accepting p -inputs and interacting with other module through its q -outputs.	21
2.1	Model order reduction.	29
2.2	Measuring error of approximation.	29
3.1	Illustrates the <i>uniform stability</i> ; uniformity implies the σ -bound is independent of t_0 .	43
3.2	A decaying-exponential bound independent of t_0 .	44
4.1	Illustration of Lipschitz property.	80
4.2	Model reduction methods for nonlinear dynamical systems categorized into four classes.	84
4.3	Illustration of the state space of a planar system, where $\mathbf{x}_i$ are the expansion points on the <i>training trajectory</i> A. Because solutions B and C are in the vicinity ball of the expansion states, they can be efficiently simulated using a TPWL model, however this can not be true for the solutions D and E.	92
4.4	Nonlinear Balanced model reduction.	99

5.1	Reduced-modeling of multiport linear networks representing N -conductor TL.	108
5.2	Illustration of forming clusters of active and victim lines in a multiconductor transmission line system.	109
5.3	Linear (RLC) subcircuit π accompanied with the reduced model $\hat{\Psi}$	112
5.4	The overall network comprising the reduced model, embedded RLC subcircuit, and nonlinear termination.	113
5.5	Illustration of strongly coupled lines bundled together as active lines in the clusters.	118
5.6	The frequency-spectrum of the minimum eigenvalue of $\Phi(s)$ containing 32 clusters.	124
5.7	The enlarged region near the x-axis of Fig. 5.6 (illustrating eigenvalues extending to the negative region, indicating passivity violation).	125
5.8	Spectrum of $\Phi(s)$ versus frequency with proper clustering to improve passivity (no passivity violations observed).	126
5.9	The frequency-spectrum of the minimum eigenvalue of $\Phi(s)$ with clustering to improve passivity behavior (no passivity violations observed).	127
5.10	32 conductor coupled transmission line network with terminations considered in the example.	128
5.11	Sparsity pattern of reduced MNA equations using conventional PRIMA (dense).	129
5.12	Sparsity pattern of reduced MNA equations using the proposed method.	129
5.13	Transient responses at victim line near-end of line#2.	130
5.14	Transient responses at victim line near-end of line#12.	131

5.15	Transient responses at victim line far-end of line#31.	132
5.16	Cross sectional geometry (Example 2).	132
5.17	Interconnect structure with nine clusters (Example 2).	133
5.18	Minimum eigenvalue of $\Phi(s)$ while using 9 clusters (each cluster with nine lines while one of them acting as an active line).	133
5.19	Negative eigenvalue of $\Phi(s)$ (using the 9-cluster approach).	134
5.20	Illustration of the interconnect structure grouped as three clusters (each cluster with nine lines while the three of the strongly coupled lines in each of them acting as active lines [shown in red color]).	134
5.21	Eigenvalue of $\Phi(s)$ (using 3 clusters based on the proposed flexible clustering approach).	135
5.22	Minimum eigenvalues of $\Phi(s)$ (using 3 clusters based on the proposed flexible clustering approach).	135
6.1	Any state corresponding to a certain time instant can be represented by a point (e.g. A, N, E and F) on the trajectory curve (T) in the variable space. .	139
6.2	Illustration of a multidimensional adjacency ball centered at $\mathbf{x}(t_i)$, accommodating its four nearest neighboring points.	141
6.3	Illustration of false nearest neighbor (FNN), where $\hat{T}$ is the projection of T in Fig. 1.	142
6.4	Illustration of the neighborhood structure of the state $\mathbf{x}_i$ and its projection $\mathbf{z}_i$ in the state space and reduced space, respectively.	143
6.5	Displacement between two false nearest neighbors in the unfolding process.	149
6.6	(a) A lossy transmission line as a 2-port network with the terminations; (b) Modeled by 1500 lumped RLGC π -sections in cascade.	154

6.7	The percentage of the false nearest neighbors on the projected trajectory. . .	155
6.8	Transient response of the current entering to the far-end of the line when the reduced model is of order $m = 66$	156
6.9	Transient response of the current at the far-end terminal of the line when the reduced model is of order $m = 66$	157
6.10	Accuracy comparison in PRIMA models with different orders.	158
6.11	A RLC mesh as a 24-port subcircuit with the terminations.	158
6.12	The percentage of the false nearest neighbors among 1000 data points on the projected trajectory.	159
6.13	Transient responses at near-end of horizontal trace#1.	160
6.14	Transient responses at near-end of horizontal trace#10.	160
6.15	Errors from using the reduced models with different orders in the frequency domain.	161
7.1	Chua's circuit.	167
7.2	Trajectory of the Chua's circuit in the state-space (scaled time: $0 \leq t \leq$ 100) for a given initial condition.	167
7.3	The time-series plot of the system variables $(x_i(t))$ as coordinates of state space.	168
7.4	(a) Digital inverter circuit; (b) The circuit model to characterize the dy- namic behavior of digital inverter at its ports.	169
7.5	A geometric structure $\mathcal{M}$ attracting the trajectories of the circuit in Fig.7.4. . .	169
7.6	(a) The Möbus strip and (b) Torus are visualizations of 2D manifolds in $\mathbb{R}^3$. .	170

7.7	Illustration of a multidimensional adjacency ball centered at $\mathbf{x}(t_i)$ ($\mathbf{X}$), accommodating its two nearest neighboring points ($\blacktriangledown$) on the trajectory of the Chua's circuit (for $0 \leq t \leq 2$).	172
7.8	Illustration of Chua's trajectory in Fig.7.7 projected to a two-dimensional subspace, where its underlying manifold is over-contracted.	174
7.9	(left) Illustration of false nearest neighbor (FNN), where the 3-dimensional trajectory of the Chua's circuit in Fig.7.7 is projected; (right) A zoomed-in view of the projected trajectory.	174
7.10	Drastic displacement between two false nearest neighbors in the unfolding process.	176
7.11	Small displacement between every two nearest neighbors by adding a new dimension ($m + 1$ or higher), when trajectory was fully unfolded in m dimensional space.	177
7.12	Flowchart of the proposed nonlinear order estimation strategy. The gray blocks are the steps of nonlinear MOR interacting with the proposed methods.	182
7.13	(a) Diode chain circuit, (b) Excitation waveform at input.	186
7.14	The percentage of the false nearest neighbors on the projected nonlinear trajectory.	187
7.15	Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).	188
7.16	Excitation test waveform at input and comparison of the responses at nodes 3, 5 and 7, respectively.	189

7.17	(a) Nonlinear transmission line circuit model, (b) Excitation waveform at input.	190
7.18	The percentage of the false nearest neighbors on the projected nonlinear trajectory.	191
7.19	Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).	192
7.20	(a) Excitation test waveform at input, (b) Comparison of the responses at nodes 5, 50, 70, and 200, respectively.	193
7.21	Excitation waveform at input.	193
7.22	The percentage of the false nearest neighbors on the projected nonlinear trajectory.	194
7.23	Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).	194
7.24	Comparison of the responses at output nodes for the segments 30, 60 and 70 respectively.	195
B.1	Visualization of a mapping	233
B.2	Visualization of an injective mapping	234
B.3	Visualization of an surjective mapping	234
B.4	Inverse mapping $T^{-1}: \mathbf{Y} \longrightarrow \mathcal{D}(T) \subseteq \mathbf{X}$ of a bijective mapping T	235

List of Acronyms

Acronyms	Definition
ADE	Algebraic Differential Equation
AWE	Asymptotic Waveform Evaluation
BIBO	Bounded-In Bounded-Out
CAD	Computer Aided Design
CPU	Central Processing Unit
DAE	Differential-Algebraic Equation
EIG	Eigenvalue (diagonal) Decomposition
FD	Frequency Domain
FNN	False Nearest Neighbor
HSV	Hankel Singular Value
IC	Integrated Circuit
I/O	Input-Output
KCL	Kirchoff's Current Law
KVL	Kirchoff's Voltage Law
LHP	Left Half (of the complex) Plane
LHS	Left Hand Side
LTI	Linear Time Invariant (Dynamical System)
MEMS	Micro-Electro-Mechanical System
MIMO	Multi Input and Multi Output (multiport) system
MOR	Model Order Reduction

NN	Nearest Neighboring point
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
POD	Proper Orthogonal Decomposition
PRIMA	Passive Reduced-order Interconnect Macromodeling Algorithm
PVL	Padé Via Lanczos
RHP	Right Half (of the complex) Plane
RHS	Right Hand Side
RMS	Root Mean Square
SISO	Single Input and Single Output system
SVD	Singular Value Decomposition
TD	Time Domain
TF	Transfer Function
TBR	Truncated Balanced Realization
TPWL	Trajectory Piecewise Linear
VLSI	Very Large Scale Integrated circuit

List of Symbols

Symbols	Definition
$\mathbb{N}$	The field of natural numbers
$\mathbb{R}$	The field of real numbers
$\mathbb{R}_+$	The set of all positive real numbers
$\mathbb{C}$	The field of complex numbers, <i>e.g.</i> : s -plane
$\mathbb{R}^n$	The set of real column vectors of size n , $\mathbb{R}^{n \times 1}$, <i>i.e.</i> n -dimensional Euclidean space
$\mathbb{C}^n$	The set of complex column vectors of size n , $\mathbb{C}^{n \times 1}$, <i>i.e.</i> n -dimensional Euclidean space
$\mathbb{R}^{n \times m}$	The set of real matrices of size $n \times m$
$\mathbb{C}^{n \times m}$	The set of complex matrices of size $n \times m$
$\mathbb{C}_+$	The open right half plane in the complex plane; $\mathbb{C}_+ = \{s \in \mathbb{C} : \Re(s) > 0\}$
$\mathbb{C}_-$	The open left half plane in the complex plane; $\mathbb{C}_- = \{s \in \mathbb{C} : \Re(s) < 0\}$
$\underline{\mathbb{C}}_+$	The closed right half plane in the complex plane; $\underline{\mathbb{C}}_+ = \{s \in \mathbb{C} : \Re(s) \geq 0\}$
$\underline{\mathbb{C}}_-$	The closed left half plane in the complex plane; $\underline{\mathbb{C}}_- = \{s \in \mathbb{C} : \Re(s) \leq 0\}$
$\Re$ or $\Re e$	Real part of a complex number
$\Im$ or $\Im m$	Imaginary part of a complex number
$\mathcal{C}^n$	n differentiable (n -smooth)
$\mathcal{C}^\infty$	Infinitely differentiable (smooth)
$\bar{a}$ or a^*	The complex conjugate of a complex number $a \in \mathbb{C}$

$\mathbf{A}_{m \times n}$	An $m \times n$ matrix $\mathbf{A} = [a_{ij}]$, where a_{ij} is an element in i -th row and j -th column
$\mathbf{A}^T$	The transpose of matrix $\mathbf{A} = [a_{ij}]$, defined as $\mathbf{A}^T = [a_{ji}]$
$\overline{\mathbf{A}}$ or $\mathbf{A}^*$	Complex-conjugate of each entries in complex matrix $\mathbf{A} = [a_{ij}]$, defined as: $\mathbf{A}^* = \overline{\mathbf{A}} = [\bar{a}_{ij}]$
$\mathbf{A}^H$	Complex-conjugate transpose of complex matrix $\mathbf{A} = [a_{ij}]$, defined as: $\mathbf{A}^H = \overline{\mathbf{A}}^T = [\bar{a}_{ji}]$
$\mathbf{I}_n$	An $n \times n$ identity matrix $\mathbf{I} = [\iota_{ij}]$, where $\iota_{ij} = 1$, for $i = j$ and $\iota_{ij} = 0$, for $i \neq j$
$\mathbf{A}_{n \times n}^{-1}$	The inverse of the square matrix $\mathbf{A}$ such that $\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}_n$
$\emptyset$	Empty set / empty subspace
$\det(\mathbf{A})$	Determinant of matrix $\mathbf{A}$
$\text{rank}(\mathbf{A})$	Rank of matrix $\mathbf{A}$
$\dim(\mathbf{A})$	Dimension of an square matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, e.g. $\dim(\mathbf{A}) = n$
$\mathbf{A} > 0$	$\mathbf{A}$ is a positive definite matrix
$\mathbf{A} \geq 0$	$\mathbf{A}$ is a semi-positive definite matrix
$\text{colsp}(\mathbf{A})$	Column span (also called <i>range</i>) of matrix $\mathbf{A}$
$\lambda(\mathbf{A})$	Set of eigenvalues (<i>spectrum</i>) of square matrix $\mathbf{A}$
$\lambda_i(\mathbf{A})$	i -th eigenvalue of matrix $\mathbf{A}$
$\lambda_{\max}(\mathbf{A})$	Maximum eigenvalue of matrix $\mathbf{A}$, the largest eigenvalue in the spectrum of $\mathbf{A}$
$\lambda_{\min}(\mathbf{A})$	Minimum eigenvalue of matrix $\mathbf{A}$, the smallest eigenvalue in the set
$\sigma(\mathbf{A})$	Set of singular values of matrix $\mathbf{A}$
$\sigma_i(\mathbf{A})$	i -th singular value of matrix $\mathbf{A}$
$\sigma_{\max}(\mathbf{A})$	Maximum singular values of matrix $\mathbf{A}$, i.e. $= \sigma_1$
$\sigma_{\min}(\mathbf{A})$	Minimum singular values of matrix $\mathbf{A}$, i.e. $= \sigma_n$
$\lambda(\mathbf{E}, \mathbf{A})$	Set of all finite eigenvalues of the regular matrix pencil $(\mathbf{E}, \mathbf{A})$
$\left\lfloor \frac{q}{p} \right\rfloor$	$m = \max \left(\frac{q}{p} \right) \mid m \in \mathbb{N}$
$\text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$	Vector space spanned by the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$
$\text{diag}(d_1, d_2, \dots, d_n)$	Diagonal matrix with $d_1, d_2, \dots, d_n$ on its diagonal
$\text{blkdiag}\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$	Block diagonal matrix with the blocks $\mathbf{A}_1, \dots, \mathbf{A}_n$ on its diagonal

$\deg()$ Degree of polynomials with real/complex coefficients

$\sup \{ \}$ Supremum of a set

$\|\mathbf{x}\|$ Euclidean vector norm $\mathbf{x} \in \mathbb{C}^n$, $\|\mathbf{x}\| = \left(\sum_i x_i^2 \right)^{\frac{1}{2}}$

$\|\mathbf{A}\|$ The consistent matrix norm subordinate to Euclidean vector norm, i.e.

$$\max_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sigma_{max}(\mathbf{A})$$

$\|\mathbf{A}\|_F$ Frobenius norm of matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$, i.e. $\left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \left(\sum_{i=1}^n \sigma_i \right)^{1/2}$,

given $n \leq m$

$\|\mathbf{A}\|_1$ Maximum of the sum of column vectors in matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$, i.e. $\max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$

$\|\mathbf{A}\|_\infty$ Maximum of the sum of row vectors in matrix $\mathbf{A}$, i.e. $\max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$

s Complex frequency (Laplace variable), $s = \alpha + j\omega$, $\alpha, \omega \in \mathbb{R}$

$\forall$ For all

$\exists, \exists!$ There exist, there exists exactly one (uniqueness)

$\in, \notin$ Is an element of, is not an element of

$\subseteq, \mapsto$ Sub-set, maps to

$: \text{ or } |$ Such that

iff If and only if

$\triangleq, \stackrel{\text{def}}{=}$ Equals by definition, is defined as

$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{bmatrix}$ Kronecker product of matrices $\mathbf{A} \in \mathbb{C}^{n,m}$ and $\mathbf{B}$

Introduction

Signal and power integrity analysis of high-speed interconnects and packages are becoming increasingly important. However, they have become extremely challenging due to the large circuit sizes and mixed frequency-time domain analysis issues. The circuit equations, despite being large, are fortunately extremely sparse. Exploiting sparsity lowers the computational cost associated with the application of numerical techniques on circuit equations. However, after some level of complexity and scale, the simulation of circuits in their original size is prohibitively expensive. Model order reduction (MOR) has proven successful in tackling this reality and hence, has been an active research topic in the CAD area. The goal of MOR is to extract a smaller but accurate model for a given system, in order to accelerate simulations of large complex designs. In order to preserve the accuracy of these downsized models over a large bandwidth, the order of the resulting macromodels may end up being high. On the other hand, any attempt of reduction can drastically impair the sparsity of the original system. The **large number of ports** even worsen the problem of being **high-order** and **dense**. Particularly, reduction of the circuit equations for electrical networks with large number of input/output terminals often leads to very large and dense reduced models. It is to be noted that, as the number of ports of a circuit increases (e.g. in the case of large bus structures), the size of reduced models also grows proportionally. This degrades the efficiency of transient simulations, significantly undermining the advantages gained by MOR techniques.

So far, MOR techniques for linear time invariant systems have been well-developed and widely used. On the other hand, nonlinear systems present numerous challenges for MOR. A common problem in the prominently used linear and nonlinear order-reduction techniques is the “selection of proper order” for the reduced models. Determining the “minimum” possible, yet “adequate” order is of critical importance to start the reduction process. This ensures that the resulting model can still sufficiently preserve the important physical properties of the original system. For both classes of physical systems, the selection of an optimum order is important to achieve a pre-defined accuracy while not over-estimating the order, which otherwise can lead to inefficient transient simulations and hence, undermine the advantage from applying MOR.

This thesis presents solutions for the above obstacles to ensure successful model reduction of large-scale linear and nonlinear systems. For this purpose, it proposes an efficient reduction algorithm to preserve the sparsity in the reduction of linear systems with large number of ports. Furthermore, it presents the efficient algorithms to determine the optimum order for linear and nonlinear macromodels.

Contributions

The main contributions of this thesis are as follows.

- A novel algorithm is developed for efficient reduction of linear networks with large number of terminals. The new method, while exploiting the applicability of the superposition paradigm for the analysis of massively coupled interconnect structures, proposes a reduction strategy based on flexible clustering of the transmission lines in the original network to form individual subsystems. The overall reduced model is

constructed by properly combining these reduced submodels based on the superposition principle.

The important advantages of the proposed algorithm are

- i)* It yields reduced-order models that are **sparse** and **block diagonal** for multiport linear networks
- ii)* It is not dependent on the assumption of certain correlations between the responses at the external ports; thereby it is input-waveform and frequency independent. Consequently, it overcomes the accuracy degradation normally associated with the existing low-rank approximation based terminal reduction techniques.
- The proposed algorithm establishes several important properties of the reduced-order model, including (a) stability, (b) block-moment matching properties, and (c) improved passivity. It is to be noted that, the flexibility in forming multi-input clusters with different sizes, as proposed in this algorithm, has been proven to be of significant importance. It establishes the block-diagonal dominance and passivity-adherence of the reduced-order macromodel.
- A robust and efficient novel algorithm to obtain an optimally minimum order for a reduced model under consideration is presented. The proposed methodology provides a geometrical approach to subspace reduction. Based on these geometrical considerations, This method develops the idea of monitoring the behavior of the projected trajectory in the reduced subspace. To serve this purpose, the proposed algorithm adopts the concept of "False Nearest Neighbor (FNN)" to the linear MOR applications. It also devises the mathematical means and quantitative measures to observe the behavior of near neighboring points, lying on the projected trajectory, when increasing the dimension of a reduced-space. To establishing the proposed methodologies, this

thesis exceeds beyond the extensive experimental justifications. It deeply contributes to the theoretical aspects involved in these algorithms by establishing new concepts, theorems and lemmas.

- A novel and efficient algorithm is developed to obtain the minimum sufficient order that ensures the accuracy and efficiency of the reduced nonlinear model. The proposed method, by deciding a proper order for the projected subspace, ensures that the reduced model can inherit the dominant dynamical characteristics of the original nonlinear system. The proposed method also adopts the concepts and mathematical means from the False Nearest Neighbors (FNN) approach to trace the deformation of nonlinear manifolds in the unfolding process. The proposed method is incorporated into the projection basis generation algorithm to avoid the computational costs associated with the extra basis. It is devised to be general enough to work in conjunction with any intended nonlinear reduced modeling scheme such as: TPWL with a global reduced subspace, TBR, or POD, etc.

As another important contribution, this thesis derives the bounds on the neighborhood range (radius) when searching for the false neighbors. Bounding this neighborhood range helps to enhance the efficiency of the automated algorithm by narrowing down the range of possible choices for the threshold value in the ratio test.

Organization of the Thesis

This thesis is organized as follows.

Chapter 1 presents a concise background on the main subjects relevant to this work such as, dynamical systems and their modelings as well as linear and nonlinear systems which are studied from a comparative perspective. Chapter 2 reviews the general concept of MOR

and physical characteristics which should be preserved in the reduction process. The next two chapters are of an introductory nature and provide an in-depth overview of the model reduction methods for linear (Chapter 3) and nonlinear (Chapter 4) dynamical systems. Next, Chapter 5 explains the details of the proposed methodologies for reduced macro-modeling of massively coupled interconnect structures. In Chapter 6, a novel algorithm for optimum order estimation is developed for reduced linear macromodels. This is followed by Chapter 7, which presents a novel algorithm for optimum order determination for reduced nonlinear models. Chapter 8 summarizes the proposed work and outlines the direction of future research.

Appendix-A further compares the properties of nonlinear and linear systems. Appendix-B presents some concepts from linear algebra and functional analysis that are useful for studying the dynamic systems. Appendices C and D present the proofs for the theorems in Chapter-5.

Chapter 1

Background and Preliminaries

This chapter presents a quick background on the main topics relevant to the subject of this work. The main characteristics of general classes of both linear and nonlinear systems are studied in a comparative manner. It also describes the groundwork for the electrical networks and their properties as a (linear / nonlinear) dynamical system. In addition, an overview of the formulation (mathematical modeling) for electrical networks is presented. For the supplementary concepts and more details about the important nonlinear phenomena Appendix A can also be referred to.

1.1 Dynamical Systems

A dynamical system is a system which changes in time according to some rule, law, or "*evolution equation*". The intrinsic behavior of any dynamical system is defined based on the following two elements [1],

- (a) a rule or "*dynamic*", which specifies how a system evolves,
- (b) an initial condition or "*initial state*" from which the system starts.

The dynamical behavior of systems can be understood by studying their mathematical descriptions. There are two main approaches to mathematically describe dynamical systems,

- (a) **differential equations** (also referred to as “flows”),
- (b) **difference equations** (also known as “iterated maps” or shortly “maps”).

Differential equations describe the evolution of systems in continuous time, whereas iterated maps arise in problems where time is discrete [2, 3]. Differential equations are used much more widely in electrical engineering, therefore we will focus on continuous-time dynamical systems.

1.2 Linear Systems

In system theory (or functional analysis, or theory of operators), “*linearity*” is defined based on the satisfaction of two properties, *additivity* and *homogeneity*, so called “*superposition*” paradigm. For a given function (map) $\mathcal{L}$ and any inputs u_i and u_j additivity states that, $\mathcal{L}(u_i + u_j) = \mathcal{L}(u_i) + \mathcal{L}(u_j)$, and homogeneity is $\mathcal{L}(k_i u_i) = k_i \mathcal{L}(u_i)$, where k_i is any arbitrary real number. Hence, the following compact definition of linearity is generally used:

$$\mathcal{L} \left(\sum_{p=1}^n k_p u_p \right) = \sum_{p=1}^n k_p \mathcal{L}(u_p), \quad n \geq 1, \quad \forall \{k_p\}, \{u_p\} \quad (\text{Superposition}). \quad (1.1)$$

Implicit in the above is the requirement that,

- For any linear function $\mathcal{L}(0) = 0$
- For $n = 1$ i.e. homogeneity (the linear scaling), $\mathcal{L}(k u) = k \mathcal{L}(u)$
- $\mathcal{L}(u_i - u_j) = \mathcal{L}(u_i) - \mathcal{L}(u_j)$

- u_p for $p = 1, \dots, n$ should be in the space of the possible inputs or the domain of the function $\mathcal{L}$. It is also required for the domain to be closed under linear combination; i.e., $k_i u_i + k_j u_j$ must belong to the domain if u_i and u_j do [4].

Given a physical system $\mathcal{L}$ as illustrated in Fig. 1.1, let the corresponding output $\mathbf{y}(t) = \mathcal{L}(\mathbf{u}(t), \mathbf{x}(t_0))$ for any two different setups of input $u(t)$ and initial conditions $\mathbf{x}(t_0)$ be as shown in (1.2a) and (1.2b).

Figure 1.1: Illustration of linear physical system $\mathcal{L}$.

$$\left. \begin{array}{l} \mathbf{x}(t_0) = \mathbf{x}_i \\ \mathbf{u}_i(t), \quad t \geq t_0 \end{array} \right\} \Rightarrow \mathbf{y}_i(t), \quad t \geq t_0, \quad (1.2a)$$

$$\left. \begin{array}{l} \mathbf{x}(t_0) = \mathbf{x}_j \\ \mathbf{u}_j(t), \quad t \geq t_0 \end{array} \right\} \Rightarrow \mathbf{y}_j(t), \quad t \geq t_0. \quad (1.2b)$$

In system theory, $\mathcal{L}$ is called a linear system if the following two conditions in (1.3) and (1.4) hold [5]:

$$\left. \begin{array}{l} \mathbf{x}(t_0) = \mathbf{x}_i + \mathbf{x}_j \\ \mathbf{u}_i(t) + \mathbf{u}_j(t), \quad t \geq t_0 \end{array} \right\} \Rightarrow \mathbf{y}_i(t) + \mathbf{y}_j(t), \quad t \geq t_0 \quad \textbf{(additivity)} \quad (1.3)$$

and

$$\left. \begin{array}{l} \mathbf{x}(t_0) = k_i \mathbf{x}_i \\ k_i \mathbf{u}_i(t), \quad t \geq t_0 \end{array} \right\} \Rightarrow k_i \mathbf{y}_i(t), \quad t \geq t_0 \quad \textbf{(homogeneity)} \quad (1.4)$$

for any real constant k_i .

The "superposition property" is generalized as

$$\left. \begin{array}{l} \mathbf{x}(t_0) = \sum_{p=1}^n k_p \mathbf{x}_p(t_0) \\ \sum_{p=1}^n k_p \mathbf{u}_p(t) \quad t \geq t_0 \end{array} \right\} \Rightarrow \sum_{p=1}^n k_p \mathbf{y}_p(t), \quad t \geq t_0, \quad (1.5)$$

for $p \geq 1$ and any $k_p \in \mathbb{R}$, where $\mathbf{y}_p(t) = \mathcal{L}(\mathbf{u}_p(t), \mathbf{x}_p)$

1.2.1 Important Property of Linear Systems

If the input $\mathbf{u}(t)$ is zero for $t \geq t_0$, then the output will be exclusively due to the initial state $\mathbf{x}(t_0)$. This output is called the "zero-input response" and will be denoted by $\mathbf{y}_{zi}(t)$ as

$$\left. \begin{array}{l} \mathbf{x}(t_0) \\ \mathbf{u}(t) \equiv 0, \quad t \geq t_0 \end{array} \right\} \Rightarrow \mathbf{y}_{zi}(t), \quad t \geq t_0. \quad (1.6)$$

If the initial state $\mathbf{x}(t_0)$ is zero, then the output will be excited exclusively by the input. This output is called the "zero-state response" and will be denoted by $\mathbf{y}_{zs}(t)$ as

$$\left. \begin{array}{l} \mathbf{x}(t_0) = 0 \\ \mathbf{u}(t), \quad t \geq t_0 \end{array} \right\} \Rightarrow \mathbf{y}_{zs}(t), \quad t \geq t_0. \quad (1.7)$$

The additivity property implies that,

$$\begin{aligned} \text{Response due to } \left\{ \begin{array}{l} \mathbf{x}(t_0) \\ \mathbf{u}(t), \quad t \geq t_0 \end{array} \right\} &= \text{Output due to } \left\{ \begin{array}{l} \mathbf{x}(t_0) \\ \mathbf{u}(t) \equiv 0, \quad t \geq t_0 \end{array} \right\} \\ &\quad + \text{Output due to } \left\{ \begin{array}{l} \mathbf{x}(t_0) = 0 \\ \mathbf{u}(t), \quad t \geq t_0 \end{array} \right\} \end{aligned} \quad (1.8)$$

or simply

$$\text{Response } \mathbf{y}(t) = \text{zero-input response } \mathbf{y}_{zi}(t) + \text{zero-state response } \mathbf{y}_{zs}(t).$$

Thus, the response of every linear system can be decomposed into the zero-state response and the zero-input response. Furthermore, the two outputs can be studied separately and their sum yields the complete response.

1.2.2 Mathematical Modeling of Linear Systems

The main stream studies on the mathematical modeling of linear systems originally started in the area of modern control theory (1950s). Thereafter, it has been extended to other disciplines such as electrical and mechanical engineering. The mathematical representation of linear dynamical systems is generally provided: (a) by means of *system transfer function matrix* and via (b) *differential equations* (or, sometimes, integro-differential equations). The former describes only the input-output property of the system, while the latter gives further insight into the structural property of the system.

1.2.2.1 Linear Time-Invariant Standard State-Space Systems

It can be remarked that, the most straightforward way to describe the dynamics of a linear time-invariant (LTI) physical system $\mathcal{L}$ is by means of *differential dynamic system*, which is a set of ordinary differential equations of the form

$$\mathcal{L} : \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t) & \text{(state equation)} \\ \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t) + \mathbf{D} \mathbf{u}(t) & \text{(output equation)}, \end{cases} \quad (1.9a) \quad (1.9b)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the vector of n system variables and $\dot{\mathbf{x}}(t)$ denotes the derivative of $\mathbf{x}(t)$ with respect to the time variable t . $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, $\mathbf{C} \in \mathbb{R}^{q \times n}$, and $\mathbf{D} \in \mathbb{R}^{q \times p}$ define the model dynamics. $\mathbf{u}(t) \in \mathbb{R}^p$ is a the vector of the excitations at the inputs, $\mathbf{y}(t) \in \mathbb{R}^q$ is the outputs, n is the the system order, and p and q are the number of system inputs and outputs, respectively. The equation (1.9) is sometimes referred to as (standard

or normal) *state-space* realization of the system.

The *state variables* are the smallest possible subset of system variables (“state variables” $\subseteq \mathbf{x}$) that can represent the entire state of the system at any given time. In other word, the state of a system may be considered to be the minimal amount of information necessary at any time to completely characterize any possible future behavior of the system. This leads to a least order realization of the system. Realizations of least order, also called *minimal* or *irreducible* realizations, are of interest since they realize a system, using the least number of dynamical elements (minimum number of elements with memory).

1.2.2.2 Solution of Linear Systems

Theorem 1.1. *The solution of state equation (1.9a) for prescribed $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{u}(\tau)$, $\tau \geq t_0$, is unique and is given by*

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}_0 + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau) d\tau. \quad (1.10)$$

In particular, the solution of the homogeneous equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \quad (1.11)$$

is

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}_0. \quad (1.12)$$

For a complete account of the proof, reader is encouraged to directly consult a text in Ordinary Differential Equation (ODE) e.g. [6–9] or an introductory text in linear system theory e.g. [4, 5] or a reference for linear circuit theory e.g. [10].

1.2.2.3 Linear Time-Invariant Descriptor Systems

Discussion of descriptor systems originated in 1977 with the fundamental paper [11]. Since then, the modeling of dynamical systems by descriptor systems (equivalently called singular systems, or semi-state systems, or differential-algebraic systems, or generalized state-space systems) have attracted much attention due to the comprehensive applications in many fields such as electrical engineering [12, 13]. The form of linear Differential-Algebraic Equation (DAE) so called *LTI descriptor model* is represented by [14]

$$\mathcal{L} : \begin{cases} \mathbf{E} \dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t), & \text{with } \mathbf{x}(t_0) = \mathbf{x}_0, \\ \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t) + \mathbf{D} \mathbf{u}(t), \end{cases} \quad (1.13a)$$

$$(1.13b)$$

where $\mathbf{E} \in \mathbb{R}^{n \times n}$ is generally a singular matrix (i.e. $\text{rank}(\mathbf{E}) = n_0 \leq n$). This is a general state-space equations, often expected in the formulation for circuit simulation.

If there are well-identified input and output variables but little or no interest in the behavior of the internal variables of the system, a convenient description is provided by the system *impulse response* $\mathbf{h}(t)$ or its Laplace transform. This is called system *transfer function matrix*, which is (more or less) the frequency domain equivalent of the time domain input-output relation [15]. Assuming zero initial conditions, the transfer function matrix $\mathbf{H}(s) : \mathbb{C} \rightarrow \mathbb{C}^{p \times p}$ of (1.13) is defined as

$$\mathbf{H}(s) \triangleq \mathcal{L}(\mathbf{h}(t)) = \int_0^\infty \mathbf{h}(t) e^{-st} dt = \mathbf{C} (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}, \quad s = \delta + j\omega, \quad (1.14)$$

where p is the number of input/output ports.

Definition 1.1 (Regularity). A linear descriptor system (1.13a), or the matrix pair $(\mathbf{E}, \mathbf{A})$, is called *regular* if there exists a constant scalar $\lambda \in \mathbb{C}$ such that

$$\det(\lambda \mathbf{E} - \mathbf{A}) \neq 0. \quad (1.15)$$

It is also equivalently said that, the matrix pencil $(s\mathbf{E} - \mathbf{A})$ of matrix pair $(\mathbf{E}, \mathbf{A})$ is *regular*.

The regularity of systems is the condition to make the solution to descriptor systems exist and unique. In the following chapters, some special features of regular descriptor linear systems such as state response and stability will be explained.

Finite Eigenvalues

Under the regularity assumption of the matrix pair $(\mathbf{E}, \mathbf{A})$, the polynomial

$$\Delta(s) = \det(s\mathbf{E} - \mathbf{A}) \quad (1.16)$$

is not identically zero ($\Delta(s) \not\equiv 0$). This polynomial (1.16) is called the *characteristic polynomial* of the system (1.13), which is of a certain degree (e.g. $\deg \Delta(s) = n_1$). Hence, it has n_1 (or less) finite roots ($s_i = \lambda_i$) satisfying $\Delta(s) = 0$. The finite roots of the system's characteristic polynomial are called the system *poles* or finite eigenvalues of the system, or the matrix pair $(\mathbf{E}, \mathbf{A})$. Thus, the set of finite poles of the system is

$$\lambda(\mathbf{E}, \mathbf{A}) = \{\lambda_i \mid \lambda_i \in \mathbb{C}, \lambda_i \text{ is finite, } \det(\lambda_i \mathbf{E} - \mathbf{A}) = 0\} . \quad (1.17)$$

the number of finite poles is always not greater than $n_1 = \text{rank}(\mathbf{E})$ ($\leq n$) for descriptor systems. Therefore, $\lambda(\mathbf{E}, \mathbf{A})$ contains at most n_1 number of complex numbers [16].

1.3 Nonlinear Systems

Any system that does not satisfy superposition property is nonlinear. It is worth noting that, there is no unifying characteristic of nonlinear systems, except for not satisfying “additivity” and “homogeneity” properties (*cf.* 1.2). A very general structure for models of

nonlinear dynamic systems is given by a set of nonlinear differential equations as

$$\mathbf{F}(t, \dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t)) = \mathbf{0}, \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad (1.18)$$

where $\mathbf{x}(t)$ is a $n \times 1$ vector of system variables, $\dot{x}_i(t)$ ($\in \dot{\mathbf{x}}(t)$) denotes the derivative of $x_i(t)$ ($\in \mathbf{x}(t)$) with respect to the time variable t , $\mathbf{u}(t) = \{u_1(t), \dots, u_p(t)\}^T$ is a vector of specified sources applied to the inputs, and $\mathbf{F}$ is a vector function as $\mathbf{F} : \mathbb{R} \times \mathbb{R}^n \mapsto \mathbb{R}^n$. Written in scalar terms, the i -th component equation in $\mathbf{F}$ has the form

$$f_i(t, \dot{x}_1, \dots, \dot{x}_n, x_1, \dots, x_n, u_1, \dots, u_p) = 0, \quad x_j(t_0) = x_{j0} \quad (1.19)$$

for $j = 1, \dots, n$.

For a *main class* of the nonlinear systems Ψ , their dynamical behavior may be adequately characterized by a finite number of coupled first-order *nonlinear ordinary differential equations* as shown in (1.20) [17].

$$\Psi : \begin{cases} \dot{x}_1(t) = f_1(t, x_1, \dots, x_n, u_1, \dots, u_p) \\ \dot{x}_2(t) = f_2(t, x_1, \dots, x_n, u_1, \dots, u_p) \\ \vdots \\ \dot{x}_n(t) = f_n(t, x_1, \dots, x_n, u_1, \dots, u_p) \end{cases} \quad (1.20)$$

The realization in (1.20) associated with another equation which is a (possibly nonlinear) measurement function can be equivalently rewritten in the following vector notation:

$$\Psi : \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{F}(t, \mathbf{x}(t), \mathbf{u}(t)) & \text{(state equation),} \\ \mathbf{y}(t) = \mathbf{h}(t, \mathbf{x}(t), \mathbf{u}(t)) & \text{(output equation).} \end{cases} \quad (1.21a) \quad (1.21b)$$

Similar to the linear case, the equations (1.21a) and (1.21b) together are referred to as

the *(standard) state-space model*, or simply the state model. Also, the smallest possible memory that the dynamical system needs from its past to predict the entire state of the system at any given future time is called state variables $\{x_i(t) \mid x_i \in \mathbf{x}(t)\}$.

One may rightfully question the applicability of (1.21) for all possible cases of nonlinear physical systems. In the late seventies (1978) [18], it became clear that *nonlinear descriptor systems* (1.18), rather than *standard Ordinary-Differential Equations (ODE)* (1.21), are more suitable for the modeling of the nonlinear dynamic systems in many applications such as electrical networks (*cf.* Sec. 1.4).

1.3.1 Solutions of Nonlinear Systems

There are powerful analysis techniques for linear systems, founded on the basis of the superposition principle (*cf.* Sec. 1.2). As we move from linear to nonlinear systems, we are faced with a more difficult situation. The superposition principle does not hold any longer and the analysis involves mathematical tools that are more advanced in concept and involved in detail.

This will be more clear by considering the following facts.

- (a) An important property of a linear system is that, when it is excited by a sinusoidal signal, the steady-state response will be sinusoidal with the same frequency as the input. Also, the amplitude and phase of the response are functions of the input frequency. In contrast, when a nonlinear system is excited by a sinusoidal signal, the steady-state response generally contains higher harmonics (multiples of the input frequencies). In some cases, the steady-state response also contains subharmonic frequencies.
- (b) For nonlinear systems, the complete response can be very different from the sum of the zero-input response and zero-state response. Therefore, we cannot separate the

zero-input and zero-state responses when studying nonlinear systems.

- (c) It is stated in Theorem-1.1 that, any linear system has a unique solution through each point in the state space for $0 \leq t \leq \infty$. However, it is only under certain conditions that, the nonlinear system has a unique solution at each point in the state space. For a linear system, the response settles down to a unique solution after the transient dies out. Nonlinear systems, on the other hand, can exhibit many qualitatively different coexisting solutions depending on the initial state [19]. In extreme cases, a nonlinear system can show chaotic behavior.

1.3.2 Linear versus Nonlinear

Nonlinear systems differ from linear systems in several fundamental ways. In Table 1.1 a summary of their general properties and characteristics are compared. More details are provided in Appendix A.

1.4 Mathematical Modeling of Electrical Networks

Being an inseparable part of the modern era, studying the dynamical behavior and the methods of mathematical modeling of electrical/electronic networks has moved to the center of attention in the past few decades. Ever increasing size, complexity and compactness of electrical designs has been enhancing the importance of such efforts to create accurate yet efficient system equations. This has been done with the main intention of inclusion of modern complex products in simulators (and virtual design environments) to ensure more “*realistic*” and “*efficient*” simulations.

- (a) *Realistic simulations* imply that the errors of the virtual models should be small,
-

Table 1.1: Summary: general properties of linear and nonlinear systems

	Linear Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$	Nonlinear Systems $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$
Equilibrium Points: A point where the system can stay forever without moving.	Unique If $\mathbf{A} \in \mathbb{R}^{n \times n}$ has rank n (full rank), then $\mathbf{x}_{equi} = \mathbf{0}$; otherwise, the solution lies in the null space of $\mathbf{A}$.	Multiple $\mathbf{F}(\mathbf{x}_{equi}) = \mathbf{0}$, n coupled nonlinear equations in n unknowns; the number of possible solutions may vary from 0 to $+\infty$.
Escape Time: The state of an unstable system:	$\mathbf{x} \rightarrow +\infty$, as $t \rightarrow +\infty$ goes to infinity as time approaches infinity!	$\mathbf{x} \rightarrow +\infty$, in $t \leq +\infty$ can go to infinity in finite time!
Stability:	The equilibrium point is stable if all eigenvalues of $\mathbf{A}$ have negative real part, regardless of Initial Conditions (IC).	Stability about an equilibrium point: <ul style="list-style-type: none"> • Dependent on initial condition • Local vs. Global stability important • Possibility of limit cycles
Oscillation vs. Limit Cycles:	Oscillation <ul style="list-style-type: none"> • Needs conjugate poles on imaginary axis • Almost impossible to maintain • Amplitude depends on IC 	Limit Cycles <ul style="list-style-type: none"> • A unique, self-excited oscillation with fixed amplitude and frequency • A closed trajectory in the state space • Independent of IC
Forced Response	$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}(t)$ <ul style="list-style-type: none"> • The principle of superposition holds. • I/O stability $\rightarrow$ bounded-input, bounded-output • Sinusoidal input $\rightarrow$ sinusoidal output of same frequency 	$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x} + \mathbf{u}(t))$ <ul style="list-style-type: none"> • The principle of superposition does not hold in general. • The I/O ratio is not unique in general, may also not be single-valued.
Steady-State Behavior The asymptotic response, when $t \rightarrow +\infty$	Unique The response settles down to a <i>unique</i> solution, independent of IC.	Non-unique / Multi-stability Can have many different coexisting solutions depending on IC. Chaos Complicated steady-state behavior, may exhibit randomness despite the deterministic nature of the system.

The above comparison is based on [20], with enhancements and modifications.

which requires that, the important physical characteristic of the product must be taken into account in the mathematical model.

- (b) *Efficient simulations* (maybe paradoxically) imply that it is not necessary to include all minute detail of a physical design in the simulator.

The latter opens the door to a trend in the area of computational science and engineering as “model order reduction” that is the main subject in this thesis and its thorough explanation will be forthcoming. The former explains the importance of the systematic formulation of the dynamic equations for electrical networks.

The next section presents an overview of the mathematical modeling of electrical networks as dynamical systems.

1.5 Overview of Formulation of Circuit Dynamics

Electrical networks (e.g. RLC circuits) are examples of dynamical systems, whose state-space dynamics for time $t \geq 0$, can accurately be captured by a set of first-order coupled differential equations [21–23]. Since the early sixties, it is known that, descriptive equations of electrical circuits belong to the class of *differential equations* on differentiable manifolds e.g. see [23–26]. This result is related to the celebrated paper of Moser and Brayton [27] in 1964 where their equations for the description of (reciprocal and) nonlinear circuits are written in coordinates. It lasted another few years until the equations of Moser and Brayton were reformulated by Smale [22] by means of the framework of modern differential geometry. Further work was done by Matsomoto [28], Ishiraku [29] and later by others (e.g.) [30, 31] to refine this approach for describing electrical networks.

The differential-equation approach to the identification of electrical circuits immediately led to the necessity of the numerical determination of the transient response, an extremely important area often more limited by the "stiffness" phenomenon [32,33]. In the late sixties, "the time-constant problem" was an infamous source of frustration for users of computer programs for the analysis of circuits. This obstacle to construct an efficient and general purpose circuit simulator was solved mainly by contributions of Gear [34] and Sandberg and Shichman [35]. It was emphasized by Gear (1968) that *circuit equations should be considered as Differential-Algebraic Equations (DAE)* (cf. Sec. 1.2.2.3). It was not until a decade later, when Linda Petzold - a former Ph.D. student of Gear - found out in 1982 that "DAEs are not ODEs" [36]. She showed that only some of the differential-algebraic systems can be solved using numerical methods which are commonly used for solving stiff systems of ordinary differential equations. She also indicated the causes of the associated difficulties and presented solutions mostly for linear cases.

The DAEs systems to represent electrical networks are directly obtained using the modified nodal analysis (MNA) matrix formulation [37–39], which will be reviewed in the following sections.

1.5.1 MNA Formulation of Linear Circuits

In the case that, all the components in the circuit are linear and Kirchoff's laws also hold, the circuit is considered as a linear network. Time-domain realization for multi-input and multi-output (MIMO) dynamical linear circuits (Ψ) in the descriptor form resulting from MNA matrix formulation [37–39] is represented as:

$$\Psi : \begin{cases} \mathbf{C} \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{L} \mathbf{x}(t), \end{cases} \quad (1.22a)$$

$$(1.22b)$$

where $\mathbf{C}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are susceptance and conductance matrices, respectively, $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of MNA variables (the nodal voltages and some branch currents) [38] of the circuit. Also, $\mathbf{B} \in \mathbb{R}^{n \times p}$ and $\mathbf{L} \in \mathbb{R}^{q \times n}$ are the input and output matrices, associated with p inputs and q outputs, respectively.

1.5.2 MNA Formulation of Nonlinear Circuits

The DAE representation for circuits that include non-linearity is generally obtained by intuitively adding a vector of nonlinear functions representing the nonlinear elements to (1.22) based on the principals of nodal analysis. It can be rightfully questioned that, how the result of MNA formulation is related to the general class of nonlinear dynamic systems. In other words, how much of the generality is scarified by adapting the general formulation (e.g.) in (1.18). To illuminate this, we cautiously move in a reverse direction. We start from the general form in (1.18) and by considering the nature of circuits' topology we attempt to modify it.

1.5.2.1 A Reverse Approach

The differential-algebraic equations that generally characterize any possible nonlinear systems (including electrical networks) is shown in (1.18), as repeated here

$$\mathbf{F}(t, \mathbf{x}(t), \dot{\mathbf{x}}(t), \mathbf{u}(t)) = \mathbf{0}, \quad \text{with } \mathbf{x}(0) = \mathbf{x}_0. \quad (1.23)$$

Finding a solution for (1.23) in its most general form can be prohibitively complex and hence, not always possible.

Without loss of generality, nonlinear circuits can be characterized as a system of the

first order coupled differential equations in the following form:

$$\Sigma : \begin{cases} \frac{d}{dt}\mathbf{g}(\mathbf{x}(t)) = \mathcal{F}(\mathbf{x}(t), \mathbf{u}(t)) & \text{(state equation),} \\ \mathbf{y}(t) = \mathbf{h}(\mathbf{x}, \mathbf{u}) & \text{(output / measurement equation),} \end{cases} \quad (1.24a)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^p$, $\mathbf{y}(t) \in \mathbb{R}^q$, $\mathbf{g}(\mathbf{x}(t))$, $\mathcal{F}(\mathbf{x}(t), \mathbf{u}(t)) \in \mathbb{R}^n$ and $\mathbf{h}(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^q$.

A complex design consists of sub-circuits that are connected together. These sub-circuits interact with the surrounding sub-networks through its external nodes as shown in Fig. 1.2. Taking the state variable approach to MIMO nonlinear electrical networks,

Figure 1.2: Illustration of a subcircuit that accepting p -inputs and interacting with other module through its q -outputs.

there are several established methods, namely sparse-tableau and modified nodal analysis (MNA) formulation [37–39] to characterize circuits.

1.5.2.2 MNA Formulation

The electrical networks, as an important class of nonlinear system, can also be well characterized using a system of first order, differential-algebraic equation system as shown in 1.24. The general form of the nonlinear -equations in 1.24 can be adapted to the major class of the nonlinear dynamical circuits by considering the following remarks:

Remark 1.1. Being expressed in the "normal" DAE form, the nonlinearity is focused on the state variables $\mathcal{F}(\mathbf{x}(t))$. Hence, the state-space equation in (1.24a) will be

$$\Sigma : \frac{d}{dt} \mathbf{g}(\mathbf{x}(t)) = \mathcal{F}(\mathbf{x}(t)) + \mathcal{P}(\mathbf{x}(t), \mathbf{u}(t)) \quad (\text{state equation}). \quad (1.25)$$

Remark 1.2. In an electrical network (e.g.: in Fig. (1.2)), the so-called input terminals are the nodes of the circuit interfacing with the input sources. Hence, the (input) sources are directly connected to the selected (interface) nodes of the circuit. In the *nodal analysis*, the currents from the sources are added to (or subtracted from) the KCL equation for corresponding nodes. For the case of voltage sources, the node voltage is directly decided (equated) by the voltage of the source. Accordingly, it can be generalized that the effect of the sources is "linearly injected" [40,41] to the system at the associated nodes. To distribute the effect of the sources, a selection matrix $\mathbf{B}$ can be directly applied to the source vector $\mathbf{u}(t)$ to decide the sources connected to each nodes. Under this mild practical assumption, the system equations for electrical networks fall in affine form in which (1) the source term is linearly combined with the rest of the formulation and (2) nonlinearities include only in the constitutive relations for the nonlinear elements as:

$$\Sigma : \frac{d}{dt} \mathbf{g}(\mathbf{x}(t)) = \mathcal{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t) \quad (\text{state equation}). \quad (1.26)$$

Remark 1.3. Even if in a special case, a nonlinear dependence on the input $\mathbf{u}(t)$ is assumed

as $\mathbf{B} \hat{\mathcal{P}}(\mathbf{v}(t))$, the nonlinear dependence on input can often be bypassed by treating the whole term $\hat{\mathcal{P}}(\mathbf{v}(t))$ as input [42]. Thus, the network equations fall in the form shown in (1.26).

Remark 1.4. The MNA formulation generally leads to certain form for the vector function $\mathcal{F}(\mathbf{x})$ that commonly occurs as $\mathcal{F}(\mathbf{x}) = -\mathbf{G}\mathbf{x}(t) + \mathbf{F}(\mathbf{x})$, where $\mathbf{G} \in \mathbb{R}^{n \times n}$ is conductance matrix and $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$ is the vector of nonlinear functions, including all the nonlinearities in the circuit. Hence, the system equation in (1.26) falls in the following form

$$\Sigma : \frac{d}{dt} \mathbf{g}(\mathbf{x}(t)) = \mathbf{G}(\mathbf{x}) + \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t) \quad (\text{state equation}). \quad (1.27)$$

Remark 1.5. Based on a similar approach for inputs, outlined in Remark-1.2 the outputs in MNA formulation $\mathbf{y}(t)$ are simply selections (generally not many) of the the voltages and currents in $\mathbf{x}(t)$. This selection can be performed by applying a properly decided selection matrix $\mathbf{L}$ in the output equation. Let us stress that in the MNA formulation the output signals are not explicitly dependent on the inputs $\mathbf{u}(t)$. Hence, the output equation commonly occurs in the form shown in (1.28b).

In summary, the electrical systems whose dynamics are formulated based on the MNA approach at time t can be generally described by the nonlinear, first order, differential-algebraic equation system. The equations encountered often in the practical situations is of the form:

$$\Sigma : \begin{cases} \frac{d}{dt} \mathbf{g}(\mathbf{x}(t)) = -\mathbf{G}\mathbf{x}(t) + \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t), & \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{y}(t) = \mathbf{L}\mathbf{x}(t), \end{cases} \quad (1.28a)$$

$$(1.28b)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of circuit variables in time t . The vector-valued functions $\mathbf{g}$, $\mathbf{F} \in \mathbb{R}^n$ respectively represent the NL susceptances of (e.g.) nonlinear

capacitors and nonlinear inductors and of conductance from nonlinear elements such as nonlinear resistors and diodes. Also, $\mathbf{G} \in \mathbb{R}^{n \times n}$ is the conductance matrix including the contributions of linear elements (such as linear resistors) and $\mathbf{B} \in \mathbb{R}^{n \times p}$ is the distribution matrix for the excitation vector $\mathbf{u}(t) \in \mathbb{R}^p$ and $\mathbf{L} \in \mathbb{R}^{q \times n}$ is the selection (measurement) matrix that defines the output response $\mathbf{y}(t) \in \mathbb{R}^q$. [40,41,43].

In MNA formulation of a major class of nonlinear circuits, it is generally possible that we limit the reactive matrix function as $\mathbf{g}(\mathbf{x}) = \mathbf{C} \mathbf{x}(t)$, while migrating (the stamps of) all the nonlinear elements (such as nonlinear inductors and nonlinear capacitors) to the function $\mathbf{F}(\mathbf{x})$. Hence, The resulting nonlinear state-space equation will be as shown below.

$$\Sigma : \begin{cases} \mathbf{C} \frac{d}{dt}(\mathbf{x}(t)) = -\mathbf{G}\mathbf{x}(t) + \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t), & \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{y}(t) = \mathbf{L}\mathbf{x}(t), \end{cases} \quad (1.29a)$$

$$(1.29b)$$

$\mathbf{C} \in \mathbb{R}^{n \times n}$ is susceptance matrix including the stamps of linear capacitors and inductors, $\mathbf{F}(\mathbf{x})$, also, includes all the nonlinearity effect in the circuit.

Chapter 2

Model Order Reduction - Basic Concepts

One may find a variety of interpretations for the topic of model order reduction in different disciplines. The common theme in all of them is that, given a large-scale dynamical system (linear or nonlinear) with predefined input and output terminals, a small-scale system is found that approximates the behavior of the original system at its terminals. To achieve this, the concepts and techniques of mathematical approximation for large differential-equation systems come into the picture. Hence, the concepts of MOR has generally been associated with the terms such as “dimensionality reduction”, “reduced-bases approximation”, “high energy dynamic modes”, “balancing the gramians” and “state truncation”. The concept itself and even almost all the techniques have been originally introduced in mathematics, mainly in the context of differential equations. Due to the feasibility of the idea, later it has been carried over to the control area and then to the fields such as civil engineering, aerospace engineering, earthquake engineering, mechanical and Micro-Electro-Mechanical Systems (MEMS), and VLSI circuits design which is the main subject of focus in this thesis.

This chapter provides an introduction and explains the fundamental concepts relevant to the subject of this thesis, in which we consider approximations of continuous-time dynamical systems.

2.1 Motivation

The problem of model order reduction of linear and nonlinear dynamical systems has been widely studied in the last two decades and is still considered an active topic, attracting much attention. Due to the ever-enhancing capability of methods and computers to accurately (and thus complexly) model real-world systems, simulation or, more generally, computational science has been proven as a reliable method for identifying, analyzing and predicting the behavior of systems. This is to such a degree that, simulation has become an important part of today's technological world, and it is now generally accepted as the third discipline, besides the classical disciplines of theory and experiment (analytical and observational forms).

Computer simulations are now utilized in almost every physical, chemical and other processes. Computer Aided design and virtual environments have been set up for a variety of problems to ease the work of designers and engineers. In this way, new products can be designed faster, more reliably, and without having to make costly prototypes [44]. In order to speed up the computation time it is a good idea to simplify the model, either in size or in complexity. Reducing the order of a model involves reducing the size of the mathematical model, but at the same time to preserve its essential features.

2.2 The General Idea of Model Order Reduction

A major class of physical systems and phenomena can be mathematically modeled [2, 45, 46] with a set of Partial Differential Equations (PDE), which adequately describes the physical behavior of the system under consideration. The spatial discretization of the PDE yields a system of ordinary differential equations (ODE), which in turn approximates the original PDE model. The dimension of this ODE system is governed by the spatial mesh

size. Thus, the finer the mesh, the larger the dimension of the resulting system of ODEs will be that have to be solved in time. Depending on the dimension of the original PDE and the desired spatial accuracy, the number of variables can extend from hundreds to several millions. If the interest lies in the time response of the system, as in the fields of structural and fluid dynamics, now this large system of ODEs has to be integrated in time to obtain the solution of the system: solving for the time response of the system means tracking the time evolution of the many variables of the system of ODEs. Finding the time-domain response (transient behavior) of such large systems would require excessive computational effort. MOR is an immediate answer to address this issue.

Model order reduction process starts from a large system of N ODEs

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^N \quad (2.1)$$

that results from the spatial discretization of the PDE, which we want to approximate with (a simpler model or) a smaller set of m ODEs of the form shown below (2.2), while preserving the main characteristic of the original (ODEs, PDEs and hence the physical) systems.

$$\frac{d\mathbf{z}}{dt} = \mathbf{g}(\mathbf{z}, t), \quad \mathbf{z} \in \mathbb{R}^m \quad (2.2)$$

The first requirement of model order reduction is that the number of states m , i.e. the number of differential equations of the reduced model given by (2.2), is much smaller compared to the number of states (N) of the original model in (2.1),

$$m \ll N. \quad (2.3)$$

To further illustration, one may consider the transmission lines, where Maxwell's equations

(the equations describing electromagnetic fields) are applied to the geometry. The transmission line equations (Telegrapher's equations) [47] can be derived by discretizing the line into infinitesimal section of length (Δx) and assuming uniform per unit length (p.u.l.) parameters of resistance, inductance, conductance, and capacitance. The segments of the line are decided to be electrically small (much smaller than a wavelength at the excitation frequency), to the aim that, lumped-circuit approximation of the exact per-unit-length distributed-parameter model can be adequately used. As a result, a cascade structure of multiple lumped filter sections (and Kirchhoff's laws) is used to replace Maxwell's equations in analyzing transmission lines. The ODEs formulation for such large circuits, having a few thousands variables for each interconnect, can be prohibitively large, even with a moderate accuracy expectation. Chapter 5 explains, how model order reduction can be utilized to address this issue.

The idea of MOR has been proved as an useful tool to obtain efficiency in simulations while ensuring desired accuracy. Its applicability to real life problems has made it a popular tool in many branches of science and engineering. Fig. 2.1 pictorially explains this process.

2.3 Model Accuracy Measures

Attainable accuracy from the resulting reduced macromodel is an important concern in the reduction process. To decide, how well a reduced system *approximates* the original system, we require a measure to quantify the accuracy. The straightforward way is to define the error (time-domain) signal $\zeta(t)$ as the deviation between two responses, from the model and from the original system as illustrated in Fig. 2.2. The difference between outputs should be measured at the same n time instances and for the same input signal $u(t)$. For the case of linear systems such deviation can also be judged comparing the

Figure 2.1: Model order reduction.

Figure 2.2: Measuring error of approximation.

frequency response of the original system and the one from the reduced transfer function at the n frequency points throughout the frequency spectrum of interest. The results for “single-input and single-output” (SISO) systems will be a vector and for “multi-input and multi-output” (MIMO) cases it is a matrix containing the instantaneous errors at different “ports”.

The error space (the space, where error resides) is considered as metric space (definition B.3) endowed with different norms that can be properly used to characterize the error (Sec. B.1.2). Table 2.1 presents a summary of the commonly used measures to quantify the error in the context of (linear / nonlinear) MOR.

Table 2.1: Measuring reduction accuracy in time domain

Name	Definition of $\mathcal{E}$
<i>mean squared error</i>	$\frac{\ \zeta\ _e^2}{n}$, where $\ \cdot\ _e$ is Euclidean norm
<i>normalized mean squared error</i>	$\frac{\ \zeta\ _e^2}{\text{var}(y)}$, where “ var ” denotes the <i>variance</i> * of data set
<i>root mean squared error</i>	$\frac{\ \zeta\ _e}{\sqrt{n}}$
<i>normalized root mean squared error</i>	$\frac{\ \zeta\ _e}{\sqrt{\text{var}(y)}}$
<i>mean absolute error</i>	$\frac{\ \zeta\ _1}{n}$
<i>mean absolute relative error</i>	$\frac{\left\ \frac{\zeta}{y} \right\ _1}{n}$

*- For a data set $y = \{y_i\}$ including N data points, variance is computed as

$$\text{var}(y) \triangleq \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2, \text{ where } \bar{y} \text{ is the data mean } \bar{y} \triangleq \frac{1}{N} \sum_{i=1}^N y_i.$$

It is to be noted that, depending on the application, other accuracy measures can also

be considered.

2.3.1 Error in Frequency Domain

For Linear systems, the error can also be measured based on the frequency domain response. It is done in a similar fashion as the definitions in Table 2.1.

Example 2.1. $\mathcal{L}_\infty$ Error: In some applications (e.g. in TBR *c.f.* Chapter. 3) it is more feasible that a measure of error in $\mathcal{L}_\infty$ norm (ζ_∞) is used, as shown below.

$$\zeta(s) = \mathbf{Y}(s) - \hat{\mathbf{Y}}(s), \quad (2.4a)$$

$$\zeta_\infty \triangleq \|\zeta(s)\|_\infty = \|\zeta(s)\|_\infty. \quad (2.4b)$$

In a single port case, it is the “maximum absolute error” that occurs throughout the frequency range of the observation.

To shed more light on the concept, the following instance of error measure is also considered as an explanatory example.

Example 2.2. A frequency-domain error is defined as shown below. It is typically used in the parametrized reduction of the MIMO system (e.g.) in [48].

$$\mathcal{E} = \|\mathbf{E}(s_o)\|_\infty = \max_{i,j} \left| \frac{\zeta_{ij}(s_o)}{|\mathbf{Y}_{ij}(s_o)|} \right|, \quad (2.5a)$$

where

$$\mathbf{E}(s_o) = \frac{\zeta_{ij}(s_o)}{|\mathbf{Y}_{ij}(s_o)|} : \quad \|\mathbf{E}(s_o)\|_e = \|\mathbf{E}(s)\|_e|_\infty = \max_{s \in [s_{\min}, s_{\max}]} \|\mathbf{E}(s)\|_e, \quad (2.5b)$$

where

$$\mathbf{E}(s) = \frac{\zeta_{ij}(s)}{|\mathbf{Y}_{ij}(s)|}, \quad (2.5c)$$

and where

$$\zeta_{ij}(s) = Y_{ij}(s) - \hat{Y}_{ij}(s). \quad (2.5d)$$

The error in (2.5) can be determined according to the following explanations.

- At each frequency point, every entry of the absolute error matrix from (2.5d) is normalized (scaled) by the magnitude of the corresponding entry in the original response matrix, as shown in (2.5c).
- $\mathbf{E}(s_o)$ in (2.5b) is the normalized error matrix evaluated at the frequency point s_o such that, it has the maximum euclidean norm in the frequency spectrum of interest $s \in [s_{\min}, s_{\max}]$.
- The error $\mathcal{E}$ in (2.5a) is defined as the magnitude of the entry in $\mathbf{E}(s_o)$ matrix which has the largest magnitude when compared to the other entries.

This measure is adequate if one wants to preserve very small entries of the transfer function, or in cases where at certain frequencies the transfer function is infinitely large [48].

2.4 Model Complexity Measures

As we already explained, the main intention of applying model order reduction is to reduce costs of simulation of such systems. Hence, in a general sense, the (CPU) time associated with simulating the model alone or as a sub-system in a hierarchy of a system can be used as a practical measure for the complexity of the model. As models become more complex, simulation cost also rises.

It is to be noted that, while the original system (specifically in circuit theory) is highly sparse, any attempt of reduction can impair the sparsity to a certain degree. For systems

with many I/O terminals, this fact may preclude any advantage expected from MOR (*c.f.* Chapter 5).

2.5 Main Requirements for Model Reduction Algorithms

A number of requirements should necessarily be satisfied, when extracting the macromodel from a detailed physical description of the original system. The most important ones that ensure feasibility of the resulting models may be summarized as follows

- **Accuracy:** The reduction technique should lead to an adequately accurate model for the original system. The reduced model should closely follow the terminal behavior of the original system.
 - **Compactness:** It should significantly reduce the number of variables or states, as compared to the original model.
 - **System properties preservation:** For many types of problems, it is desirable that the reduced models conserve the main physical properties of the original system, i.e. *passivity* and *stability*. Due to the importance of such characteristics, further details are discussed in Sections 2.6 and 3.1.
 - **Computationally efficient:** The MOR algorithm should extract a model which is relatively inexpensive to simulate and store it in the computer's memory. The computational cost for simulating the resulting model should be much lower than the cost for the original model.
 - **Inexpensive algorithm:** The model reduction algorithm should be relatively inexpensive to apply and it is preferable to be automated. In other words, the extraction
-

process needs to be practically repeatable in any phase of design, optimization or verification directly by the designers with a reasonable cost.

The resulting reduced model should accurately mimic the dynamics of the original subsystem in operating regimes which are different from the ones used to construct the model. This assures the accuracy of the results when the submodel is embedded in the hierarchy of a design and undergoes a higher level system level simulation. This property is referred to as *Transportability* [49] in the literature of nonlinear behavioral macromodeling [19, 49–54].

2.6 Essential Characteristic of Physical Systems

Among the essential characteristic that should be passed on to the successor representative of physical systems, *stability* and *passivity* are prominent. This section reviews the basic concepts and definitions in the most general form. Due to its direct relevance to our work, later in this report, we will study the detail of stability and passivity for linear and nonlinear systems.

2.6.1 Stability of Dynamical Systems

General stability concepts and theory play a central role in studying dynamical systems and have been extensively studied in the literature of system theory [4, 5, 17, 48, 55–58]. For an arbitrary dynamical system, stability is studied in the frame of two major notions:

- (a) **Internal stability:** The internal stability considers the trajectory (response curve) of an autonomous system

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)), \quad (2.6)$$

that is a time invariant system without any input. The response of autonomous systems (2.6) is only due to the nonzero initial state $\mathbf{X}_0$, that is also called, zero-input response. It is referred to as internal because such stability is decided based on the internal dynamics of the system without any outside intervention.

- (b) **External stability:** The external stability (so-called, Input-Output stability) is connected with how much the system amplifies the input signals to provide the output.

The following stabilities are well-known in literature and have been extensively studied [17][ch.:4,5,7,8], [4, 5, 57, 58].

2.6.2 Internal Stability

For an autonomous system in (2.6), internal stability can be studied as global or local stability of “equilibrium states”. An equilibrium point is a state that the autonomous system can maintain for an infinite time.

2.6.2.1 Local Stability of Equilibrium Points

Definition 2.1 (Lyapunov stability). An equilibrium state of (2.6) x_{eq} is called Lyapunov stable (or simply stable), if $\forall \epsilon > 0, \exists \delta = \delta(\epsilon) > 0$, such that

$$\|\mathbf{x}(t_0) - \mathbf{x}_{eq}\| < \delta \implies \|\mathbf{x}(t) - \mathbf{x}_{eq}\| < \epsilon, \quad \forall t > t_0 \quad (2.7)$$

otherwise it is not stable.

This states that, an equilibrium point is called (Lyapunov) stable if all solutions starting at nearby points stay nearby and remains bounded as $t \rightarrow \infty$ (without necessarily going to zero); otherwise, it is unstable.

Definition 2.2 (Asymptotic stability). An equilibrium state of (2.6) x_{eq} is asymptotically stable if it is Lyapunov stable and δ can be chosen such that

$$\|\mathbf{x}(t_0) - \mathbf{x}_{eq}\| < \delta \implies \lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{x}_{eq}. \quad (2.8)$$

It is said that, an equilibrium point is asymptotically stable if all solutions starting at nearby points not only stay nearby (i.e. Lyapunov stable), but also tend to the equilibrium point x_0 as time approaches infinity. Loss-less (conservative) systems, such as pure LC circuits, are examples of (Lyapunov) stable but are not asymptotically stable systems.

In 1892, Lyapunov showed that certain other functions (hence the name Lyapunov functions) could be used instead of energy to determine stability of an equilibrium point as shown in the following Lyapunov theorem. For convenience, we state the theorem for the case when the equilibrium point is at the origin of $\mathbb{R}^n$; that is, $\mathbf{x}_{eq} = \mathbf{0}$. There is no loss of generality in doing so because any equilibrium point can be shifted to the origin via a change of variables.

Theorem 2.1 (Lyapunov's stability theorem [17, 59]). Let $\mathbf{x}_{eq} = \mathbf{0}$ be an equilibrium point for $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ and $\mathcal{D} \subset \mathbb{R}^n$ be a domain containing $\mathbf{x}_{eq} = \mathbf{0}$ (origin). Let $V : \mathcal{D} \mapsto \mathbb{R}$ be a continuously differentiable function such that

$$V(\mathbf{0}) = 0 \text{ and } V(\mathbf{x}) > 0 \text{ in } \mathcal{D} - \{\mathbf{0}\}. \quad (2.9)$$

If

$$\dot{V}(\mathbf{x}) \leq 0, \text{ in } \mathcal{D} \quad (2.10)$$

then, $\mathbf{x} = \mathbf{0}$ is stable. Moreover, if

$$\dot{V}(\mathbf{x}) < 0 \text{ in } \mathcal{D} - \{\mathbf{0}\} \quad (2.11)$$

then $\mathbf{x} = \mathbf{0}$ is asymptotically stable.

In Theorem 2.1, $\dot{V}(\mathbf{x})$ denotes the derivative of $V(\mathbf{x})$ along the trajectories of

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}) \ \dots \ f_i(\mathbf{x}) \ \dots \ f_n(\mathbf{x})]^T, \quad (2.12)$$

given by

$$\begin{aligned} \dot{V}(\mathbf{x}) &= \sum_{i=1}^n \frac{\partial V}{\partial x_i} \dot{x}_i = \sum_{i=1}^n \frac{\partial V}{\partial x_i} f_i(\mathbf{x}) = \left[\frac{\partial V}{\partial x_1}, \ \frac{\partial V}{\partial x_2}, \ \dots, \ \frac{\partial V}{\partial x_n} \right] \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} \\ &= \frac{\partial V}{\partial \mathbf{x}} \mathbf{F}(\mathbf{x}). \end{aligned} \quad (2.13)$$

It is to be noted that, Lyapunov stability theorem's conditions are *only sufficient*. They do not say whether the given conditions are also necessary. Hence, failure to satisfy the conditions for stability or asymptotic stability does not mean that the equilibrium is not stable or asymptotically stable. It only means that such a stability property cannot be established by using the Lyapunov function shown in Theorem 2.1. Whether the equilibrium point is stable (asymptotically stable) or not can be determined only by further investigation. For more details one can consult [17].

2.6.2.2 Global Stability of Equilibrium Points

Theorem 2.2 (Global asymptotic stability theorem [17]). *Let $\mathbf{x} = \mathbf{0}$ be an equilibrium point for $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$. Let $V : \mathbb{R}^n \mapsto \mathbb{R}$ be a continuously differentiable function such that*

$$V(\mathbf{0}) = 0 \quad \text{and} \quad V(\mathbf{x}) > 0 \quad \forall \mathbf{x} \neq \mathbf{0} \quad (2.14)$$

$$\|\mathbf{x}\| \rightarrow \infty \Rightarrow V(\mathbf{x}) \rightarrow \infty \quad (2.15)$$

$$\dot{V}(\mathbf{x}) < 0 \quad \forall \mathbf{x} \neq \mathbf{0} \quad (2.16)$$

then $\mathbf{x} = \mathbf{0}$ is globally asymptotically stable.

2.6.3 External Stability

An input $u(t)$ is said to be bounded if there exists a constant α such that $|u(t)| \leq \alpha < \infty$ for $-\infty < -T \leq t < \infty$, where T is any arbitrary time point prior to the reference time $t = 0$. Or, it can be said that a bounded signal does not grow to positive or negative infinity. We shall say that a causal system is *BIBO-stable* (bounded input - bounded output stable) if a bounded input necessarily produces a bounded output $|y(t)| < \gamma$ for $-T \leq t < \infty$.

2.6.4 Passivity of a Dynamical Model

Another important characteristic for the models is passivity. It is said that, reduced model for a passive original system should preserve passivity. It is important because, stable but non-passive models may lead to unstable systems when connected to other passive components. On the other hand, a passive macromodel, when terminated with any arbitrary passive load, always guarantees the stability of the overall network. Generally speaking, a model of a circuit is passive if it does not generate energy. This notion ultimately depends

on the nature of the input and output signals as shown below.

Definition 2.3 (Passivity [17]). The system $y = h(t, u)$ is

- passive if $u^T y \geq 0$,
- lossless if $u^T y = 0$,

for all (t, u) .

2.7 The Need for MOR for Electrical Circuits

The ever increasing demand for higher data rates, lower power, and multifunction capable products are necessitating newer generations of complex and denser electronic circuits and systems. With the rapid increase in signal-speeds and decreasing feature sizes, high-frequency effects become the dominant factors limiting overall performance of microelectronic systems. The high complexity in design and the necessity of capturing the high-speed effects lead to extremely large models, to ensure sufficient accuracy in simulations. The processing cost for simulation of such large models can be prohibitive. It is to be noted that, the simulation of these models generally needs to be repeated many times during design, optimization and verification processes. Initial interest in model reduction (MOR) techniques stemmed from efforts to improve the simulation efficiency by reducing the circuit complexity while producing a good approximation for the input-output behavior of large structures. Hence, MOR is specifically useful when a compact macromodel is required to represent the signal behavior at the ports of the circuit block in a higher level simulation. The order reduction problems (in general) can be categorized as *linear MOR* and *nonlinear MOR*. Reduced-order modeling is well established for linear circuit systems such as electrical interconnect, whereas the available techniques for reduction of nonlinear circuits are limited in number and the scope of application. The complex nature of nonlinear phenomena makes nonlinear model order reduction a challenging area.

Chapter 3

Model Order Reduction for Linear Dynamical Systems

The previous chapter (2) presented a general discussion on computational science, and the need for compact models of phenomena observed in nature and industry. It was remarked that, the basic motivation for system approximation is the need for simplified models of dynamical systems, which capture the main features of the original complex models such as stability. This need arises from limited accuracy, computational power, and storage capabilities. The main goal of the efforts made in the field of linear model (order) reduction has been aimed at creating such simplified (but adequately accurate) models that can be properly incorporated in higher level simulations in place of the original complex models.

This thesis is mainly focused on the application of the MOR techniques on electrical circuit simulation problems such as the ones arising in the current high-performance VLSI designs (see [60–62], but also work of others). For this purpose, the current chapter is dedicated to the subject of compact modeling of general linear time invariant (LTI) systems such as on-chip interconnects which are modeled as linear RLCM circuits. Due to the practical importance, providing ingenious solutions to reduce the complexity of resulting macromodels has been an active area of research within the last three decades. As a result,

a rich body of literature is available covering the linear MOR techniques [44, 63–66] for compact modeling and analysis of linear circuits. It reviews some fundamental concepts and techniques of model reduction for linear time invariant (LTI) circuits. Chapters 5 and 6 present numerical examples of the application of Linear MOR on the Linear circuits and address some associated issues.

3.1 Physical Properties of Linear Dynamical Systems

There are intrinsic properties for the physical linear systems such as causality, stability, and passivity. To obtain model from a reduction process whose behavior stays faithful to the original system, conservation of the main physical characteristics of original systems is necessary. In other words, the compact models should inherit the essential properties from the original systems, among which the following characteristics are important.

3.1.1 Stability of Linear Systems

Stability is regarded as one of the most important properties of dynamical systems. It deals with the *boundedness* properties and the *asymptotic behavior* (as $t \rightarrow \infty$) of the transient solution of a zero-input state-equation (1.13) with respect to initial condition disturbances. Hence, it is often crucial [67] that reduced-order models inherit the stability properties of the original system. Due to its application in circuit analysis, the stability of continuous-time (LTI) descriptor (DAE) systems(1.13) is explained. The equation (1.13) is repeated below for ease of reference,

$$\mathcal{L} : \begin{cases} \mathbf{E} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), & \text{with } \mathbf{x}(t_0) = \mathbf{x}_0, \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t). \end{cases} \quad (3.1a)$$

$$(3.1b)$$

The standard state-space equations (1.9) can be considered as a special case of descriptor systems (3.1), where $\mathbf{E}$ is invertible ($\text{rank}(\mathbf{E}) = n = \dim(\mathbf{A})$). Hence, the analogous results can be obtained for state-space systems.

3.1.1.1 Internal Stability

To study the internal stability of descriptor linear systems, one needs only to consider the system state-equations 1.9a or 3.1a under zero-input conditions (homogeneous form).

The first stability notion based on the boundedness of solutions of (3.1) is *uniform stability*. Because solutions are linear in the initial state (*cf.* 1.2), it is convenient to express the bound as a linear function of the initial state.

Definition 3.1 (Uniform Stability [58]). The linear descriptor equations (3.1) is called *uniformly stable* if there exists a finite positive constant γ such that for any t_0 and $\mathbf{x}_0$ the corresponding solution satisfies

$$\|\mathbf{x}(t)\| \leq \gamma \|\mathbf{x}_0\|, \quad t \leq t_0. \quad (3.2)$$

The adjective *uniform* in the definition refers precisely to the fact that σ must not depend on the choice of initial time, as illustrated in Fig. 3.1. This figure also depicts the applicability of the concept to a general class of linear systems including *time-varying* types, where the system matrices are time dependent, i.e. $\mathbf{A}(t)$ and $\mathbf{B}(t)$.

For the case of standard state-space systems (1.9), where an state-transition matrix can be explicitly defined as shown in (3.3) [5], it is natural to begin by characterizing the stability of the linear state equation (1.9a) in terms of bounds on the transition matrix.

$$\Phi(t, \tau) = e^{\mathbf{A}(t-\tau)}, \quad \forall t, \tau \quad (3.3)$$

Figure 3.1: Illustrates the *uniform stability*; uniformity implies the σ -bound is independent of t_0 .

Theorem 3.1. *The linear system is uniformly stable if and only if there exist a finite positive constant γ such that*

$$\|\Phi(t, \tau)\| \leq \gamma \quad (3.4)$$

for all t, τ such that $t \geq \tau$.

The behavior of the state-transient matrix as defined in (3.4), is totally determined by the eigenvalues of matrix A (poles). In regards to this, from the boundedness of the response characterized by uniform stability, the following conclusion is readily discernible.

Theorem 3.2 (Marginal stability [5]). *The causal system $\dot{x} = Ax$ is marginal stable if and only if all real parts of eigenvalues of A are nonpositive (zero or negative) and those with zero real parts are simple roots of the minimal polynomial of A .*

Marginal stability requires that, all pure imaginary eigenvalues of A to be simple and only occur in 1×1 blocks in the Jordan form of A . This ensures that, the state $x(t)$, resulting from any nonzero initial state, will remain bounded under zero-input conditions.

Next, a stability property for (3.1a) that addresses both boundedness and asymptotic behavior of solutions is considered. It implies uniform stability, and imposes an additional

requirement that all solutions approach zero exponentially as $t \rightarrow \infty$.

Definition 3.2 (Uniform Exponential Stability [16,58]). System (3.1a) is called *uniformly exponentially stable* if there exist finite positive constants α, β , such that for any t_0 and $\mathbf{x}_0$ the corresponding solution satisfies

$$\|\mathbf{x}(t)\| \leq \alpha e^{-\beta(t-t_0)} \|\mathbf{x}_0\|, \quad t \geq t_0 \quad (3.5)$$

where the scalar β is called the decay rate.

This is illustrated in Fig. 3.2. Similar to Fig.-3.1 this figure also infers the applicability of the concept to a general class of linear systems including *time-varying*. Definition 3.5 is

Figure 3.2: A decaying-exponential bound independent of t_0 .

equivalent to $\lim_{t \rightarrow +\infty} \mathbf{x}(t) = 0$, so called *asymptotically stable*. This implies that the state $\mathbf{x}(t)$ resulting from any nonzero initial state $\mathbf{x}(t_0)$, given enough time, will decay to zero state under a zero-input condition.

The asymptotic stability can also be examined with regards to the system poles through the following theorem.

Theorem 3.3 ([16, 68]). *The regular descriptor linear system (3.1) is asymptotically stable if and only if*

$$\lambda(\mathbf{E}, \mathbf{A}) \subset \mathbb{C}_- = \{\lambda_i \mid \lambda_i \in \mathbb{C}, \det(\lambda_i \mathbf{E} - \mathbf{A}) = 0, \Re(\lambda_i) < 0\}. \quad (3.6)$$

It is to be noted that, $\lambda(\mathbf{E}, \mathbf{A}) \subset \mathbb{C}_-$ stands for the field of finite poles with negative real parts.

The above facts are strongly desirable to be extended to standard state-space systems (1.9a), leading to the following conclusion.

Theorem 3.4 (Asymptotic stability [5]). *The causal state-space system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is asymptotically stable if and only if all the eigenvalues of $\mathbf{A}$ have strictly negative real parts, that is, $\mathbf{A}$ is **Hurwitz**.*

The celebrated Lyapunov stability theory is another suitable mean to analyze the stability of linear systems.

Theorem 3.5 (Lyapunov Stability [4, 5, 63]). *The causal system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is asymptotically stable, i.e. $\Re\{\lambda(\mathbf{A})\} < 0$ for all eigenvalues of $\mathbf{A}$, if and only if for any given positive-definite symmetric matrix $\mathbf{Q}$ there exists a positive-definite (symmetric) matrix $\mathbf{P}$ that satisfies*

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = -\mathbf{Q}. \quad (3.7)$$

A Lyapunov equation theory for the stability of descriptor linear systems has also been established in the descriptor linear systems literature (e.g. [69] and the references therein). It is referred to as the generalized Lyapunov matrix equations, and falls in the following form [68, 70],

Theorem 3.6 (Generalized Lyapunov Stability [68] [4,5,63]). *A causal descriptor system (3.1a) is regular and asymptotically stable if and only if for any given positive-definite symmetric matrix $\mathbf{Q}$ there exists a positive-definite (symmetric) matrix $\mathbf{P}$ that satisfies*

$$\mathbf{A}^T \mathbf{P} \mathbf{E} + \mathbf{E}^T \mathbf{P} \mathbf{A} = -\mathbf{E}^T \mathbf{Q} \mathbf{E}. \quad (3.8)$$

satisfying

$$\text{rank}(\mathbf{E}^T \mathbf{P} \mathbf{E}) = \text{rank}(\mathbf{P}) = r. \quad (3.9)$$

3.1.1.2 External Stability

A causal system is externally stable if a bounded input, $u(t) < M_1$ for $-\infty < -T \leq t < \infty$, produces a bounded output $y(t) < M_2$, $-T \leq t < \infty$. A necessary and sufficient condition for bounded-input and bounded-output (BIBO) stability is defined based on its impulse response $h(\cdot)$ as presented in the following theorem.

Theorem 3.7 (BIBO stability [4, 5, 63]). *A SISO system is bounded-input and bounded-output (BIBO) stable if and only if its impulse response $h(t)$ is absolutely integrable in $[0, \infty)$, i.e.*

$$\int_0^{\infty} |h(t)| dt \leq M < \infty, \quad (3.10)$$

for some constant M . In discussing external stability, we shall assume zero initial conditions.

3.1.2 Passivity of Linear Systems

Roughly speaking, passive systems are systems that do not generate energy internally. In other words, the energy dissipated in the system is never greater than the energy supplied

to it.

3.1.2.1 Hybrid Case (Admittance and Impedance)

According to the positive-real lemma, an asymptotically stable network is passive if its transfer-function matrix $\mathbf{H}(s)$ (in **admittance** or **impedance** form) is positive real. Strictly speaking, fulfillment of the conditions in the following theorem implies that the underlying state-space description is a representation of a passive system [10, 71–73].

Theorem 3.8. *An impedance/admittance matrix $\mathbf{H}(s)$ represents a passive linear system if and only if*

$$\text{Each element of } \mathbf{H}(s) \text{ is analytic in } \mathbb{C}_+ \text{ } (\forall s : \Re(s) > 0) , \quad (3.11a)$$

$$\overline{\mathbf{H}}(s) = \mathbf{H}(\bar{s}) \quad (3.11b)$$

$$\begin{aligned} \Phi(s) = \mathbf{H}^H(s) + \mathbf{H}(s) &\geq 0; \quad \forall s : \Re(s) > 0, \\ \text{where } \Phi(s) &\text{ is positive semi-definite.} \end{aligned} \quad (3.11c)$$

The second condition (3.11b) ensures that all the coefficients in numerator and denominator polynomials are real and hence, the associated impulse response is also real. The third condition (3.11c) is a generalization of the fact that “a passive one-port impedance/admittance must have a positive real part” to the matrix case for multiport systems. A matrix that fulfills these three conditions is said to be *positive real*. For the physical systems which are asymptotically stable, positive-realness can be equivalently investigated by checking one equation as shown in (3.12),

$$\Phi(s) = \mathbf{H}(s) + \mathbf{H}^T(\bar{s}) \geq 0 \quad \forall s : s \in \mathbb{C}_+. \quad (3.12)$$

3.1.2.2 Scattering Case (s-parameters)

For a scattering representation, the first and the second condition are still valid. Only, it should be noted that for scattering representations no poles are allowed on the imaginary axis [10]; accordingly, condition two must hold for $s \in \underline{\mathbb{C}}_+$. Hence, a scattering matrix transfer function $\mathbf{H}(s)$ represents a passive linear system if it fulfills the following conditions [10, 72]

$$\text{Each element of } \mathbf{H}(s) \text{ is analytic in } \underline{\mathbb{C}}_+ \text{ } (\forall s : \Re[s] \leq 0) , \quad (3.13a)$$

$$\overline{\mathbf{H}}(s) = \mathbf{H}(\overline{s}) , \quad (3.13b)$$

$$\Phi(s) = \mathbf{I} - \mathbf{H}^H(s) \mathbf{H}(s) \geq 0; \quad \forall s : s \in \underline{\mathbb{C}}_+ \quad (3.13c)$$

where $\Phi(s)$ is positive semi-definite.

A matrix fulfilling these three conditions is said to be *bounded real*. A matrix is positive semi-definite when all its spectrum (eigenvalues) are a non-negative values. To investigate this for (3.13c) we mathematically have

$$\lambda(\mathbf{I} - \mathbf{H}^H \mathbf{H}) = 1 - \lambda(\mathbf{H}^H \mathbf{H}) \geq 0 , \quad (3.14)$$

or equivalently

$$\lambda(\mathbf{H}^H \mathbf{H}) \leq 1 , \quad (3.15)$$

which directly implies,

$$\lambda_{max}(\mathbf{H}^H \mathbf{H}) \leq 1 . \quad (3.16)$$

Knowing $\lambda_i(\mathbf{H}^H \mathbf{H}) = \sigma_i^2(\mathbf{H})$ [74], from (3.16) it is

$$\sigma_{max}^2(\mathbf{H}) \leq 1 , \quad (3.17)$$

and hence,

$$\sigma_{\max}(\mathbf{H}) \leq 1. \quad (3.18)$$

Theorem 3.9 (The Courant-Fischer Theorem for Singular Values [75]). *Suppose $\mathbf{H} \in \mathbb{C}^{m \times n}$, $n \leq m$, has singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. Then for $k = 1, \dots, n$*

$$\sigma_k = \min_{\dim(S)=n-k+1} \max_{\substack{\mathbf{x} \in S \\ \mathbf{x} \neq \mathbf{0}}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \max_{\dim(S)=k} \min_{\substack{\mathbf{x} \in S \\ \mathbf{x} \neq \mathbf{0}}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}. \quad (3.19)$$

From the above theorem, it is straightforward to conclude the following

Corollary 3.1. *Given $\mathbf{H}(s) \in \mathbb{C}^{p \times p}$ for any $s \in \mathbb{C}_+$, it is*

$$\sigma_{\max}(\mathbf{H}(s)) = \max_{\mathbf{x} \in \mathbb{C} - \{0\}} \frac{\|\mathbf{H}(s)\mathbf{x}(s)\|_2}{\|\mathbf{x}(s)\|_2} \triangleq \|\mathbf{H}(s)\|_2, \quad \forall s : \Re(s) \geq 0. \quad (3.20)$$

Considering (3.20) from corollary 3.1, an alternative form for passivity condition is obtained as shown in (3.21)

$$\|\mathbf{H}(s)\|_2 \leq 1. \quad (3.21)$$

To conclude, it is highlighted that, stability and passivity are the physical properties of the original system that should be preserved in a model order reduction process. It is important because, stable but non-passive models may lead to unstable systems when connected to other passive components. On the other hand, a passive macromodel, when terminated with any arbitrary passive load, always guarantees the stability of the overall network.

3.2 Linear Order Reduction Algorithms

There are a number of methods that have been developed for model order reduction of large electrical networks. Among those, the followings are widely used and known to be

successful in accomplishing a compact model representation.

A. Polynomial approximations of the transfer functions:

- i) Explicit moment matching techniques and asymptotic waveform evaluation (AWE)

B. Projection based techniques:

- i) Krylov-subspace methods
- ii) SVD based methods
 - Truncated balanced state-space representation (TBR)
 - Proper Orthogonal Decomposition (POD)

C. Non-projection based MOR methods:

- i) Hankel optimal model reduction
- ii) Singular perturbation
- iii) Transfer function fitting methods
- iv) Model reduction via convex optimization

3.3 Polynomial Approximations of Transfer Functions

The transfer function of a linear multiport network $\mathbf{H}(s)$ is a complex-valued matrix function. It linearly relates input to output at each complex-frequency point as $\mathbf{Y}(s) = \mathbf{H}(s) \mathbf{U}(s)$ when the initial condition is zero. For SISO systems, $H(s)$ is a complex-valued scalar function, that is defined as the following ratio,

$$H(s) = \frac{Y(s)}{U(s)}. \quad (3.22)$$

Let (3.22) be any asymptotically stable transfer function whose singularities fall in $\mathbb{C}_-$. It can be expanded using Taylor series at any frequency point s_0 ($\in \mathbb{C}_+$).

$$H(s) = \sum_{n=0}^{\infty} m_n (s - s_0)^n, \quad (3.23)$$

where

$$m_n = \frac{1}{n!} \times \left. \frac{d_n H(s)}{d s^n} \right|_{s=s_0}. \quad (3.24)$$

Expansion at origin of the complex plane can also be considered as

$$H(s) = \sum_{n=0}^{\infty} m_n s^n, \quad \text{where } m_n = \frac{1}{n!} \times \left. \frac{d_n H(s)}{d s^n} \right|_{s=0}. \quad (3.25)$$

As defined in (1.14), system *transfer function* $H(s)$ is obtained from Laplace transformation of the systems impulse response as

$$\begin{aligned} H(s) &\triangleq \int_0^{\infty} h(t) e^{-st} dt = \int_0^{\infty} h(t) \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} t^n \right) dt = \\ &\sum_{n=0}^{\infty} \left(\frac{(-1)^n}{n!} \int_0^{\infty} t^n h(t) dt \right) s^n. \end{aligned} \quad (3.26)$$

Hence, comparing (3.25) and (3.26), the coefficient can be defined as shown in (3.27) and hence the name, moments.

$$m_n = \frac{(-1)^n}{n!} \int_0^{\infty} t^n h(t) dt. \quad (3.27)$$

For Electrical Networks: Laplace transformation can also applied to solve the linear DAE equations from MNA formulation (1.22) for the electrical network. The corresponding

equations in *Laplace Domain* is given by (c.f. Chapter 5):

$$\begin{cases} \mathbf{C}s\mathbf{X}(s) + \mathbf{G}\mathbf{X}(s) = \mathbf{B}\mathbf{U}(s) \\ \mathbf{Y}(s) = \mathbf{L}\mathbf{X}(s), \end{cases} \quad (3.28a)$$

$$(3.28b)$$

where $\mathbf{X}(s) \in \mathbb{C}^n$, $\mathbf{U}(s) \in \mathbb{C}^m$ and $\mathbf{I}(s) \in \mathbb{C}^p$. Combining (3.28a) and (3.28b), the output $\mathbf{Y}(s)$ is related to the input $\mathbf{U}(s)$ through a transfer function as

$$\mathbf{H}(s) = \mathbf{L} (\mathbf{G} + s\mathbf{C})^{-1} \mathbf{B}. \quad (3.29)$$

Assuming a regular system (defined in 1.1), let $s_0 \in \underline{\mathbb{C}}_+$ be a properly selected expansion point at which matrix pencil $(\mathbf{G} + s_0\mathbf{C})$ is nonsingular. From (3.29), it is equivalently

$$\mathbf{H}(s) = \mathbf{L} (\mathbf{I} + (s - s_0)\mathbf{A})^{-1} \mathbf{R}, \quad (3.30)$$

where

$$\mathbf{A} \triangleq (\mathbf{G} + s_0\mathbf{C})^{-1} \mathbf{C} \quad (3.31)$$

and

$$\mathbf{R} \triangleq (\mathbf{G} + s_0\mathbf{C})^{-1} \mathbf{B}. \quad (3.32)$$

The j -th moment of the function at s_0 is defined as

$$\mathbf{M}_j(s_0) = \mathbf{L}\mathbf{M}_j(s_0) = \mathbf{L} \mathbf{A}^j \mathbf{R}, \quad \text{for } j = \{0, 1, 2, \dots\}. \quad (3.33)$$

3.3.1 AWE Based on Explicit Moment Matching

Asymptotic waveform evaluation (AWE) algorithm was first proposed in [76], where explicit moment matching was used based on Padé approximation to obtain a reduced order

rational function, sharing the few (e.g. m) leading moments with the original system. In AWE, the Padé approximant is obtained via explicit computation of the first $2m$ moments of $H(s)$ [39, 77, 78]. The AWE method shows that the wildly popular interconnect delay model, the Elmore delay [79], is just the first order of moments of a circuit. We have to keep in mind that, (a) accuracy cannot be guaranteed in the whole domain and (b) the AWE method is numerically unstable for higher-order moment approximation. [39] introduced some remedial methods to overcome this problem by frequency shifting and expanding around $s = \infty$. A more effective method introduced in [80]. It is based on (a) carrying out the multiple-point expansions along the imaginary axis (called frequency hopping) and (b) combining the expansion results which takes higher computational costs.

3.4 Projection-Based Methods

In Sec. 3.3.1, we tried to find a direct approximation of the transfer function by explicitly matching m leading moments. However, explicit moment matching approaches (namely AWE) suffer from numerical ill-conditioning in their equations. A more elegant solution to the numerical problem of AWE is to use projection based model order reduction methods, which are based on implicit moment matching [65]. There are several numerically stable methods based on the projection on subspaces and implicit moment matching [81–85]. The idea is to first reduce the number of state variables by projecting the vector of state variables $\mathbf{X}$ on a subspace spanned by the column vectors of an orthogonal matrix $\mathbf{Q}$ whose dimension $m \ll n$, where n is the original order. Let there exist z variables in reduced space such that it can be projected back to the original space with a minimal error $\zeta(t)$, as

$$\mathbf{x}(t) = \mathbf{Q} \mathbf{z}(t) + \zeta(t), \quad (3.34)$$

where at any time, $\mathbf{x}(\cdot) \in \mathbb{R}^n$ and $\mathbf{Q}\mathbf{z}(\cdot) \in \text{colsp}(\mathbf{Q})$. It is desirable for the error vector $\zeta(t)$ to not have any component in the reduced subspace, i.e. $\text{colsp}(\mathbf{Q})$. This requires confining the error vector $\zeta(t)$ to the *orthogonal complement subspace* of $\text{colsp}(\mathbf{Q})$, i.e. $\zeta(\cdot) \in \text{colsp}(\mathbf{Q})^\perp$, where

$$\text{colsp}(\mathbf{Q})^\perp \triangleq \{ \mathbf{w} \mid \mathbf{w} \in \mathbb{R}^n, \mathbf{w}^T \mathbf{v} = 0, \forall \mathbf{v} \in \text{colsp}(\mathbf{Q}) \}. \quad (3.35)$$

It is straightforward to mathematically establish that, the orthogonal complement of the column-space of $\mathbf{Q} \in \mathbb{R}^m$ is the null-space of $\mathbf{Q}^T$. It is a set of all possible solution vectors for $\mathbf{Q}^T \zeta(\cdot) = 0$.

Substituting $\mathbf{x}(t)$ from (3.34) in (1.22a), we get

$$\mathbf{C} \frac{d}{dt} (\mathbf{Q}\mathbf{z}(t) + \zeta(t)) + \mathbf{G}(\mathbf{Q}\mathbf{z}(t) + \zeta(t)) - \mathbf{B}\mathbf{u}(t) = 0 \quad (3.36)$$

and

$$\mathbf{C} \frac{d}{dt} \mathbf{Q}\mathbf{z}(t) + \mathbf{G}\mathbf{Q}\mathbf{z}(t) - \mathbf{B}\mathbf{u}(t) = \mathbf{C} \frac{d}{dt} \zeta(t) + \mathbf{G}\zeta(t). \quad (3.37)$$

Considering $\mathbf{x}(t) \approx \tilde{\mathbf{x}} = \mathbf{Q}\mathbf{z}$, using this approximated solution $\tilde{\mathbf{x}}$ leads to a *residual error* as

$$\mathcal{R}(\zeta(t)) \triangleq \mathbf{C} \frac{d}{dt} \mathbf{Q}\mathbf{z}(t) + \mathbf{G}\mathbf{Q}\mathbf{z}(t) - \mathbf{B}\mathbf{u}(t) = \mathbf{C} \frac{d}{dt} \zeta(t) + \mathbf{G}\zeta(t). \quad (3.38)$$

Multiplying both sides of (3.38) by $\mathbf{Q}^T$ and using the orthogonality property as $\mathbf{Q}^T \zeta(\cdot) = 0$, we get

$$\begin{aligned} \mathbf{Q}^T \mathcal{R}(\zeta) &= (\mathbf{Q}^T \mathbf{C} \mathbf{Q}) \frac{d}{dt} \mathbf{z}(t) + (\mathbf{Q}^T \mathbf{G} \mathbf{Q}) \mathbf{z}(t) - (\mathbf{Q}^T \mathbf{B}) \mathbf{u}(t) = \\ &= \mathbf{C} \frac{d}{dt} \mathbf{Q}^T \zeta(t) + \mathbf{G} \mathbf{Q}^T \zeta(t) = 0, \end{aligned} \quad (3.39)$$

and hence,

$$(\mathbf{Q}^T \mathbf{C} \mathbf{Q}) \frac{d}{dt} \mathbf{z}(t) + (\mathbf{Q}^T \mathbf{G} \mathbf{Q}) \mathbf{z}(t) - (\mathbf{Q}^T \mathbf{B}) \mathbf{u}(t) = 0. \quad (3.40)$$

For this purpose, the idea is simply to make the residual error in the differential equation (1.22a) small when the approximated solution obtained from the reduced model is used. This is achieved by making the error vector in solution “orthogonal” to the subspace spanned by the column vectors in $\mathbf{Q}$ (the subspace of the solution $\mathbf{z}$). This is the so-called *Petrov-Galerkin* [86, pp. 9] scheme in solving differential equations.

Next, the approximated output is obtained as

$$\mathbf{y}(t) = \mathbf{L}^T \mathbf{Q} \hat{\mathbf{x}}(t). \quad (3.41)$$

For the resulting reduced set of differential equations in (3.40) (*reduced order model*) and its associated output equation (3.41)), the reduced MNA matrices are

$$\begin{aligned} \hat{\mathbf{C}} &= \mathbf{Q}^T \mathbf{C} \mathbf{Q}, & \hat{\mathbf{G}} &= \mathbf{Q}^T \mathbf{G} \mathbf{Q}, \\ \hat{\mathbf{B}} &= \mathbf{Q}^T \mathbf{B}, \text{ and } & \hat{\mathbf{L}} &= \mathbf{L} \mathbf{Q}. \end{aligned} \quad (3.42)$$

The next step in any subspace projection techniques is to find a proper choice for orthogonal basis to span the reduced projection space as $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m] \in \mathbb{R}^{n \times m}$ such that, having $\mathbf{Z} \in \text{colsp}(\mathbf{Q})$ a properly accurate approximation as $\mathbf{X} \approx \mathbf{Q} \mathbf{Z}$ can be derived. It is possible to use eigenvectors when diagonalization of the dynamic matrix in minimal state-space representation is possible. Another approach could be to compute the basis using time series data from all states of systems (in POD [41], [44, Chap.5], [66, Chap.10], [87]). Alternatively, one may try balancing the system’s controllability and observability Gramians (in TBR [88]) by using singular value decomposition (SVD) to choose the $\mathbf{Q}$.

Among all the possibilities, the use of Krylov-subspaces is also worth studying.

3.4.1 General Krylov-Subspace Methods

The most successful algorithms for reduction of large-scale linear systems have been projection-based approaches [64, 89, 90]. Among all, the (block) Krylov subspace based projection method [64, 82, 84, 90, 91] is the most commonly used method in model order reduction. The orthogonal projection matrix that maps the n -dimensional state space of original system into a m -dimensional subspace is constructed as follows:

$$\begin{aligned} \text{colsp}\{\mathbf{Q}\} &= \mathcal{K}_m(\mathbf{A}, \mathbf{R}) \\ &= \text{span}\{\mathbf{R}, \mathbf{A}\mathbf{R}, \dots, \mathbf{A}^{m-1}\mathbf{R}\}, \end{aligned} \quad (3.43)$$

for $\mathbf{A}$ and $\mathbf{R}$ defined in (3.31) and (3.32). A linear system of much smaller order is derived by a variable change as $\mathbf{x} = \mathbf{Q}\mathbf{z}$ and multiplying $\mathbf{Q}^T$ on both sides of the differential matrix equations of reduced variables. For the moment preservation properties for the Krylov-subspace based methods [90, Theorem-31, pp.36] can be consulted.

3.4.1.1 Arnoldi Algorithm

The Arnoldi method is the classic method to find a set of orthonormal vectors as a basis for a given Krylov subspace in (3.43). The Arnoldi process was originally introduced in the field of mathematics by W.E. Arnoldi in 1951 [92]. The Arnoldi algorithm for the applications in model order reduction of RLC network was introduced in [93]. Given Krylov subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{R})$, the Arnoldi method using the “*modified Gram-Schmidt*” orthogonalization [94] calculates the columns of projection matrices. The Arnoldi algorithms is known to be a numerically efficient iterative method. For the practical implementations of Arnoldi (SISO) and Block-Arnoldi (MIMO) algorithms, one can refer to [47, 64, 65, 94].

3.4.1.2 Padé via Lanczos (PVL)

Although the Krylov subspace method (using Arnoldi process) possibly is the widely employed one, the Padé via Lanczos (PVL) method was the first projection-based method [81]. The classical Lanczos process [95] is an iterative procedure for the successive reduction of any square matrix to a sequence of tridiagonal matrices. It is a numerically stable method to compute eigenvalues of generalizable matrices. Lanczos is employed to compute the Krylov subspace in PVL in the sense of an oblique projection. Later it was proved that the reduced system implicitly matches the original system to a certain order of moments [96].

Later on, the multiport Padé via Lanczos (MPVL) algorithm [97] was developed which is an extension of PVL to general multiple-input multiple-output systems. The MPVL algorithm computes a matrix Padé approximation to the matrix-valued transfer function of the multiple-input multiple-output system, using Lanczos-type algorithm for multiple starting vectors [98].

The PVL method was also extended to deal with circuits with symmetric matrices by means of the SyPVL algorithm [99]. Similarly, its multiport counterpart (SyMPVL) was introduced in [100].

3.4.1.3 PRIMA

On the idea of utilizing the Krylov subspace, PRIMA (*passive reduced-order interconnect macromodeling algorithm*) was introduced in [85] as a direct extension of the block Arnoldi technique. PRIMA guarantees passivity in the resulting reduced model. The attraction of the method is mainly due to its passivity preservation which is promised with simple formulation. To shed more light on the method, let (1.22) be the time-domain MNA equations for a given linear RLC dynamic (MIMO) system, where in susceptance matrix $\mathbf{C}$ and conductance matrix $\mathbf{G}$ the rows corresponding to the current variables are negated [85]. Also,

for a p -port system (i.e. $m = p$) of size n , $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the column vectors, including the voltage sources and the output currents at p ports, respectively. The input matrix $\mathbf{B}$ is a selector matrix consisting of “1”s, “-1”s and “0”s and is related to the output selector matrix $\mathbf{L}^T = \mathbf{B}$. By exploiting an orthogonal projection matrix $\mathbf{Q}$, defined by the basis from Krylov subspace, a change of variable $\mathbf{z} = \mathbf{Q}^T \mathbf{x}$ is applied in (1.22) to find a reduced-order model based on a congruence transformation as:

$$\begin{cases} \hat{\mathbf{C}} \frac{d}{dt} \mathbf{z}(t) + \hat{\mathbf{G}} \mathbf{z}(t) = \hat{\mathbf{B}} \mathbf{u}(t), \\ \hat{\mathbf{y}}(t) = \hat{\mathbf{L}} \mathbf{z}(t). \end{cases} \quad (3.44a)$$

$$(3.44b)$$

The reduced model while preserving the main properties of the original system provides an output $\hat{\mathbf{y}}(t)$ that appropriately approximates the original response $\mathbf{y}(t)$. For the resulting macromodel in (3.44), the reduced MNA matrices are

$$\begin{aligned} \hat{\mathbf{C}} &= \mathbf{Q}^T \mathbf{C} \mathbf{Q}, & \hat{\mathbf{G}} &= \mathbf{Q}^T \mathbf{G} \mathbf{Q}, \\ \hat{\mathbf{B}} &= \mathbf{Q}^T \mathbf{B}, \text{ and } & \hat{\mathbf{L}} &= \mathbf{L} \mathbf{Q}. \end{aligned} \quad (3.45)$$

It is proved that the reduced system (3.44) of order q preserves the first $\left\lfloor \frac{q}{p} \right\rfloor$ block moments of the original network (1.22) [64, 85]. This implies that for a desired predefined accuracy, the order of the reduced system should be increased with the increase in the number of ports p .

3.4.2 Truncated Balance Realization (TBR)

One of the alternative methods for model order reduction of LTI systems is by means of control-theoretical-based truncated balance realization (TBR) methods [65, 88, 101–116]. In TBR which is a SVD-based approach, the weak uncontrollable and unobservable state

variables are truncated to achieve the reduced models for linear VLSI systems. For a formal definition of controllability and observability any references in the area of linear system theory can be fruitfully consulted with [5, Chapter 6], [57, Chapter 3], and [58, Chapter 9].

3.4.2.1 Standard / Conventional TBR

The TBR procedure is centered around information obtained from the controllability Grammian $\mathbf{W}_c$ and the observability Grammian $\mathbf{W}_o$. These two Gramians are Hermitian positive definite matrices that can be uniquely [88, 101] obtained from solving the following Lyapunov equations. Given a state-space model in descriptor form as shown in (3.1), let $\mathbf{E} = \mathbf{I}$. This is as a matter of convenience, while formulation for singular $\mathbf{E}$ is also possible (*cf.* generalized Lyapunov equation (3.8) in Theorem-3.6).

$$\mathbf{A} \mathbf{W}_c + \mathbf{W}_c \mathbf{A}^T = -\mathbf{B} \mathbf{B}^T \quad (3.46a)$$

$$\mathbf{A}^T \mathbf{W}_o + \mathbf{W}_o \mathbf{A} = -\mathbf{C} \mathbf{C}^T \quad (3.46b)$$

The eigenvalues $\lambda(\mathbf{W}_c \mathbf{W}_o)$ are called the *Hankel singular values*. In particular “small” Hankel singular values correspond to internal dynamic modes that have a weak effect on the input-output behavior of the system and are therefore, close to non-observable or non-controllable or both [65].

A complete TBR algorithm [103] is shown as Algorithm 1.

Balancing transformation matrix $\mathbf{T}$ is obtained in step-5 (of Algorithm 1). Under a similarity transformation, as shown in step-6 a balanced system is obtained whose both Gramians become equal and diagonal as

$$\mathbf{W}_c = \mathbf{W}_o = \mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n), \quad \text{where } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n. \quad (3.47)$$

Algorithm 1: Standard TBR Algorithm.

-
- input** : Original Model ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$)
output: Reduced Macromodel ($\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$)
- 1 Solve $\mathbf{A} \mathbf{W}_c + \mathbf{W}_c \mathbf{A}^T = -\mathbf{B} \mathbf{B}^T$ for $\mathbf{W}_c$;
 - 2 Solve $\mathbf{A}^T \mathbf{W}_o + \mathbf{W}_o \mathbf{A} = -\mathbf{C} \mathbf{C}^T$ for $\mathbf{W}_o$;
 - 3 Compute Cholesky factors $\mathbf{W}_c = \mathbf{L}_c \mathbf{L}_c^T$ and $\mathbf{W}_o = \mathbf{L}_o \mathbf{L}_o^T$;
 - 4 Compute **SVD** of Cholesky factors $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{L}_o^T \mathbf{L}$, where $\mathbf{\Sigma}$ is diagonal positive and $\mathbf{U}$, $\mathbf{V}$ have orthonormal columns;
 - 5 Compute the balancing transformation matrices
 $\mathbf{T} = \mathbf{L}_c \mathbf{V} \mathbf{\Sigma}^{-1/2}$, $\mathbf{T}^{-1} = \mathbf{\Sigma}^{-1/2} \mathbf{U}^T \mathbf{L}_o^T$;
 - 6 Form the balanced realization as $\tilde{\mathbf{A}} = \mathbf{T}^{-1} \mathbf{A} \mathbf{T}$, $\tilde{\mathbf{B}} = \mathbf{T}^{-1} \mathbf{B}$, $\tilde{\mathbf{C}} = \mathbf{C} \mathbf{T}$;
 - 7 Select reduced model order and partition $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ conformally;
 - 8 Truncate $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ to form the reduced realization $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$ and it is $\hat{\mathbf{D}} = \mathbf{D}$;
-

One may partition $\mathbf{\Sigma}$ into

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix}, \quad (3.48)$$

where the singular values $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \dots, \sigma_m)$ and $\mathbf{\Sigma}_2 = \text{diag}(\sigma_{m+1}, \dots, \sigma_n)$. It is seen that, $\mathbf{\Sigma}_1$ corresponds to the “strong” sub-systems to be retained and $\mathbf{\Sigma}_2$ the “weak” sub-systems to be deleted [111]. Conformally partitioning the transformed matrices as

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{bmatrix}, \quad \tilde{\mathbf{C}}_1 = \begin{bmatrix} \tilde{\mathbf{C}}_1 & \tilde{\mathbf{C}}_2 \end{bmatrix} \quad (3.49)$$

Hence, the reduced model is defined as

$$\hat{\mathbf{A}} = \tilde{\mathbf{A}}_{11}, \quad \hat{\mathbf{B}} = \tilde{\mathbf{B}}_1, \quad \hat{\mathbf{C}} = \tilde{\mathbf{C}}_1. \quad (3.50)$$

Error Bounds

One of the attractive aspects of TBR methods is that computable error bounds are available for the reduced model. This bounded model reduction error is the prominent characteristic for TBR in comparison to the projection methods based on the (explicit) moment matching, namely Krylov subspace methods. The error in the transfer function of the order-reduced model is bounded by [101, 102]:

$$\sigma_m \leq \left\| H(s) - \hat{H}(s) \right\|_{\infty} \leq 2 \sum_{i=m+1}^n \sigma_i. \quad (3.51)$$

Computational Complexity

For a macromodel of order m ($m \ll n$), the CPU cost is analyzed in the Table 3.1.

Table 3.1: Time complexities of standard TBR.

Operation	Cost
<i>Computation of the Gramians $\mathbf{W}_c$ and $\mathbf{W}_o$ by solving Lyapunov equations in steps 1-2</i>	$\mathcal{O}(n^3)$
<i>Two Cholesky decomposition in Step-3* (for sparse equations)</i>	$2 \mathcal{O}(n^\beta)$ (typically, $1.1 \leq \beta \leq 1.5$)
<i>SVD in Step-4 to compute m leading singular values*</i>	$\mathcal{O}(n m^2)$
<i>Linear matrix solving tasks (for sparse equations)</i>	$\mathcal{O}(n^\alpha)$ (typically, $1.1 \leq \alpha \leq 1.2$)
<i>Forming transformation matrices in step-5</i>	$\mathcal{O}(n m^2)$
<i>Similarity transformation in step-6</i>	$\mathcal{O}(n m)$
Total cost: $\mathcal{O}(n^3 + n^\beta + n m^2 + n m)$	

* Based on the algorithms in [117].

It is seen that, the computational cost for solving Lyapunov equations $\mathcal{O}(n^3)$ is dominant.

Hence, the bottleneck in balanced truncation methods is the computational complexity for solving the Lyapunov equations. There are both direct and iterative ways to solve Lyapunov equations. Specifically, the efficient numerical solvers in [108, 109, 118] and newly developed method in [119] can be named. They are based on the *alternated direction implicit (ADI)* method [120, 121]. Despite all the advancement in solution techniques, the complexity cost for solving Lyapunov equations is still noticeably high, prohibiting the TBR method from reducing large systems.

It is generally remarked that, the TBR methods can produce nearly optimal models but they are more computationally expensive than projection-based methods.

3.4.2.2 Passive Truncated Balance Realization

To preserve passivity in TBR the following two cases should be considered.

Positive Real TBR (PR-TBR)

The positive-real lemma, states that $\mathbf{H}(s)$ is positive real if and only if there exist matrices $\mathbf{X}_c = \mathbf{X}_c^T \geq 0$, $\mathbf{J}_c$, $\mathbf{K}_c$ as well as $\mathbf{X}_o = \mathbf{X}_o^T \geq 0$, $\mathbf{J}_o$, $\mathbf{K}_o$ such that the following two sets of *Lur'e equations* are satisfied [122].

$$\mathbf{A}\mathbf{X}_c + \mathbf{X}_c\mathbf{A}^T = -\mathbf{K}_c\mathbf{K}_c^T, \quad (3.52a)$$

$$\mathbf{X}_c\mathbf{C}^T - \mathbf{B} = -\mathbf{K}_c\mathbf{J}_c^T, \quad (3.52b)$$

$$\mathbf{J}_c\mathbf{J}_c^T = \mathbf{D} + \mathbf{D}^T, \quad (3.52c)$$

and its dual set,

$$\mathbf{A}^T\mathbf{X}_o + \mathbf{X}_o\mathbf{A} = -\mathbf{K}_o^T\mathbf{K}_o, \quad (3.53a)$$

$$\mathbf{X}_o\mathbf{B} - \mathbf{C}^T = -\mathbf{K}_o^T\mathbf{J}, \quad (3.53b)$$

$$\mathbf{J}_o^T\mathbf{J}_o = \mathbf{D} + \mathbf{D}^T. \quad (3.53c)$$

Matrices $\mathbf{X}_c$ and $\mathbf{X}_o$ are analogous to the controllability and observability Gramians, respectively.

Algebraic Riccati Equations (ARE)

By combining (3.52a), (3.52a) and (3.52a) and similarly by combining (3.53a), (3.53a) and (3.53a), the following two equations (3.54a) and (3.54b) are respectively obtained. They are so-called, algebraic Riccati equation (ARE) [123].

$$\mathbf{A} \mathbf{X}_c + \mathbf{X}_c \mathbf{A}^T + (\mathbf{B} - \mathbf{X}_c \mathbf{C}^T)(\mathbf{D} + \mathbf{D}^T)^{-1}(\mathbf{B} - \mathbf{X}_c \mathbf{C}^T)^T = \mathbf{0}, \quad (3.54a)$$

$$\mathbf{A}^T \mathbf{X}_o + \mathbf{X}_o \mathbf{A} + (\mathbf{C}^T - \mathbf{X}_o \mathbf{B})(\mathbf{D} + \mathbf{D}^T)^{-1}(\mathbf{C}^T - \mathbf{X}_o \mathbf{B})^T = \mathbf{0}, \quad (3.54b)$$

The algorithm for positive-real TBR [111, 112] is shown as Algorithm 2.

Similar to the standard TBR method, this method also has error bounds.

Algorithm 2: Positive -Real TBR (PR-TBR) Algorithm.

input : Original Model ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$)

output: Reduced Macromodel ($\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$)

- 1 Solve set of equations (3.52) (or equivalently (3.54a)) for $\mathbf{X}_c$;
 - 2 Solve set of equations (3.53) (or equivalently (3.54b)) for $\mathbf{X}_o$;
 - 3 Proceed with steps 3-8 in Algorithm 1, substituting $\mathbf{X}_c$ for $\mathbf{W}_c$ and $\mathbf{X}_o$ for $\mathbf{W}_o$;
-

Bounded Real TBR (BR-TBR)

As previously explained, passivity for the systems identified by their s -parameters representation is ensured by preserving the bounded realness in the order reduction process. To guarantee the bounded realness of the reduced model $\hat{\mathbf{H}}(s)$ from TBR, a similar procedure to PR-TBR is followed. It mainly includes obtaining the two system Gramians by solving the two sets of modified Lur'e equations, so-called bounded-real equations. For further details, [111] can be consulted with.

3.4.2.3 Other Extensions of TBR

Spectrally Weighted Balanced Truncation (SBT)

Enns [102, 105] extended the TBR method to include frequency weightings. The resulting method is known as the *frequency weighted balanced truncation*. In this method, by using a chosen weighting function, the error in the reduced model can be minimized and bounded [106] in some frequency range of interest. With only one weighting present, the stability of reduced-order models is guaranteed. However, in case of double-sided weightings, Enns' method may yield unstable models for stable original systems. Several modifications to Enns' technique have been proposed to overcome this shortcoming. Wang's technique [107, 124], in addition to guaranteeing stability in the case of double-sided weightings, also provide frequency response error bounds. Generalization of this technique to include passivity preservation was presented in [116].

Poor Man's TBR (2005)

An empirical TBR method, named *poor man's TBR*, was proposed to improve the scalability of the TBR methods, which shows the connection with the generalized projection-based reduction methods [113, 114].

3.4.3 Proper Orthogonal Decomposition (POD) Methods

Proper orthogonal decomposition (POD), also known as Karhunen-Loève decomposition [125] or principal component analysis (PCA) [126], provides a technique for analyzing multidimensional data. It is also a method that derives reduced models by linear projection. This method essentially constructs an orthonormal projection basis from the snapshots of the state vectors at N different time-points to form a data matrix as

$\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_N)] \in \mathbb{R}^{n \times N}$, where generally $N \ll n$. These time samples of states are obtained during a transient simulation of the system with some input excitation [127]. After obtaining the projection matrix from POD approaches, it is used to generate a reduced model via a standard projection scheme. More details on the time-domain POD and its formulation are presented in the next chapter.

3.4.3.1 Frequency-Domain POD

For linear (LTI) systems, the POD basis can be obtained efficiently by taking advantage of linearity and the frequency domain. The *frequency-domain POD* methods [87, 128–131] have been developed to obtain reduced model using this frequency domain data. Hence, in place of transient response, the snapshots of frequency response corresponding to some frequency points of interest are used. For the problems that applying appropriate time simulation to obtain snapshots faces difficulties, this frequency-domain approach is well appreciated. The POD snapshots can, therefore, be obtained by choosing a set of sample frequencies $\{\omega_i\}$ based on the frequency contents of the problems of interest and solving the frequency-domain system to obtain the responses $\mathbf{X}(\omega_i) = (j\omega_i \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{B}$. The resulting complex response can be used in a frequency-domain POD analysis as in [128], or the real and imaginary part of each complex response can be used similar to the snapshots in a time domain POD analysis as in [131].

The general idea in frequency-domain POD methods is outlined in Algorithm 3.

The benefits of POD methods are the followings.

- The time domain samples $\mathbf{X}_t = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)]$ are easy to obtain using existing numerical solvers for a system (linear or nonlinear). Extracting the frequency domain response is also trivially possible using existing solvers. In both cases, one can take advantage of the sparsity of system matrices and fast solvers.

Algorithm 3: An Outline of Frequency-Domain POD Algorithm.

-
- input** : Original Model $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$
output: Reduced Macromodel $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}})$
- 1 Select the real frequency points of interest $\omega_k, k = 1, \dots, N$;
 - 2 Compute the original system response at frequencies of interest $\mathbf{X}(\omega_i) = (j\omega_i \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{B}$ for $i = 1, 2, \dots, N$ and store them in an ensemble of complex snapshot, $\mathbf{X}_s = [\mathbf{X}(\omega_1), \mathbf{X}(\omega_2), \dots, \mathbf{X}(\omega_N)]$;
 - 3 Construct the correlation matrix $\mathbf{R} = \frac{1}{N} (\mathbf{X}_s^H \mathbf{X}_s)$, ($\in \mathbf{C}^{N \times N}$);
 - 4 Solve the matrix eigenvalue problem, $\mathbf{R} \mathbf{T}_i = \lambda_i \mathbf{T}_i$;
 - 5 Form the basis vectors as $\mathbf{v}_i = \frac{1}{N} \sum_{i=1}^N T_i \mathbf{X}(\omega_i)$;
 - 6 Construct the projection matrix, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$, ($m \leq N$), whose column vectors span the reduced subspace;
 - 7 Form reduced system's matrices, $\hat{\mathbf{A}} = \mathbf{V}^T \mathbf{A} \mathbf{V}$, $\hat{\mathbf{B}} = \mathbf{V}^T \mathbf{B}$, $\hat{\mathbf{C}} = \mathbf{C} \mathbf{V}$, and $\hat{\mathbf{D}} = \mathbf{D}$.
-

- Simple to implement
- In practice, it works quite reliably
- POD does not neglect the nonlinearities of the original vector-field. Indeed, it has a straightforward generalization for nonlinear systems (see the next chapter)

POD is general because it can take any trajectory of state variables. This advantage of POD is also its limitation. Because the POD basis is generated from the system response with a specific input, the reduced model is only guaranteed to be close to the original system when the input is close to the modeling input. For this purpose, the excitation signals should be carefully decided such that its frequency spectrum covers the major frequency range of interest for the intended application.

It is worth to remark that, despite the above objection, model reduction via POD is quite popular and is the method of choice in many fields, mainly due to its simplicity of implementation and promising accuracy.

3.5 Non-Projection Based MOR Methods

Non-projection methods do not employ construction of any projection matrices. The following are several most commonly used methods of this kind.

3.5.1 Hankel Optimal Model Reduction

The task of Hankel optimal model reduction of a matrix transfer function $\mathbf{H}(s)$ calls for finding a stable reduced system $\hat{\mathbf{H}}(s)$ of order less than a given positive integer m , such that the Hankel norm [101] $\|\zeta(s)\|_H$ of the absolute error $\zeta(s) = \mathbf{H}(s) - \hat{\mathbf{H}}(s)$ is minimal. Since Hankel operator $\mathcal{H}$ represents a “part” of the total LTI system with transfer matrix $\mathbf{H}(s)$, Hankel norm is never larger than $\mathcal{L}_\infty$ norm. Hence, Hankel optimal model reduction setup can be viewed as a relaxation of the “original” ($\mathcal{L}_\infty$ optimal) model reduction formulation. While no acceptable solution is available for the $\mathcal{L}_\infty$ case, Hankel optimal model reduction has an elegant and algorithmically efficient solution [101, 132–135].

3.5.2 Singular Perturbation

As it will be shown in Chapter 6, projection-based MOR methods can be interpreted as performing a coordinate transformation of the original system’s state space to a lower dimension subspace. For example, in TBR, such space transformation leads to a balanced system for which we “truncate” the system’s states. This mathematically can be seen as effectively setting the last $(n - m)$ states to zero. As an alternative, one can instead set the derivatives of the states to be discarded to zero. This procedure is called *state residualization*, which is the same as a *singular perturbation approximation* [48]. For more details and the mechanics of the method, [48, 63] can be consulted. For the basic properties of singular perturbation for balanced systems, [136] can be referred to.

3.5.3 Transfer Function Fitting Method

With the ever increasing operating frequencies, obtaining analytical models for high-speed modules becomes difficult and consequently, the characterization based on terminal responses (obtained through measurements or EM simulations) becomes increasingly popular [137]. Beside time-domain characterization, linear devices and subsystems can also be characterized in frequency domain, which is usually more feasible in applications. Such approaches demand development of fast and accurate physical system identification algorithms so as to embed the resulting model in a transient simulation environment. In order to make time-domain simulations feasible, one can construct a state-space model that approximates the sampled transfer function of the system. Such methods can be treated as model reduction, as starting from a characterization of the original system, we obtain a minimal model as an approximant.

The following methods fall into this category.

3.5.3.1 Rational Fitting Methods

For linear time invariant (LTI) systems, these methods are used to find polynomial coefficients of the numerator and the denominator of approximate rational transfer functions through iterative application of linear least squares [138].

3.5.3.2 Vector Fitting Methods

Several algorithms have been developed in the recent years for physical system identification of networks characterized by tabulated data. One of the popular techniques among these is the Vector Fitting (VF) algorithm. The vector fitting method was originally introduced in [139] and its extension in [140] to work with frequency domain data. Later, the method was extended in [141–143] to utilize the time domain data directly. Recently,

in [144, 145] the z -domain vector-fitting algorithms have also been developed. All these methods are to obtain reduced order (and generally minimal) models for the physical systems based on the iterative application of Linear Least Squares (LLS), where unknowns are systems' poles and residues. Starting with an initial guess of poles, an accurate set of poles is computed by fitting a scaled function through pole relocation iterations. After the poles have been identified, the residues of the transfer function are finally calculated by solving the corresponding least squares problem with known poles. Although there is no convergence proof for these methods, they usually work well in practice.

3.5.3.3 Original Formulation

In the original formulation of vector fitting [139], the objective is to approximate a given frequency response $H(s)$ with a rational function

$$H(s) \approx \sum_{n=1}^N \frac{k_n}{s - p_n} + d + hs \quad (3.55)$$

where terms d and h are optional. The vector fitting first identifies the poles of $H(s)$ by solving the following linear problem in the least squares sense

$$\delta(s)H(s) = \hat{H}(s); \quad (3.56)$$

with

$$\delta(s) = \sum_{n=1}^N \frac{r_n}{s - a_n} + 1 \quad (3.57)$$

and

$$\hat{H}(s) = \sum_{n=1}^N \frac{\hat{r}_n}{s - a_n} + \hat{d} + \hat{h}s \quad (3.58)$$

where $\{a_n\}$ is a set of initial poles, all poles and residues in (3.57) and (3.58) are real or come in complex conjugate pairs while $\hat{d}$ and $\hat{h}$ are real. The relocated (improved) poles are equal to the zeros of $\delta(s)$ and are obtained as

$$\{a_n\} = \text{eig}(\mathbf{A} - \mathbf{B}\mathbf{C}^T) \quad (3.59)$$

where $\mathbf{A}$ is a diagonal matrix holding the poles $\{a_n\}$, $\mathbf{B}$ is a column vector of ones, and $\mathbf{C}$ is a column vector holding the residues r_n .

This procedure can be applied in an iterative manner where (3.56)-(3.59) are solved repeatedly with the new poles from (3.59). After the poles have been identified, the residues of (3.55) are finally calculated by solving the corresponding least squares problem with known poles [139, 146].

Later (2006), a modification of the VF formulation was introduced in [147, 148], which improves the ability of VF to relocate poles to better positions, thereby improving its convergence performance and reducing the importance of the initial pole set. This is achieved by replacing the high-frequency asymptotic requirement of the vector fitting scaling function (3.57) with a more relaxed condition as

$$\delta(s) = \sum_{n=1}^N \frac{r_n}{s - a_n} + \tilde{d}. \quad (3.60)$$

Another noteworthy improvement, to speed-up the vector fitting for multiport systems using a common set of poles was introduced in [149]. This is achieved by applying the QR decomposition to the LS equations formed in each iteration of VF.

3.5.3.4 Generalized Formulation

In general, the frequency domain VF methods are used to find rational transfer function

$$H(s) = \frac{N(s)}{D(s)} = \frac{\sum_{k=1}^N N_k \Phi_k(s)}{\sum_{r=1}^D D_r \Phi_r(s)} \quad s = j2\pi f \quad (3.61)$$

which approximates the spectral response of a system over some predefined frequency range of interest $[f_{min}, f_{max}]$. This reduces to finding the real-valued coefficients N_k , D_r and the poles for basis functions $\Phi_i(s)$, where the numbers N and D represent the order of numerator and denominator, respectively.

Rational least-squares approximation is essentially a nonlinear problem, and corresponds to minimizing the following cost function [140, 150, 151]

$$\arg \min_{N_k, D_r} \sum_{l=0}^{N_{freq}} \left| H(s_l) - \frac{N(s_l)}{D(s_l)} \right|^2 = \arg \min_{N_k, D_r} \sum_{l=0}^{N_{freq}} \left(\frac{1}{|D(s_l)|^2} |D(s_l)H(s_l) - N(s_l)|^2 \right). \quad (3.62)$$

By taking Levi's approach [152], the problem is simplified to minimizing the summation of the squared weighted error as shown below, that is nonquadratic in the system parameters

$$\arg \min_{N_k, D_r} \sum_{l=0}^{N_{freq}} |D(s_l)H(s_l) - N(s_l)|^2. \quad (3.63)$$

3.5.3.5 Sanathanan-Koerner

Advocating the Sanathanan-Koerner interactive weighted LLS estimator [153], the model parameters at the t -th iteration are calculated by minimizing the weighted linear cost function

$$\arg \min_{N_k^{(t)}, D_r^{(t)}} \left(\sum_{l=0}^{N_{freq}} \frac{1}{|D^{(t-1)}(s_l)|^2} |D^{(t)}(s_l)H(s_l) - N^{(t)}(s_l)|^2 \right) \quad (3.64)$$

where for the first iteration $D^{(0)}(s) = 1$.

3.5.3.6 Basis Functions for Original VF

In the original VF (Sec. 3.5.3.3), to make sure that the transfer function has real-valued coefficients, the adopted partial fractions are used as basis functions $\Phi_i(s)$ in (3.61). These basis are obtained as a linear combination of the partial fractions for any pair of complex conjugate poles. The adopted nonorthogonal basis functions are

- For real poles, $p_n = -\alpha_n \in \mathbb{R}$:

$$\Phi_n(s) = \frac{1}{s - p_n} \quad (3.65)$$

- For complex conjugate poles, $p_{n,n+1} = -\alpha_n \pm j\omega_n$, $\alpha_n, \omega_n \in \mathbb{R}$:

$$\Phi_n(s) = \frac{1}{s - p_i} + \frac{1}{s - p_i^*} \quad (3.66)$$

$$\Phi_{n+1}(s) = \frac{j}{s - p_n} - \frac{j}{s - p_n^*} \quad (3.67)$$

3.5.3.7 Orthonormal Vector Fitting

In [140], the Orthonormal Vector Fitting (OVF) technique was developed to approximate frequency domain responses. The OVF method uses orthonormal rational functions for

$\Phi_i(s)$ to improve the numerical stability of the method. This reduces the numerical sensitivity of the system equations to the choice of starting poles and limits the overall macro-modeling time. To ensure the resulting transfer function has real-valued coefficients, the orthonormal functions in the following format are used

- For stable real poles, $p_n = -\alpha_n \in \mathbb{R}$:

$$\Phi_n(s) = \sqrt{-2p_n} \left(\prod_{i=1}^{n-1} \frac{s + p_i^*}{s - p_i} \right) \frac{1}{s - p_n} \quad (3.68)$$

- For stable complex conjugate poles, $p_{n,n+1} = -\alpha_n \pm j\omega_n$, $\alpha_n, \omega_n \in \mathbb{R}$:

$$\Phi_n(s) = \sqrt{-2\Re(p_n)} \left(\prod_{i=1}^{n-1} \frac{s + p_i^*}{s - p_i} \right) \frac{s + |p_n|}{(s - p_n)(s - p_n^*)} \quad (3.69)$$

$$\Phi_{n+1}(s) = \sqrt{-2\Re(p_n)} \left(\prod_{i=1}^{n-1} \frac{s + p_i^*}{s - p_i} \right) \frac{s - |p_n|}{(s - p_n)(s - p_n^*)} \quad (3.70)$$

3.5.3.8 z -Domain Vector Fitting

In [144] z -domain vector-fitting (ZDVF) is proposed to fit transfer functions using frequency or time-domain response data. This method is a reformulation of the original vector fitting method in the z -domain. It has an advantage of faster convergence and better numerical stability compared to the s -domain VF. The fast convergence of the method reduces the overall macromodel generation time.

The z -domain response of a dynamic system can be constructed from the the time domain response or by using s -to- z bilinear transformation. The z -domain response $H(z)$ of any linear time-invariant passive network can be represented using rational function. For an D -th order system response, the rational function is

$$H(z) = \frac{N(z)}{D(z)} = \frac{\sum_{k=0}^N b_k z^{-k}}{\sum_{r=1}^D a_r z^{-r}} = \frac{\sum_{k=1}^N N_k \Phi_k(z)}{\sum_{r=1}^D D_r \Phi_r(z)} \quad (3.71)$$

Using the following basis functions conforms to the requirement of real-valued time-domain response from the resulting transfer function $H(z)$ (3.71) (or from its equivalent in s -domain).

- For real poles, $p_n = \alpha_n \in \mathbb{R}$ and $|P_n| < 1$:

$$\Phi_n(z) = \frac{1}{z - p_n} \quad (3.72)$$

- For complex conjugate poles, $p_{n,n+1} = \alpha_n \pm j\beta_n$, $\alpha_n, \beta_n \in \mathbb{R}$ and $|P_n| < 1$:

$$\Phi_n(z) = \frac{1}{z - p_i} + \frac{1}{z - p_i^*} \quad (3.73)$$

$$\Phi_{n+1}(z) = \frac{j}{z - p_n} - \frac{j}{z - p_n^*} \quad (3.74)$$

3.5.3.9 z -Domain Orthonormal Vector Fitting

An advanced macromodeling tool based on z -domain Orthonormal Vector Fitting (ZD-OVF) was developed in [145, 154] for fast and accurate macromodeling of linear subnetworks using either frequency or time-domain tabulated data. This algorithm extends the inherent advantages of orthonormal basis to z -domain VF formulations. Hence, it further improves the numerical stability of the ZVF method and significantly reduces the numerical sensitivity of the system equations to the choice of starting poles. This directly leads to an improvement of the overall macromodeling time. To this end, the following modified z -domain Takenaka-Malmquist orthonormal bases are developed.

- For real poles, $p_n = \alpha_n \in \mathbb{R}$ and $|P_n| < 1$:

$$\Phi_n(z) = \sqrt{1 - |p_n|^2} \left(\prod_{i=1}^{n-1} \frac{1 - p_i^* z}{z - p_i} \right) \frac{1}{z - p_n} \quad (3.75)$$

- For complex conjugate poles, $p_{n,n+1} = \alpha_n \pm j\beta_n$, $\alpha_n, \beta_n \in \mathbb{R}$ and $|P_n| < 1$:

$$\Phi_n(s) = \frac{1}{\sqrt{2}} |1 + p_n| \sqrt{1 - |p_n|^2} \left(\prod_{i=1}^{n-1} \frac{1 - p_i^* z}{z - p_i} \right) \frac{1 - z}{(z - p_n)(z - p_n^*)} \quad (3.76)$$

$$\Phi_{n+1}(s) = \frac{1}{\sqrt{2}} |1 - p_n| \sqrt{1 - |p_n|^2} \left(\prod_{i=1}^{n-1} \frac{1 - p_i^* z}{z - p_i} \right) \frac{1 + z}{(z - p_n)(z - p_n^*)} \quad (3.77)$$

3.5.3.10 State-Space Realization from Poles and Residues

The real-valued minimal LTI state-space realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ using the poles and residues is an important step in the iterations of the vector fitting methods. It has been corroborated that the zeros of the denominator expression become the improved poles to start the next iteration and for the final transfer function. Calculating the zeros can be done through state-space realization of the denominator. An example was provided for the formulation of original VF in the Sec. 3.5.3.3. For the details of such realizations for other vector fitting methods, the above corresponding reference can be referred to. It is to be noted that, the representation of a transfer function in state-space form is obviously not unique.

Also, to incorporate the resulting multiport macromodels for subnetworks in higher level spice-like simulations, generating an equivalent circuit for the macromodels is important. This can be achieved in two steps. First a state-space representation is obtained. For the details [151] and [64, Sec. 7.3] can be referred to. The second step is to synthesize an equivalent circuit, the details of which can be found in [64, Sec. 7.4].

3.5.3.11 Recursive Convolution

Another way of converting a frequency-domain description to a time-domain model is through convolution, which, in general, has a quadratic CPU-time cost. If the frequency-domain descriptions are in terms of poles and residues, we can exploit this fact and evaluate the convolutions in a recursive manner so that the computational cost is constant regardless of the time [64]. The recursive convolution method is efficient and easy to implement. For further details [64, Sec. 7.5] can be referred to.

3.5.3.12 Quasi-Convex Optimization Method:

This method was originally introduced in [155] and uses more rigorous techniques to obtain guaranteed stable models. It can be used to obtain parameterized models, which preserve additional properties such as passivity [156].

3.6 Other Alternative Methods

It is to be noted that, the existing methods are not strictly limited to the presented categories. There is a variety of recent reduction techniques that have been proposed in the area of linear MOR. These techniques aim at obtaining the best results by combining advantages from different methods. As an example, one can consider *interpolatory model reduction* techniques [137, Sec.1-7, part I] and [157, 158], which has recently (2012) attracted attention. Interested readers can also consult (e.g.) [83, 87, 112, 159–163] for some other examples of such techniques.

Chapter 4

Model Order Reduction for Nonlinear Dynamical Systems

Model order reduction of nonlinear systems follows the model order reduction for linear systems. However, compared to the reduction of linear systems, nonlinear model reduction are much less developed and are far more challenging to develop and analyze. The problem of nonlinear model reduction deals with approximations of the large nonlinear dynamic systems represented in the form of a nonlinear differential equations. This is mainly to reduce costs of simulating large systems, a goal that can only be attained through answering both the following two sub-problems

- a.* Reducing the dimensionality of the state vector,
- b.* Finding ways to efficiently calculate nonlinear functions and derivatives.

While an elaborated formulation of nonlinear systems is presented in Chapter-1, it is concisely revisited here.

For a broad class of engineering problems, the following nonlinear models consisting of a system of state equations (DAE) (4.1a) along with output equations (4.1b) as shown

below, are ample to represent their nonlinear dynamical behavior.

$$\begin{cases} \frac{d}{dt}\mathbf{g}(\mathbf{x}(t)) = \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t), & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{L}\mathbf{x}(t) \end{cases} \quad (4.1a)$$

$$(4.1b)$$

where system variables $\mathbf{x}(t) \in \mathbb{R}^n$, nonlinear vector functions $\mathbf{g}(\mathbf{x})$, $\mathbf{F}(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, $\mathbf{L} \in \mathbb{R}^{q \times n}$, $\mathbf{u}(t) \in \mathbb{R}^p$, and $\mathbf{y}(t) \in \mathbb{R}^q$. Nonlinear electrical networks can also be characterized by a set of coupled nonlinear first order differential equations representing the dynamical behavior of the system variables [21–23, 25, 27, 164–166]. In the context of circuit simulation, these equations are directly obtained from the circuit netlist using the modified nodal analysis (MNA) matrix formulation [37–39, 167] in the form

$$\begin{cases} \mathbf{C} \frac{d}{dt}\mathbf{x}(t) + \mathbf{G}\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t), & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{L}\mathbf{x}(t) \end{cases} \quad (4.2a)$$

$$(4.2b)$$

where $\mathbf{C}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are susceptance and conductance matrices including the contribution of linear elements, respectively.

4.1 Physical Properties of Nonlinear Dynamical Systems

Similar to the linear case, there are important conditions that, nonlinear differential equations representing dynamics of a physical systems have to satisfy. It is also desirable to have such inherent characteristics of the original system passed on to the lower order approximant obtained in the MOR process. This section reviews some of these important properties mainly related to the “Lipschitz Continuity” and the “stability” of dynamical nonlinear systems. The former is the most important condition to ensure the “existence and uniqueness” of the response for a physical system. The latter is significantly important in

analyzing the local/global behavior of nonlinear systems.

4.1.1 Lipschitz Continuity

For any given nonlinear system defined in $\mathcal{D}$ as an initial value problem in the following form

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t, \mathbf{x}(t)), \quad \text{with } \mathbf{x}(t_0) = \mathbf{x}_0, \quad (4.3)$$

existence and uniqueness can be ensured by imposing some constraints on the vector field function $\mathbf{F}(t, \mathbf{x})$ in (4.3). The key constraint for this is *Lipschitz condition*.

Definition 4.1 (Lipschitz condition [7, 168]). Consider the function $\mathbf{F}(t, \mathbf{x}) = [f_1(t, \mathbf{x}), f_2(t, \mathbf{x}), \dots, f_n(t, \mathbf{x})]^T$ with $\mathbf{F} : (\mathbb{R} \times \mathbb{R}^n) \mapsto \mathbb{R}^n$, $|t - t_0| \leq a$, and $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathcal{D} \subset \mathbb{R}^n$, where $\mathcal{D}$ is an open and connected set; $\mathbf{F}(t, \mathbf{x})$ satisfies the *Lipschitz condition* with respect to $\mathbf{x}$ if in $[t_0 - a, t_0 + a] \times \mathcal{D}$ we have

$$\|\mathbf{F}(t, \mathbf{x}_1) - \mathbf{F}(t, \mathbf{x}_2)\| \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|, \quad (4.4)$$

with $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and L a constant. Positive constant L is called the *Lipschitz constant*.

Given function $\mathbf{F}(\mathbf{x}) : \mathcal{D} \mapsto \mathbb{R}^n$ is defined in an open and connected set $\mathcal{D} \subset \mathbb{R}^n$, and each component of $\mathbf{F}(\mathbf{x})$ does not explicitly depend on time. A system of differential equations having this $\mathbf{F}(\mathbf{x})$ as its vector field function, i.e.

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)), \quad \text{with } \mathbf{x}(t_0) = \mathbf{x}_0, \quad (4.5)$$

is referred to as *autonomous* system. For autonomous systems (4.5), Lipschitz condition is also defined in a similar manner to the definition in 4.1.

A locally Lipschitz function on an open and connected domain $\mathcal{D}$ is Lipschitz on every

compact (closed and bounded) subset of $\mathcal{D}$ [17]. According to the domain over which the Lipschitz condition holds, it is categorized as *locally Lipschitz* and *globally Lipschitz*.

A geometrical interpretation for Lipschitz property of function $f(x)$ ($f : \mathbb{R} \mapsto \mathbb{R}$), is illustrated in Fig. 4.1. It implies that on a plot of $f(x)$ versus x , a straight line joining any two points of $f(x)$ cannot have a slope (4.6), whose absolute value is greater than L .

$$\frac{|f(x_2) - f(x_1)|}{|x_2 - x_1|} \leq L. \quad (4.6)$$

Figure 4.1: Illustration of Lipschitz property.

4.1.2 Existence and Uniqueness of Solutions

The existence theorem originally contributed by A. L. Cauchy (1789-1857). Since then, many different forms of existence theorem have been established in the literature of dynamical systems.

Theorem 4.1 (Existence-Uniqueness [7]). Consider the initial value problem given in (4.3) with $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^n$, $|t - t_0| \leq a$; $\mathcal{D} = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\| \leq d\}$, where a and d are positive constants. The vector function $\mathbf{F}(t, \mathbf{x})$ satisfies the following conditions:

a) $\mathbf{F}(t, \mathbf{x})$ is continuous in $G = [t_0 - a, t_0 + a] \times \mathbf{D}$;

b) $\mathbf{F}(t, \mathbf{x})$ is Lipschitz continuous in $\mathbf{x}$.

Then the initial value problem has one and only one solution for $|t - t_0| \leq \min(a, \frac{d}{M})$ with $M = \sup_G \|\mathbf{F}\|$.

Proof. For the proof [8, 9, 169] can be referred to. ■

4.1.3 Stability of Nonlinear Systems

This section briefly reviews several definitions and results concerning the stability of nonlinear dynamical systems that have been commonly used for the stability analysis of the reduced nonlinear models in the literature (e.g. [170–174]). For some more details on the fundamental concepts and definitions of stability analysis, Chapter 2.6.1 can be referred to. However, it needs to be remarked that, the stability analysis for nonlinear systems is not nearly as simple as it was for linear systems, presented in Sec. 3.1.1.

Consider the nonlinear dynamical system

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t)) \quad (4.7)$$

with equilibrium point $\mathbf{x}_{eq}$ such that $\mathbf{F}(\mathbf{x}_{eq}, \mathbf{0}) = \mathbf{0}$. For (4.7) the local behavior analysis (*cf.* Sec. 2.6.2.1) at an equilibrium point turns out to be a less sophisticated task, compared to the difficulties associated with the global stability analysis (*cf.* Sec. 2.6.2.2). In order to determine the local behavior of a nonlinear system at an equilibrium point, it is sufficient to consider the linearizations of the nonlinear model about that equilibrium point and analyze the stability of the local approximate model. This is referred to as Lyapunov's indirect method.

Theorem 4.2 (Lyapunov's indirect method [55, 170]). *If the linearized system*

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t), \quad \text{where } \mathbf{A} = \left. \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{eq}, \mathbf{u}=\mathbf{0}} \quad (4.8)$$

is asymptotically stable, then $\mathbf{x}_{eq}$ is a locally asymptotically stable equilibrium of the system (4.5).

Thus, the equilibrium of the nonlinear system is stable if the Jacobian matrix $\mathbf{A}$ has eigenvalues with a strictly negative real part.

Consider the nonlinear dynamical system

$$\mathbf{C}\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{y}(t) = \mathbf{L}\mathbf{x}, \quad (4.9)$$

which may arise when modeling analog circuits using modified nodal analysis. Assume that the descriptor matrix $\mathbf{C}$ is nonsingular and the system has a unique equilibrium point $\mathbf{x}_{eq}$. This equilibrium point is said to be *exponentially stable* if all solutions to the autonomous system (i.e. $\mathbf{u} = \mathbf{0}$, $\forall t$) for any arbitrary initial condition $\mathbf{x}_0$ converge to the equilibrium point exponentially fast. Without a loss of generality we may transform the coordinate system such that $\mathbf{x}_{eq} = \mathbf{0}$.

Definition 4.2 (Exponentially stable [171]). The equilibrium $\mathbf{x}_{eq} = \mathbf{0}$ is said to be exponentially stable if there exist constants $r, a, b > 0$ such that

$$\|\mathbf{x}(t_0 + t)\| \leq a \|\mathbf{x}_0\| e^{-bt}, \quad \forall t, t_0 \geq 0, \quad \forall \mathbf{x}_0 \in \mathbb{B}_r \subseteq \mathbb{R}^n \quad (4.10)$$

Here, $\mathbb{B}_r$ is a ball with radius r centered at $\mathbf{x}_{eq}$.

Exponential stability can be proven through Lyapunov functions.

Theorem 4.3 ([170, 175]). *The equilibrium point $\mathbf{x}_{eq} = \mathbf{0}$ of system (4.9) is exponentially stable if there exist constants $\lambda_1, \lambda_2, \lambda_3 > 0$ and a continuously differentiable Lyapunov function $L(\mathbf{x})$ such that*

$$\lambda_1 \mathbf{x}^T \mathbf{x} \leq L(\mathbf{x}) \leq \lambda_2 \mathbf{x}^T \mathbf{x} \quad (4.11)$$

$$\frac{\partial}{\partial t} L(\mathbf{x}) \leq -\lambda_3 \mathbf{x}^T \mathbf{x} \quad (4.12)$$

$$\forall t \geq 0, \forall \mathbf{x} \in \mathbb{B}_r \subseteq \mathbb{R}^n.$$

Definition 4.3 (Globally exponentially stable [171]). If $\mathbb{B}_r = \mathbb{R}^n$ (in Def. 4.2 or Theorem 4.3), then the equilibrium point is globally exponentially stable.

External stability refers to the input-output system (4.9) and concerns the system's ability to amplify signals from input $\mathbf{u}$ to output $\mathbf{y}$. Qualitatively, the system is said to be externally stable if the system's output $\mathbf{y}(t)$ can be bounded in some measure by a linear function of the system's input $\mathbf{u}(t)$ in that same measure.

Definition 4.4 (Small-signal finite-gain L_p stable [171, 175]). System (4.9) is said to be small-signal finite-gain L_p stable if there exist $r_p > 0$ and $\gamma_p < \infty$ such that

$$\|\mathbf{y}\|_p \leq \gamma_p \|\mathbf{u}\|_p \quad (4.13)$$

for all $t > t_0$, given initial state $\mathbf{x}(0) = \mathbf{0}$ and input $\mathbf{u}(t)$ such that $\|\mathbf{u}\|_\infty < r_p$. if $r_p = \infty$, then the system is finite-gain L_p stable.

Theorem 4.4 (Small-signal finite-gain L_p stable [170, 175]). *Suppose $\mathbf{x}_{eq} = \mathbf{0}$ is an exponentially stable equilibrium of system (4.9). If $\mathbf{F}(\mathbf{x}, \mathbf{u})$ is continuously differentiable and $\mathbf{F}(\mathbf{x}, \mathbf{u})$ is locally Lipschitz continuous at $(\mathbf{x}_e, \mathbf{u} = \mathbf{0})$, then system (4.9) is small-signal finite-gain L_p stable.*

Definition 4.5 (Finite-gain L_p stable). If a small-signal finite-gain L_p stable system (4.9) in Theorem-4.4 is globally exponentially stable ($\mathbb{B}_r = \mathbb{R}^n$), the system is finite-gain L_p stable.

4.2 Nonlinear Order Reduction Algorithms

In Fig. 4.2 four different classes of the existing model order reduction techniques for nonlinear systems in (4.1) (or its counterpart (4.1)) are displayed.

Figure 4.2: Model reduction methods for nonlinear dynamical systems categorized into four classes.

4.2.1 Projection framework for Nonlinear MOR - Challenges

The projection-based approaches have been the most successful algorithms for reduction of large-scale linear systems (*cf.* Chapter 3 and references therein e.g. [64, 89, 90]). In contrast, a direct application of the projection framework to the large nonlinear systems can generally face some challenges. Consider the nonlinear system in (4.1). We may

formally apply the projection recipe to this system of equations which leads to a “reduced” order system of equations in the following form

$$\begin{cases} \frac{d}{dt} \mathbf{W}^T \mathbf{g}(\mathbf{V} \mathbf{z}(t)) = \mathbf{W}^T \mathbf{F}(\mathbf{V} \mathbf{z}(t)) + \mathbf{W}^T \mathbf{B} u(t), & \mathbf{z}_0 = \mathbf{V} \mathbf{x}(t_0) \\ \mathbf{y}(t) \approx \mathbf{L} \mathbf{V} \mathbf{z}(t). \end{cases} \quad (4.14a)$$

$$(4.14b)$$

First, it is not at all clear how to choose left and right projection matrices $\mathbf{W}$ and $\mathbf{V}$, and even less is known about efficient computation of the response from the resulting models. In the general case, interpreting the terms $\mathbf{W}^T \mathbf{g}(\mathbf{V} \mathbf{z}(t))$ and $\mathbf{W}^T \mathbf{F}(\mathbf{V} \mathbf{z}(t))$ as “reduced” models is problematic. Since $\mathbf{g}(\mathbf{x})$ and $\mathbf{F}(\mathbf{x})$ are nonlinear function, the only way to computing $\mathbf{W}^T \mathbf{g}(\mathbf{V} \mathbf{z}(t))$ and $\mathbf{W}^T \mathbf{F}(\mathbf{V} \mathbf{z}(t))$ may be to (1) explicitly construct $\tilde{\mathbf{x}} = \mathbf{V} \mathbf{z}$, (2) evaluate $\tilde{\mathbf{g}} = \mathbf{g}(\tilde{\mathbf{x}})$ and $\tilde{\mathbf{F}} = \mathbf{F}(\tilde{\mathbf{x}})$, and (3) compute $\mathbf{W}^T \tilde{\mathbf{g}}$ and $\mathbf{W}^T \tilde{\mathbf{F}}$. As a result, an efficient simulation is not guaranteed. For example, in a nonlinear circuit simulation, even for circuits with tens of thousands of nodes, roughly half the simulation time is spent on evaluating the nonlinear functions $\mathbf{g}(\mathbf{x})$ and $\mathbf{F}(\mathbf{x})$. Thus, regardless of the reduction in the size of the state space, since the original nonlinear functions must be evaluated, the efficiency gained from the order reduction will be at most a factor of two or three [176]. Consequently, in order to obtain efficient, low-cost reduced order models for nonlinear systems the following two issues need to be addressed by a MOR technique:

- (a) Constructing low-order projection basis $\mathbf{W}$ and $\mathbf{V}$, which approximate the dominant (‘useful’) parts of the state-space to accommodate the dynamics of the system
- (b) Applying the low-cost, yet feasible approximate representation of system’s nonlinearity (associated with $\mathbf{g}(\mathbf{x})$ and $\mathbf{F}(\mathbf{x})$).

The first of the issues is developing suitable projection bases for linear systems. This has received a considerable amount of attention, leading to successful approaches based (e.g.) on Proper Orthogonal Decomposition that is explained in the Sec. 4.2.4. The techniques

such as linearization, quadratic methods presented in Sec. 4.2.2, and TPWL approach in Sec. 4.2.3 are used to address the problem of finding cost-efficient representations of nonlinearity.

4.2.2 Nonlinear Reduction Based on Taylor Series

The very first practical approaches to nonlinear model reduction were based on using Taylor series expansions of nonlinear functions $\mathbf{F}(\mathbf{x}(t))$ and $\mathbf{g}(\mathbf{x}(t))$ in [42, 176–180]. The original system should be expanded around some (initial, equilibrium) state $\mathbf{x}_0$ on a representative trajectory, using a multidimensional Taylor series expansion technique. The Taylor expansion of the functions are

$$\begin{aligned} \mathbf{F}(\mathbf{x}) = & \mathbf{F}(\mathbf{x}_0) + \mathbf{G}_1 (\mathbf{x} - \mathbf{x}_0) + \mathbf{G}_2 (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) + \\ & \mathbf{G}_3 (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) + \dots \end{aligned} \quad (4.15)$$

and

$$\begin{aligned} \mathbf{g}(\mathbf{x}) = & \mathbf{g}(\mathbf{x}_0) + \mathbf{C}_1 (\mathbf{x} - \mathbf{x}_0) + \mathbf{C}_2 (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) + \\ & \mathbf{C}_3 (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) \otimes (\mathbf{x} - \mathbf{x}_0) + \dots \end{aligned} \quad (4.16)$$

where $\otimes$ is the Kronecker product, $\mathbf{C}_1 = \left. \frac{\partial}{\partial \mathbf{x}} \mathbf{g} \right|_{\mathbf{x}_0}$ and $\mathbf{G}_1 = \left. \frac{\partial}{\partial \mathbf{x}} \mathbf{F} \right|_{\mathbf{x}_0}$ ($\in \mathbb{R}^{n \times n}$) are the Jacobian matrices and $\mathbf{C}_2, \mathbf{G}_2$ ($\in \mathbb{R}^{n \times n^2}$) are Hessian tensors that represent the second order contributions and in general $\mathbf{C}_k, \mathbf{G}_k$ ($\in \mathbb{R}^{n \times n^k}$) correspond to tensors with the k -th order effect contributions.

4.2.2.1 Linearization Methods

Linearization is the earliest and most straightforward approach to model nonlinear systems. In this approach, the strategy is based on linearizing the system around some point in its state-space. For this purpose, $\mathbf{F}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are assumed to be smooth enough so that it can be expanded into Taylor series. This means that the sum of infinite terms in (4.15) and (4.16) are truncated to only the first order terms or linear components. Substituting the truncated approximant series for $\mathbf{F}(\mathbf{x})$ (4.15) and $\mathbf{g}(\mathbf{x})$ (4.16) in (4.1), we get

$$\frac{d}{dt} (\mathbf{g}(\mathbf{x}_0) + \mathbf{C}_1 (\mathbf{x} - \mathbf{x}_0)) = \mathbf{F}(\mathbf{x}_0) + \mathbf{G}_1 (\mathbf{x} - \mathbf{x}_0) + \mathbf{B}\mathbf{u}(t) \quad (4.17)$$

and next,

$$\mathbf{C}_1 \frac{d}{dt} \mathbf{x} = \mathbf{G}_1 \mathbf{x} - \mathbf{G}_1 \mathbf{x}_0 + \mathbf{F}_0 + \mathbf{B}\mathbf{u}(t) \quad (4.18)$$

where $\mathbf{F}_0 \triangleq \mathbf{F}(\mathbf{x}_0)$.

It can be equivalently rewritten as

$$\mathbf{C}_1 \frac{d}{dt} \mathbf{x} = \mathbf{G}_1 \mathbf{x} + \underbrace{\begin{bmatrix} \mathbf{B}, & \mathbf{F}_0 - \mathbf{G}_1 \mathbf{x}_0 \end{bmatrix}}_{\tilde{\mathbf{B}}} \underbrace{\begin{bmatrix} \mathbf{u}(t) \\ 1 \end{bmatrix}}_{\tilde{\mathbf{u}}(t)}. \quad (4.19)$$

It is desirable to match m moments of the transfer function of the reduced model with those of the transfer function of the original system (4.19) at the complex frequency point s_0 . For this purpose, the projection matrix for order reduction is constructed using the Arnoldi algorithm based on the Krylov subspace as follows

$$\text{colspan}\{\mathbf{V}\} = \text{Ker}\{\mathbf{A}, \mathbf{R}\} = \text{span}\{\mathbf{R}, \mathbf{A}\mathbf{R}, \dots, \mathbf{A}^m \mathbf{R}\}, \quad (4.20)$$

where $\mathbf{R} \triangleq (s_0 \mathbf{C}_1 - \mathbf{G}_1)^{-1} \tilde{\mathbf{B}}$ and $\mathbf{A} \triangleq (s_0 \mathbf{C}_1 - \mathbf{G}_1)^{-1} \mathbf{C}_1$.

Using the projection matrix $\mathbf{V}$ and a variable change as $\mathbf{x}(t) = \mathbf{V} \mathbf{z}(t)$, the reduced linearized model is

$$\begin{cases} \hat{\mathbf{C}} \frac{d}{dt} \mathbf{z}(t) = \hat{\mathbf{G}} \mathbf{z}(t) + \hat{\mathbf{B}} \mathbf{u}(t), \\ \hat{\mathbf{y}}(t) = \hat{\mathbf{L}} \mathbf{z}(t), \end{cases} \quad (4.21a)$$

$$(4.21b)$$

where

$$\begin{aligned} \hat{\mathbf{C}} &= \mathbf{V}^T \mathbf{C}_1 \mathbf{V}, & \hat{\mathbf{G}} &= \mathbf{V}^T \mathbf{G}_1 \mathbf{V}, \\ \hat{\mathbf{B}} &= \mathbf{V}^T \tilde{\mathbf{B}}, & \hat{\mathbf{L}} &= \mathbf{L} \mathbf{V}. \end{aligned} \quad (4.22)$$

4.2.2.2 Quadratic Methods

Quadratic method is an improvement over the previous approach (*cf.* Sec. 4.2.2.1) where the second order term of the expansion in (4.15) is also included for the state-space approximation of the nonlinear system. Given $\mathbf{x}_0 = \mathbf{0}$ as the expansion point, the approximants for nonlinear functions in (4.1) are

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{0}) + \mathbf{G}_1 \mathbf{x} + \mathbf{x}^T \mathbf{G}_2 \mathbf{x}, \text{ and } \mathbf{g}(\mathbf{x}) \approx \mathbf{g}(\mathbf{0}) + \mathbf{C}_1 \mathbf{x}. \quad (4.23)$$

where $\mathbf{C}_1$ and $\mathbf{G}_1$ are the Jacobian of $\mathbf{g}$ and $\mathbf{F}$, respectively, evaluated about origin and $\mathbf{G}_2$ is an $n \times n^2$ Hessian tensor whose entries are given by

$$\mathbf{g}_{2i,j,k} = \frac{1}{2} \frac{\partial^2 f_i}{\partial x_j \partial x_k}. \quad (4.24)$$

and hence,

$$\mathbf{C}_1 \frac{d}{dt} \mathbf{x} = \mathbf{G}_1 \mathbf{x} + \mathbf{x}^T \mathbf{G}_2 \mathbf{x} + \underbrace{\begin{bmatrix} \mathbf{B}, & \mathbf{F}_0 \end{bmatrix}}_{\tilde{\mathbf{B}}} \underbrace{\begin{bmatrix} \mathbf{u}(t) \\ 1 \end{bmatrix}}_{\tilde{\mathbf{u}}(t)}. \quad (4.25)$$

Having a desirable projection matrix $\mathbf{V}$, the reduced model is

$$\mathbf{V}^T \mathbf{C}_1 \mathbf{V} \frac{d}{dt} \mathbf{z} = \mathbf{V}^T \mathbf{G}_1 \mathbf{V} \mathbf{z} + \mathbf{V}^T \mathbf{z}^T \mathbf{V}^T \mathbf{G}_2 \mathbf{V} \mathbf{z} + \mathbf{V}^T \tilde{\mathbf{B}} \tilde{\mathbf{u}}(t). \quad (4.26)$$

In tensorial notation, it can be equivalently rewritten as

$$(\mathbf{V}^T \mathbf{C}_1 \mathbf{V}) \frac{d}{dt} \mathbf{z} = (\mathbf{V}^T \mathbf{G}_1 \mathbf{V}) \mathbf{z} + \mathbf{V}^T \mathbf{G}_2 (\mathbf{V} \otimes \mathbf{V}) (\mathbf{z} \otimes \mathbf{z}) + (\mathbf{V}^T \tilde{\mathbf{B}}) \tilde{\mathbf{u}}(t), \quad (4.27a)$$

$$\mathbf{y}(t) \approx \mathbf{L} \mathbf{V} \mathbf{z}(t). \quad (4.27b)$$

From the reduction process above, it is seen that the quadratic method is more precise than the traditional linearization method. To obtain the projection matrix $\mathbf{V}$, the simple approach is to consider the Krylov subspaces defined by the linear part of the representation by ignoring the second order term. There are also specially designed methods for the extraction of the projection matrix, known as the quadratic projection methods [42, 180–183]. They are based on considering the effect of the linear and second order components in Taylor series expansion for defining the corresponding Krylov subspace and to construct the orthogonal vector basis.

4.2.2.3 Bilinearization Reduction Method

This section reviews the bilinearization reduction method based on the approach proposed in [176]. This method uses the terms from the Taylor expansion in (4.15) up to the quadratic

term to approximate the nonlinear functions. For simplicity of description, consider the expanded state-space model of the nonlinear system as

$$\frac{d}{dt}\mathbf{x} = \mathbf{G}_1\mathbf{x} + \mathbf{G}_2\mathbf{x} \otimes \mathbf{x} + \mathbf{B}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{L}\mathbf{x}. \quad (4.28)$$

Following the computational steps in [176, 184] an approximate bilinear system is obtained as

$$\frac{d}{dt}\mathbf{x}^\otimes = \mathbf{A}^\otimes\mathbf{x}^\otimes + \mathbf{N}^\otimes\mathbf{x}^\otimes + \mathbf{B}^\otimes\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{L}^{\otimes T}\mathbf{x}, \quad (4.29)$$

where

$$\mathbf{x}^\otimes = \begin{bmatrix} \mathbf{x} \\ \mathbf{x} \otimes \mathbf{x} \end{bmatrix}, \quad \mathbf{B}^\otimes = \begin{bmatrix} \mathbf{B} \\ 0 \end{bmatrix}, \quad \mathbf{L}^\otimes = \begin{bmatrix} \mathbf{L} \\ 0 \end{bmatrix}, \quad (4.30)$$

$$\mathbf{A}^\otimes = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{0} & \mathbf{G}_1 \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{G}_1 \end{bmatrix}, \quad \mathbf{N}^\otimes = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{B} & \mathbf{0} \end{bmatrix}. \quad (4.31)$$

The resulting bilinear systems (4.29) is of much larger dimension than the original nonlinear system. For bilinear systems, Volterra-series expression [185] and multimoment expansions [176] are the key to applying the Krylov subspace based MOR. For further details [184, 186] can be referred to.

4.2.3 Piecewise Trajectory based Model Order Reduction

The main drawback of Taylor-series based MOR methods for nonlinear systems is that these methods typically expand the nonlinear operator about a single state (*cf.* Sections 4.2.2.1, 4.2.2.2, and 4.2.2.3). Therefore the generated models are only accurate locally, a fact that makes them useful only for weakly nonlinear systems. In order to overcome this weak nonlinearity limitation, TPWL approach was first proposed in [187] and then extended in several ways (e.g.) [171, 172, 188–198]. The central idea in all these approaches is to use a collection of expansions around states visited by a given training trajectory. Hence, one may categorize them in a class of methods that can be referred to as “piecewise trajectory based” methods.

TPWL: The idea in TPWL [187, 188, 199] is to represent a nonlinear system as a collage of *linear models* in adjoining polytopes, centered around *expansion points* $\mathbf{x}_i$, in the state space as illustrated in Fig. 4.3.

Given a nonlinear system as shown below

$$\begin{cases} \mathbf{C} \frac{d}{dt} \mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t)) + \mathbf{B}u(t), & \mathbf{x}(t_0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{L} \mathbf{x}(t). \end{cases} \quad (4.32a)$$

$$(4.32b)$$

For the sake of simplicity in the form of equations for TPWL, in (4.32), we considered $\mathbf{g}(\mathbf{x}) = \mathbf{C}$, where $\mathbf{C}$ is a constant matrix. The steps of **TPWL** approach can be summarized as follows.

1. Finding the nonlinear system trajectory in response to a properly chosen training input.
2. Locating meaningful states as linearisation points (LPs) along the trajectory at which

Figure 4.3: Illustration of the state space of a planar system, where $\mathbf{x}_i$ are the expansion points on the *training trajectory* A. Because solutions B and C are in the vicinity ball of the expansion states, they can be efficiently simulated using a TPWL model, however this can not be true for the solutions D and E.

local approximations are to be created, as

$$\text{LP} : \mathcal{X} = \{\mathbf{x}_i : \mathbf{x}_i = \mathbf{x}(t_i), \text{ for } i = 1, \dots, \mathcal{M}\}. \quad (4.33)$$

3. Linearizing nonlinear function $\mathbf{F}(\mathbf{x})$ at selected LPs results in an approximation for the original nonlinear system as shown in Sec. 4.2.2.1. The resulting linearized nonlinear system is

$$\mathbf{C} \frac{d}{dt} \mathbf{x} = \mathbf{G}_i (\mathbf{x} - \mathbf{x}_i) + \mathbf{F}_i + \mathbf{B} \mathbf{u}(t), \quad \forall \mathbf{x}_i \in \mathcal{X}, \quad (4.34)$$

where $\mathbf{F}_i \triangleq \mathbf{F}(\mathbf{x}_i)$.

4. Finding the dominant subspace of the system, in which the system dynamics lie. This is achieved by generating projection basis $\mathbf{V}_i$ for each local LTI model and

calculating a common subspace $\mathbf{V}$ through aggregation of the local subspaces as $\mathbf{V}_{agg} = \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_M\}$. Then, reorthogonalization of the column vectors in $\mathbf{V}_{agg}$ using SVD construct a new basis $\mathbf{V}$. The order of the global reduced subspace $\text{colspan}(\mathbf{V})$ is usually larger than each $\mathbf{V}_i$ but much smaller than the size of the original system.

5. Performing the linear model reduction using $\mathbf{V}$ on the local linear submodels and the output equation,

$$\hat{\mathbf{C}} \frac{d}{dt} \mathbf{z} = \hat{\mathbf{G}} (\mathbf{z} - \mathbf{z}_i) + \hat{\mathbf{F}}_i + \hat{\mathbf{B}} \mathbf{u}(t), \quad \forall \mathbf{x}_i \in \mathcal{X}, \quad \mathbf{y}(t) \approx \hat{\mathbf{C}} \mathbf{z}, \quad (4.35)$$

where

$$\begin{aligned} \hat{\mathbf{C}} &= \mathbf{V}^T \mathbf{C} \mathbf{V}, & \hat{\mathbf{B}} &= \mathbf{V}^T \mathbf{B}, & \hat{\mathbf{L}} &= \mathbf{L} \mathbf{V}, \\ \hat{\mathbf{G}}_i &= \mathbf{V}^T \mathbf{G}_i \mathbf{V}, & \hat{\mathbf{F}}_i &= \mathbf{V}^T \mathbf{F}(\mathbf{x}_i). \end{aligned} \quad (4.36)$$

6. The final reduced model is the weighted combination of all the reduced models

$$\hat{\mathbf{C}} \frac{d}{dt} \mathbf{z} = \sum_{i=1}^M \omega_i(\mathbf{z}) \left(\hat{\mathbf{G}}_i (\mathbf{z} - \mathbf{z}_i) + \hat{\mathbf{F}}_i + \hat{\mathbf{B}} \mathbf{u}(t) \right), \quad \mathbf{y}(t) \approx \hat{\mathbf{C}} \mathbf{z}, \quad (4.37)$$

where $\omega_i(\mathbf{z})$ is the weighting function.

On the reduction of the linear submodels: In the steps-4 and 5, basically, any MOR-technique for linear problems can be applied to the linear submodels. In the original approach [199], a Krylov-based reduction using the Arnoldi-method was proposed. [189] introduced Truncated Balanced Reduction (TBR) to TPWL and [200] proposed using Poor Man's TBR. Proper Orthogonal Decomposition (POD) was also used in [194] as linear MOR kernel. For comparison of different linear MOR strategies when applied to problems

in circuit simulation [201–204] can be referred to.

Determination of the weights: In the step-6, the method deals with the combination of weighted reduced linear submodels. To ensure, and at the same time, to limit the dominance of each submodel to its own segment, the weighting functions of choice naturally require to have steep gradients. To this end, the original work in [199] suggested a scheme that is depending on the absolute distance of a state to the linearization points. The importance of each single model is weighted by

$$\omega_i(\mathbf{x}) = e^{-\frac{\beta}{\gamma} \|\mathbf{x} - \mathbf{x}_i\|_2}, \quad \text{with } \gamma = \min_i \|\mathbf{x} - \mathbf{x}_i\|_2. \quad (4.38)$$

where β decides the pace of the decay for weighting functions. A typical value may be chosen as $\beta = 25$. To guarantee a convex combination, the weights are normalized such that $\sum_i \omega_i(\mathbf{x}) = 1$.

The TPWL has excellent global approximations for large signal analysis because of the piecewise nature but has limited local accuracy for small signal analysis. Intuitively, when the excitation is small enough to keep the states stay within one region, the system reduces to a pure LTI model, and no distortions could be captured. Nonlinearities induced exclusively by the nonlinear weight function $\omega_i(\mathbf{z})$ are generated only when states cross boundaries. Recently, some works [191–193] have greatly extended the original TPWL method, making it more scalable and practical. However, there is still less evidence in literature to show the usage of the generated macromodel in other analysis, such as dc, ac, HB, etc. [196].

TPWP: To address the above shortcomings, a method proposed in [196]. It combines the trajectory-based techniques and the weakly nonlinear MOR algorithms. This method is dubbed PWP because of its reliance on 'PieceWise Polynomials'. It follows the TPWL methodology, but instead of using purely linear representations, it approximate each region

with higher order (tensor) polynomials. The PWP claims the possibility of exploiting any existing polynomial MOR technique (e.g. in [205–207]) to perform the weakly nonlinear reduction for each piecewise region.

4.2.4 Proper Orthogonal Decomposition (POD) Methods

Proper orthogonal decomposition (POD), also known as Karhunen-Loève decomposition [125] or principal component analysis (PCA) [126], provides a technique for analyzing multidimensional data. The original concept goes back to [208] and it has been developed in many application areas such as: image processing, fluid dynamics and electrical engineering. Application of POD to dynamical system model reduction calls for using system full state response. This method essentially constructs an orthonormal projection basis from the orthogonalized snapshots of the state (/ data) vectors $\mathbf{x}(t)$ obtained during simulation of some *training input*. After obtaining the projection matrix from POD approaches, it is used to generate a reduced model via a standard projection scheme. Clearly, the choice of the initial excitation function(s) (see [127] and references therein) and the data set from the associated simulation(s) play a crucial role in the POD process.

4.2.4.1 Method of Snapshots

Among all the possibilities, the most prominent approach is known to be the *method of snapshots*. In this method, the POD basis vectors is calculated through performing a SVD of the matrix $\mathbf{x}(t)$, as

$$\mathbf{X}_t = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)] = \mathbf{U}\mathbf{\Sigma}\mathbf{V}. \quad (4.39)$$

The first columns of $\mathbf{U}$ are the POD basis vectors corresponding to the highest singular values. Those are also the eigenvectors of the correlation matrix $\mathcal{C} = \mathbf{X}_t \mathbf{X}_t^T \in \mathbb{R}^{n \times n}$.

The details of the method of snapshot are presented in the associated algorithm table in Chapter 7.

4.2.4.2 Sirovitch Method

If n is very large, it might be prohibitive to use the above approach. Taking advantage of the fact that $N \ll n$, is another method introduced by Sirovich [209]. In this method, the correlation matrix $\mathcal{C}$ is replaced with the following temporal covariance matrix

$$\mathbf{R} = \frac{1}{N} \mathbf{X}_t^T \mathbf{X}, \quad (4.40)$$

which is only $N \times N$.

The algorithm for the Sirovitch method is shown in the following Algorithm-4.

It is to be noted that, for the step-5 of the algorithm 4, we should decide the number of POD basis vectors that are capturing a certain percent of system energy. POD reduces the model in favor of the states containing most of the system “energy”, the so-called dominant dynamics.

POD is general because it can take any trajectory of state variables. This advantage of POD is also its limitation. Because the POD basis is generated from system response with a specific input, the reduced model is only guaranteed to be close to the original system when the input is close to the modeling input. For this purpose, the excitation signals should be carefully decided such that its frequency spectrum is rich enough to excite all dynamics important to the intended model used in the application.

Despite the reservations, model reduction via POD is quite popular and is the method of choice in many fields, mainly due to its simplicity of implementation and promising accuracy. Another advantage of this method is that, it can also be applied to highly complex

Algorithm 4: method of Sirovitch for POD

-
- input** : Original Model ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$)
output: Reduced Macromodel ($\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{D}}$)
- 1 Simulate the original system of order n to obtain N snapshots of state vector $\mathbf{X}_t = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)] \in \mathbb{R}^{n \times N}$;
 - 2 Calculate the mean of the snapshots: $\bar{\mathcal{X}}_i = \frac{1}{N} \sum_{k=1}^N x_i(t_k)$;
 - 3 Obtain new snapshot ensemble with zero mean for each state: $\bar{x}_i(t_k) = x_i(t_k) - \bar{\mathcal{X}}_i$ for $k = 1, \dots, N$;
 - 4 Form the matrix of the new snapshots: $\bar{\mathbf{X}}_t = [\bar{x}_i(t_k)] \in \mathbb{R}^{n \times N}$ for $i = 1, \dots, n$ while $1 \leq k \leq N$;
 - 5 Construct the temporal covariance matrix $\mathbf{R} = \frac{1}{N} \bar{\mathbf{X}}_t^T \bar{\mathbf{X}}_t$, where each entries in $\mathbf{R} = \left[\frac{r_{ij}}{N} \right] \in \mathbb{R}^{N \times N}$ is $r_{ij} = \bar{\mathbf{x}}^T(t_i) \bar{\mathbf{x}}(t_j)$;
 - 6 Calculate the POD eigenvectors $\mathbf{T}_i$ and eigenvalues λ_i of $\mathbf{R}$;
 - 7 Rearrange eigenvalues (and corresponding eigenvectors) in descending order;
 - 8 Find the number of POD basis vectors capturing a certain percent of energy of the ensemble:
 - 9 **while** $\left(\frac{\mathcal{E}}{\sum_{m=1}^N \lambda_m} \right) (\%) < \text{"certain percent"}$ **do**
 - 10 $\mathcal{E} = \mathcal{E} + \lambda_m$;
 - 11 $m = m + 1$;
 - 12 Form the order reduction projection matrix $\mathbf{Q} = [\mathbf{T}_1, \dots, \mathbf{T}_m]$;
 - 13 Project the governing equations onto the reduced basis as $\hat{\mathbf{A}} = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$,
 $\hat{\mathbf{B}} = \mathbf{Q}^T \mathbf{B}$, $\hat{\mathbf{C}} = \mathbf{C} \mathbf{Q}$ $\hat{\mathbf{D}} = \mathbf{D}$;
 - 14 **Remark:** The m most energetic (normalized) POD basis are $\Phi_i = \frac{\mathbf{V}_i}{\|\mathbf{V}_i\|}$, where $\mathbf{V}_i = \bar{\mathbf{X}}_t^T \mathbf{T}_i$ for $i = 1, \dots, m$;
-

linear systems in a straightforward manner.

4.2.4.3 Missing Point Estimation

The missing point estimation (MPE) was proposed in [210] to reduce the cost of updating system information in the solution process of time varying systems arising in computational fluid dynamics. In [41, 211–214] the MPE approach was brought forward for the circuit simulation.

4.2.5 Empirical Balanced Truncation

Balanced truncation is one of the well known methods for model reduction of linear systems (see Sec. 3.4.2). The balanced reduction is accomplished by Galerkin projection onto the states associated to the largest Hankel singular values. It was expanded by Scherpen to locally asymptotically stable nonlinear systems mainly based on the controllability and observability functions and their corresponding singular values [215, 216]. Since then, many results on nonlinear balanced truncation techniques for reduction of finite dimensional nonlinear systems have been developed (e.g.) in [217–219]. However, it is not clear how these approaches can be applied to dynamic systems with high dimensions. In nonlinear balanced truncation, for (affine) nonlinear systems, the controllability and observability functions were shown to be solutions of Hamilton-Jacobi-Bellman and Lyapunov type equations, respectively. Undesirably, solving them is a computationally expensive task. After these “Gramian” functions are computed, an appropriate nonlinear coordinate transformation to “diagonalize” and balance the system is necessary. Its computation turns out to be prohibitively challenging. Then as usual, truncating the weakly controllable and observable states yields the reduced model. Nonlinear balancing has been introduced in theory with strong mathematical support, but no general purpose algorithm exists. Practically, due to the required numerical effort, The method is still difficult to apply to systems with general nonlinearities, and it is not clear how it can be applied systematically by means of numerical computations. Hence, only models with very moderate size have so far been considered.

As highlighted above, being too computationally intensive to compute, it is not satisfactory to reduce nonlinear systems based on linear gramians and nonlinear energy functions. In [220] a hybrid method was developed to tackle this issue using “empirical gramians”,

which can be computed from simulation (or experimental) data for realistic operating conditions. In *empirical balanced truncation* it is possible that instead of creating the reduced subspace with only one relevant input and initial state, several training trajectories are created and the reduced subspace is built in a similar way. Since in this concise review, repeating the involved mathematical formulation does not serve the purpose of clarification in any ways, a flowchart of the algorithm is presented in the following Fig. 4.4. For detailed formulation, [220–222] can be referred to.

Figure 4.4: Nonlinear Balanced model reduction.

- Covariance matrix is computed from data collected along system trajectories. These trajectories represent the system behavior under some input, starting from different initial conditions.
- For *Empirical controllability gramian* $\mathbf{W}_C$, see [221, Definition-6]
- For *Empirical observability gramian* $\mathbf{W}_O$, see [221, Definition-7]

4.2.6 Summary

The properties of the available nonlinear model order reduction algorithms is summarized and presented in 4.1.

Table 4.1: Comparison of properties of the available nonlinear model order reduction algorithm

Nonlinear MOR methods	Advantages	Disadvantages
Linearization	Simple implementation and fast model extraction, Full-system simulation is not necessary	Very limited accuracy, Only applicable to weakly nonlinear systems with small signal excitation, Can not capture any nonlinear distortions
Quadratic Methods	Improve accuracy over linearized models, Full-system simulation is not necessary	Reduction process is more involved (compared to linearization), Still limited to weakly nonlinear systems with small signal excitation, Can not capture high order nonlinear distortions
Bilinearization & Volterra Series	Moment matching, Full-system simulation is not necessary	Increased dimension of the state vector, Not applicable to DAEs
TPWL	Cheap reduced model evaluations	Requires full system simulation for some training input, High memory Usage, Low accuracy for highly nonlinear systems, Poor local accuracy for small signal analysis, Deciding the expansion points is by heuristics
POD	Straightforward implementation, High accuracy, Different inputs/initial values for modeling are possible	Limited speed up (MPE can help), No global error estimation
Empirical Balanced Truncation	Good approximation, Different inputs/initial values are possible	Most expensive model extraction, No speed-up, No global error estimation, Only applicable to the systems with very moderate size

Chapter 5

Reduced Macromodels of Massively Coupled Interconnect Structures via Clustering

There are challenging issues that arise in the model order reduction of networks with large number of input/output terminals. The direct application of the conventional Model Order Reduction (MOR) techniques on a multiport network often leads to inefficient transient simulations due to the large and dense reduced models. This chapter explains the details of a new, robust and practical algorithm to address this prohibitive issues.

5.1 Introduction

As signal rise times drop into the sub-nanosecond range, interconnect effects such as ringing, signal delay, distortion, and crosstalk can severely degrade the signal integrity. To provide sufficient accuracy, these effects must be captured by appropriate models and included during simulations. However, simulation of MTLs suffers from the major difficulties of excessive CPU time and mixed frequency/time problem. This is because, MTLs are best described in the frequency-domain whereas SPICE-like circuit simulators are mainly based on time-domain ODE formulations/solutions. To address these difficulties, several

techniques have been proposed in the literature, such as the ones based on waveform-relaxation [223–225] or macromodeling [226–230] approaches. In the case of waveform-relaxation based approaches, the input/output terminations as well as the input stimuli of the circuit are part of the simulation process. If the terminations or the stimuli are changed, then the entire simulation process including the waveform relaxation part has to be repeated to obtain the new results. In contrast, the macromodelling approach is independent of the terminations or the stimuli (i.e., the developed macromodel can be used in conjunction with any termination or the stimuli and the macromodel generation part does not need to be repeated every time). However, in order to preserve the accuracy of these macromodels over a large bandwidth, order of the resulting macromodels may typically end up being high. This problem (high-order) is further worsened in the presence of large number of coupled lines. Consequently, direct utilization of these macromodels in the simulation process is not practically efficient, as it leads to prohibitively excessive CPU time requirements. To improve the efficiency of simulations, the order of discretized models needs to be reduced while ensuring that the resulting downsized model can still sufficiently preserve the important physical properties of the original system. To serve this purpose, several numerically stable techniques based on implicit moment matching and congruence transformation (*cf.* Chapter-3) can be found in the literature.

Generally, by applying any of the above reduction techniques for networks with a small number of ports, reduced models can be obtained with sizes much smaller than the original circuit. However, as the number of ports of a circuit increases (as in the case of large bus structures), the size of reduced models also grows proportionally. As a result, use of these reduced models degrades the efficiency of transient simulations [65], significantly undermining the advantages gained by model order reduction techniques. This is because, to achieve a desired (predefined) accuracy, for every increase in the number of ports, the order of the reduced system should be increased proportional to the number of block moments.

Hence, the order of the model depends not only on the order of approximation (number of the block moments), but also on the number of the ports. Moreover, in the reduced-order model, the number of non-zero entries is also increasing rapidly with the number of the ports [231]. Therefore, the equations describing the reduced model are generally denser than the original system representation.

Recently, several attempts have been made to confront this problem via port-compression [231–236]. Early studies in [231, 232] reveal that, there may exist a large degree of correlation between various input and output terminals. Incorporating this correlation information in the matrix transfer function at the I/O ports of the reduced model during the model-reduction process became the common theme in the existing terminal-reduction methods. However, the major difficulty in port-compression algorithms such as SVD MOR [231] and RecMOR [232] is that the correlation relationship is frequency-dependent and in many cases is also input-dependent. As a consequence, such a reduction can lead to accuracy loss. To address this issue, foundation and initial results of a general clustering algorithm have been presented by the authors of this manuscript in [237], where a flexible scheme that consists of multi-input clusters was used. Later, in [238], Zhang et al. presented a similar idea based on splitting the system into subsystems, with each subsystem excited by a single input signal. From the conceptual point of view, the algorithm of [238], which requires constraining each subsystem to be single input as well as imposing the condition that the reduced subsystem to be of equal size can be considered as special case of the algorithm in [237].

A novel algorithm is presented in this chapter for efficient reduction of linear networks with large number of terminals. The new method while exploiting the applicability of the superposition paradigm [5, 239] to the analysis of massively coupled interconnect structures [237], proposes a reduction strategy based on flexible clustering of the transmission lines in the original network to form individual subsystems. Each subsystem consists of all the

lines in the interconnect structure where only a subset of the lines act as the aggressor (active) lines at a time. The overall reduced model is constructed by properly combining these reduced submodels based on the superposition principle. As a result, the contribution of the inputs of each cluster is included in evaluating the behavior of all the other clusters. The reduced submodel is obtained by applying the order-reducing projection to subsystems containing a dedicated cluster of active lines. The new contributions of this work include establishing several important properties of the reduced-order model, including a) stability b) block-moment matching properties and c) improved passivity. It is to be noted that, the flexibility in forming multi-input clusters with different sizes that was provided by [237] (unlike [238], which was limited to single input and subsystems of equal size) proved to be of significant importance while establishing the block-diagonal dominance and passivity-adherence of the reduced-order macromodel.

An important advantage of the proposed algorithm is that, for multiport interconnect networks, it yields reduced-order models that are sparse and block diagonal. The proposed algorithm is not dependent on the assumption of certain correlation between the responses at the external ports; thereby it is input-waveform and frequency independent. Consequently, it overcomes the accuracy degradation normally associated with the low-rank approximation based terminal reduction techniques [231–236].

5.2 Background and Preliminaries

This section first presents a brief overview of time-domain realization for multi-input and multi-output (MIMO) dynamical systems. The equations for the realization of interest are reviewed in the descriptor form as they appear in modified nodal analysis (MNA) matrix formulation [21, 37–39, 167]. Also, the complex-valued matrix transfer function representation for the systems will be presented followed by the definition of the corresponding

frequency-domain block-moments, as it will be useful in the later part of this paper when elaborating on the properties of the proposed method. We also briefly review PRIMA [85] as an example of order reduction technique based on congruence transformation and its moment matching property as relevant to the proposed algorithm here.

5.2.1 Formulation of Circuit Equations

Let the time-domain modified nodal analysis (MNA) matrix formulation for a linear RLC MIMO circuits be represented as:

$$\Psi : \begin{cases} \mathbf{C} \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t) \\ \mathbf{i}(t) = \mathbf{L} \mathbf{x}(t), \end{cases} \quad (5.1a)$$

$$(5.1b)$$

where $\mathbf{C}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are susceptance and conductance matrices, respectively, $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of MNA variables (the nodal voltages and some branch currents) of the circuit. Also, $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{L} \in \mathbb{R}^{p \times n}$ are the input and output matrices, associated with m inputs and n outputs, respectively.

Applying Laplace transformation to the dynamic equations (5.1a) and output equations (5.1b), the corresponding frequency-domain representation is given by:

$$\Psi : \begin{cases} \mathbf{C} s \mathbf{X}(s) + \mathbf{G} \mathbf{X}(s) = \mathbf{B} \mathbf{U}(s) \\ \mathbf{I}(s) = \mathbf{L} \mathbf{X}(s), \end{cases} \quad (5.2a)$$

$$(5.2b)$$

where $\mathbf{X}(s) \in \mathbb{C}^n$, $\mathbf{U}(s) \in \mathbb{C}^m$ and $\mathbf{I}(s) \in \mathbb{C}^p$.

Combining (5.2a) and (5.2b), the corresponding complex-valued matrix transfer function in s -domain is obtained as

$$\mathbf{H}(s) = \mathbf{L} (\mathbf{G} + s \mathbf{C})^{-1} \mathbf{B}. \quad (5.3)$$

Let $s_0 \in \mathbb{C}$ be a properly selected expansion point such that the matrix pencil $(\mathbf{G} + s_0\mathbf{C})$ is nonsingular. Eq. (5.3) can be rewritten as

$$\mathbf{H}(s) = \mathbf{L} (\mathbf{G} + s_0\mathbf{C} + (s - s_0)\mathbf{C})^{-1} \mathbf{B} = \mathbf{L} (\mathbf{I} + (s - s_0)\mathbf{A})^{-1} \mathbf{R}, \quad (5.4)$$

where

$$\mathbf{A} \triangleq (\mathbf{G} + s_0\mathbf{C})^{-1} \mathbf{C} \quad (5.5)$$

and

$$\mathbf{R} \triangleq (\mathbf{G} + s_0\mathbf{C})^{-1} \mathbf{B}. \quad (5.6)$$

The matrix function $\mathbf{H}(s)$ in (5.3) can be expanded in Taylor series around s_0 as:

$$\mathbf{H}(s) = \mathbf{L} \sum_{j=0}^{\infty} (-1)^j \underline{\mathbf{M}}_j(s_0) (s - s_0)^j \quad (5.7)$$

where the j -th moment of the function at s_0 is defined as

$$\mathbf{M}_j(s_0) = \mathbf{L} \underline{\mathbf{M}}_j(s_0) = \mathbf{L} \mathbf{A}^j \mathbf{R}, \quad (\text{for all } j). \quad (5.8)$$

5.2.2 Model-Order Reduction via Projection

Any suitable projection based order-reduction method can be used in conjunction with the proposed method in this paper. Without loss of generality, we will use the PRIMA algorithm in the rest of this paper. However, we should emphasize that other reduction techniques can also be equally used without degrading the merits of our proposed method. A brief description of the PRIMA algorithm is given in this section.

By exploiting an orthogonal projection matrix $\mathbf{Q}$, a change of variable $\tilde{\mathbf{z}} = \mathbf{Q}^T \mathbf{x}$ is applied on (5.1) to find a reduced-order model based on a congruence transformation

as [85]:

$$\tilde{\Psi} : \begin{cases} \tilde{\mathbf{C}} \frac{d}{dt} \tilde{\mathbf{z}}(t) + \tilde{\mathbf{G}} \tilde{\mathbf{z}}(t) = \tilde{\mathbf{B}} \mathbf{u}(t) \\ \tilde{\mathbf{i}}(t) = \tilde{\mathbf{L}} \tilde{\mathbf{z}}(t) . \end{cases} \quad (5.9a)$$

$$(5.9b)$$

The reduced model while preserving the main properties of the original system provides an output $\tilde{\mathbf{i}}(t)$ that appropriately approximates the original response $\mathbf{i}(t)$. For the resulting macromodel in (5.9), the reduced MNA matrices are

$$\begin{aligned} \tilde{\mathbf{C}} &= \mathbf{Q}^T \mathbf{C} \mathbf{Q}, & \tilde{\mathbf{G}} &= \mathbf{Q}^T \mathbf{G} \mathbf{Q}, \\ \tilde{\mathbf{B}} &= \mathbf{Q}^T \mathbf{B}, \text{ and } & \tilde{\mathbf{L}} &= \mathbf{L} \mathbf{Q}. \end{aligned} \quad (5.10)$$

Orthogonal projection matrix $\mathbf{Q}_{n \times q}$ above is obtained using block Arnoldi process as an implicit moment matching method [64, 90] such that the q column vectors of $\mathbf{Q}$ spans the same space with $\mathcal{M}$ -block moments of system denoted by $\mathcal{K}_{\mathcal{M}}(\mathbf{A}, \mathbf{R})$ as the sequel shown below [82]

$$\text{colspan} \{ \mathbf{Q} \} = \mathcal{K}_{\mathcal{M}}(\mathbf{A}, \mathbf{R}) = \text{span} \{ \mathbf{R}, \mathbf{A}\mathbf{R}, \dots, \mathbf{A}^{(\mathcal{M}-1)}\mathbf{R} \}. \quad (5.11)$$

The reduced system (5.9) of order q preserves the first $\mathcal{M} = \lfloor q/m \rfloor$ block moments of the original network (5.1) [85]. This implies that, for the same model accuracy, increasing the number of ports directly leads to proportionally larger order for the reduced system.

5.3 Development of the Proposed Algorithm

In this section, details of the proposed clustering-based algorithm for macromodeling of multi-port, large-order dynamical linear systems with emphasis on massively coupled

transmission line structures are presented.

For a given N -conductor interconnect structure (Fig. 5.1), the associated time-domain modified nodal analysis (MNA) matrix formulation is presented in (5.1), where $m = p = 2N$. Accordingly, for large bus structures, the number of external (input/output) terminals to the network is proportionally large.

Figure 5.1: Reduced-modeling of multiport linear networks representing N -conductor TL.

5.3.1 Formulation of Submodels Based on Clustering

Let Ψ_i , $i = 1, 2, \dots, K$, represents the i -th cluster of the system in (5.1), where K is the total number of clusters (Fig. 5.2). Each cluster Ψ_i consists of a group of (α_i) active lines (with inputs) and $(N - \alpha_i)$ lines for which the inputs are disabled. It is to be noted that, clustering is performed such that none of any two clusters share a common input (or common active line), hence, $N = \frac{1}{2} \sum_{j=1}^K m_i$, where m_i is the number of the input(s) to Ψ_i . However, all clusters share the same $2N$ output terminals.

To identify the submodels using admittance (y -)parameters, the m_i inputs of Ψ_i are excited by voltage sources, while all other terminals are grounded. The corresponding output currents at all $2N$ terminals are noted. The system of MNA equations for this submodel can be written as

$$\Psi_i : \begin{cases} \mathbf{C} \frac{d}{dt} \mathbf{x}_i(t) + \mathbf{G} \mathbf{x}_i(t) = \mathbf{B}_i \mathbf{u}_i(t) \\ \mathbf{i}_i(t) = \mathbf{L} \mathbf{x}_i(t) . \end{cases} \quad (5.12a)$$

$$(5.12b)$$

Figure 5.2: Illustration of forming clusters of active and victim lines in a multiconductor transmission line system.

It is to be noted that, for the submodels Ψ_i , the same $\mathbf{C}$, $\mathbf{G}$, and $\mathbf{L}$ matrices are recycled from the original system (5.1). Also, $\mathbf{B}_i$ is only a selection of the columns from original $\mathbf{B}$ matrix. Hence, avoiding the repetitive stamping process for subsystems leads to a significant speed-up while constructing the reduced-order model.

Next, the order of each subnetwork can be reduced using a suitable projection based algorithm. Using (5.9) and (5.10), (5.12) is reduced in the form:

$$\hat{\Psi}_i : \begin{cases} \hat{\mathbf{C}}_i \frac{d}{dt} \mathbf{z}_i(t) + \hat{\mathbf{G}}_i \mathbf{z}_i(t) = \hat{\mathbf{B}}_i \mathbf{u}_i(t) \\ \hat{\mathbf{i}}_i(t) = \hat{\mathbf{L}}_i \mathbf{z}_i(t), \end{cases} \quad (5.13a)$$

$$(5.13b)$$

where the associated reduced MNA matrices are

$$\hat{\mathbf{C}}_i = \mathbf{Q}_i^T \mathbf{C} \mathbf{Q}_i, \quad \hat{\mathbf{G}}_i = \mathbf{Q}_i^T \mathbf{G} \mathbf{Q}_i,$$

$$\hat{\mathbf{B}}_i = \mathbf{Q}_i^T \mathbf{B}_i, \text{ and } \hat{\mathbf{L}}_i = \mathbf{L} \mathbf{Q}_i. \quad (5.14)$$

Using superposition, an approximant $\hat{\mathbf{i}}(t)$ for the original responses $\mathbf{i}(t)$ can be obtained as

$$\mathbf{i}(t) \approx \hat{\mathbf{i}}(t) = \sum_{i=1}^K \hat{\mathbf{i}}_i(t) = \sum_{i=1}^K \hat{\mathbf{L}}_i \mathbf{z}_i(t) = \mathbf{L} \sum_{i=1}^K \mathbf{Q}_i \mathbf{z}_i(t). \quad (5.15)$$

5.3.2 Formulation of the Reduced Model Based on Submodels

Based on (5.15), the reduced model of the original system is obtained by superposing K reduced submodels as

$$\hat{\Psi} : \begin{cases} \hat{\mathbf{C}} \frac{d}{dt} \mathbf{z}(t) + \hat{\mathbf{G}} \mathbf{z}(t) = \hat{\mathbf{B}} \mathbf{u}(t) \\ \hat{\mathbf{i}}(t) = \hat{\mathbf{L}} \mathbf{z}(t), \end{cases} \quad (5.16a)$$

$$(5.16b)$$

where $\mathbf{z}(t) = [\mathbf{z}_1(t), \dots, \mathbf{z}_K(t)]^T$ and the concatenated projection matrix $\hat{\mathbf{Q}}$ for the reduction process is defined as

$$\hat{\mathbf{Q}} \triangleq \text{blkdiag}(\mathbf{Q}_i), \quad \text{for } i = 1, \dots, K. \quad (5.17)$$

Using (5.15) and (5.17), the output matrix $\hat{\mathbf{L}}$ in (5.16b) is obtained as

$$\hat{\mathbf{L}} = \hat{\mathbf{L}} \hat{\mathbf{Q}}, \quad (5.18)$$

where

$$\hat{\mathbf{L}} \triangleq \left[\begin{array}{ccc} 1 & \dots & 1 \end{array} \right]_{1 \times K} \otimes \mathbf{L}, \quad (5.19)$$

the operator $\otimes$ denotes the Kronecker product of matrices and operator “*blkdiag*” forms a block-diagonal matrix with its operand matrices located along the diagonal. Similarly,

other system matrices for the resulting superposed reduced model (5.16) can be obtained as

$$\hat{\mathbf{C}} = \hat{\mathbf{Q}}^T \hat{\mathbf{C}} \hat{\mathbf{Q}}, \quad \text{where} \quad \hat{\mathbf{C}} \triangleq \mathbf{I}_{K \times K} \otimes \mathbf{C} \quad (5.20)$$

$$\hat{\mathbf{G}} = \hat{\mathbf{Q}}^T \hat{\mathbf{G}} \hat{\mathbf{Q}}, \quad \hat{\mathbf{G}} \triangleq \mathbf{I}_{K \times K} \otimes \mathbf{G}, \quad (5.21)$$

and

$$\hat{\mathbf{B}} = \hat{\mathbf{Q}}^T \hat{\mathbf{B}}, \quad \hat{\mathbf{B}} \triangleq \text{blkdiag}(\mathbf{B}_i), \quad (5.22)$$

for $i = 1, \dots, K$,

where $\mathbf{I}_{K \times K}$ signifies an identity matrix of size $(K \times K)$. The resulting reduced model in (5.16) is of the size $(q \times q)$ (as in (5.9)), however, with the important advantages of being block diagonal and sparse.

- *MNA formulation of linear subnetwork π containing the reduced macromodel $\hat{\Psi}$:*

As shown in Fig. 5.3, the resulting model (5.16) can be embedded in a design consisting of surrounding lumped RLC elements. To simulate the whole circuit, the equations of the (embedded) reduced model (5.16) are combined with the MNA equations of the rest of the circuit in subnetwork π . Having realizations in a descriptor form with real matrices, the resulting reduced macromodel $\hat{\Psi}$ is directly stamped into the MNA matrix as

$$\begin{bmatrix} \left(\mathbf{G}_\pi + \mathbf{C}_\pi \frac{d}{dt} \right) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & -\hat{\mathbf{L}} \\ \mathbf{0} & -\hat{\mathbf{B}} & \mathbf{0} & \left(\hat{\mathbf{G}} + \hat{\mathbf{C}} \frac{d}{dt} \right) \end{bmatrix} \begin{bmatrix} \mathbf{x}_\pi \\ \mathbf{v}_P \\ \mathbf{i}_P \\ \mathbf{z} \end{bmatrix} - \begin{bmatrix} \mathbf{J}_\pi \\ \mathbf{J}_P \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \mathbf{0}. \quad (5.23)$$

Figure 5.3: Linear (RLC) subcircuit π accompanied with the reduced model $\hat{\Psi}$.

In (5.23), $\mathbf{v}_P$ and $\mathbf{i}_P$ respectively are the voltages and currents at the ports of the reduced model $\hat{\Psi}$, interfacing with the rest of the subnetwork π . $\mathbf{x}_\pi$ contains all the voltages at the nodes of subnetwork π followed by the extra variables (currents) associated with the voltage sources and inductors in the subnetwork. Vectors $\mathbf{J}_\pi$ and $\mathbf{J}_P$ denote the sources connected to the subnetwork π and the ports of macromodel, respectively. $\mathbf{I}$ is an identity matrix of size $2N$. Here, $\mathbf{z}$ denotes the extra variables that are introduced from the inclusion of the reduced model into the circuit. Also, $\mathbf{C}_\pi$ and $\mathbf{G}_\pi \in \mathbb{R}^{n_\pi \times n_\pi}$ are susceptance and conductance matrices, respectively, describing the lumped elements of subnetwork π .

- *Formulation of overall circuit including nonlinear subnetwork Φ :*

In the presence of nonlinear elements, the nonlinear subnetwork Φ should be also included in the time-domain MNA representation of the overall network and be simulated along with the rest. For simplicity, let the linear components be grouped into a single subnetwork π as shown in Fig. 5.4. Without loss of generality, the MNA equations for the network Φ can be written as [47]

$$\mathbf{G}_\Phi \mathbf{x}_\Phi(t) + \mathbf{C}_\Phi \frac{d}{dt} \mathbf{x}_\Phi(t) + \mathbf{L}_\Phi \mathbf{i}_\pi + \mathbf{F}(\mathbf{x}_\Phi(t)) - \mathbf{J}_\Phi = \mathbf{0}, \quad (5.24)$$

Figure 5.4: The overall network comprising the reduced model, embedded RLC subcircuit, and nonlinear termination.

where $\mathbf{F}(\mathbf{x}_\Phi(t)) \in \mathbb{R}^{n_\Phi}$ is nonlinear vector describing the nonlinear elements in Φ , $\mathbf{J}_\Phi(t)$ includes the independent sources to subnetwork Φ . Also, $\mathbf{i}_\pi$ denotes the port currents entering the linear subnetwork π and $\mathbf{L}_\Phi$ is a selector matrix that maps $\mathbf{i}_\pi$ to the vector of unknowns $\mathbf{x}_\Phi \in \mathbb{R}^{n_\Phi}$ in subnetwork Φ .

5.4 Properties of the Proposed Algorithm

In this section, important properties of the proposed macromodeling methodology are discussed.

5.4.1 Preservation of Moments

In this section, it will be shown that the proposed macromodeling algorithm preserves the first $\mathcal{M}$ block moments of the transfer-function matrix of the original system. This is the same number of moments which are matched in the conventional projection-based methods such as classical block Arnoldi reduction and PRIMA. For the purpose of proving the moment preservation property of the method, following definitions and theorems are developed. Applying Laplace transform to the circuit equations (5.12a) of each unreduced subsystem $\Psi_i(t)$, and assuming the initial condition $x(t_0) = 0$, we obtain the input-to-state transfer function for i^{th} cluster as

$$\underline{\mathbf{H}}_i(s) = (\mathbf{G} + s\mathbf{C})^{-1} \mathbf{B}_i. \quad (5.25)$$

Also, following the similar steps, an approximant for the transfer function in (5.25) can be obtained from the corresponding reduced subsystem $\hat{\Psi}_i(t)$ in (5.13) using the associated order-reducing projection matrix $\mathbf{Q}_i$ as

$$\hat{\underline{\mathbf{H}}}_i(s) = \mathbf{Q}_i \left(\hat{\mathbf{G}}_i + s\hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{B}}_i, \quad (5.26)$$

and the transfer function for the reduced model in (5.16) is

$$\hat{\mathbf{H}}(s) = \hat{\mathbf{L}} \left(\hat{\mathbf{G}} + s\hat{\mathbf{C}} \right)^{-1} \hat{\mathbf{B}}. \quad (5.27)$$

Considering the transfer functions in (5.3), (5.25) and their approximants in (5.27) and (5.26), respectively, the following theorems are developed.

Theorem 5.1. *Consider the input-to-state transfer function $\underline{\mathbf{H}}_i(s) = (\mathbf{G} + s\mathbf{C})^{-1} \mathbf{B}_i$, for a subsystem Ψ_i in (5.12), associated with the cluster of m_i inputs $\mathbf{u}_i$, while the other excitations are disabled. Also, consider the approximated input-to-state transfer function obtained from the corresponding reduced submodel $\hat{\Psi}_i(t)$ with its order-reducing projection matrix $\mathbf{Q}_i$ as $\hat{\underline{\mathbf{H}}}_i(s) = \mathbf{Q}_i (\hat{\mathbf{G}}_i + s\hat{\mathbf{C}}_i)^{-1} \hat{\mathbf{B}}_i$. The input-to-state transfer function for the original subsystem $\underline{\mathbf{H}}_i(s)$, and its approximant $\hat{\underline{\mathbf{H}}}_i(s)$ share the same first $\mathcal{M}_i = \lfloor q_i/m_i \rfloor$ block moments.*

Theorem 5.2. *The input-to-output transfer function for the original (unreduced) system $\mathbf{H}(s) = \mathbf{L}(\mathbf{G} + s\mathbf{C})^{-1} \mathbf{B}$ and its approximant $\hat{\mathbf{H}}(s) = \hat{\mathbf{L}}(\hat{\mathbf{G}} + s\hat{\mathbf{C}})^{-1} \hat{\mathbf{B}}$ from the proposed method share the same $\mathcal{M}$ first block moments, where $\mathcal{M} = \min_{i=1,\dots,K} (\mathcal{M}_i)$ and K is the number of the subsystems.*

The proof of Theorems 5.1 and 5.2 are given in Appendices C and D, respectively.

5.4.2 Stability

It is desirable and often crucial that reduced-order models inherit the essential properties of the original linear dynamical system. One such crucial property is stability [67]. We consider a macromodel resulting from the proposed algorithm (5.16) whose transfer function is of the form shown in (5.27). The poles of this transfer function are located where the kernel $(\hat{\mathbf{G}} + s\hat{\mathbf{C}})^{-1}$ is singular. Therefore, to describe the stability of the model, the spectrum of its matrix pencil $(\hat{\mathbf{G}} + s\hat{\mathbf{C}})$ should be considered. Given that we are modeling physical systems, it is assumed that the associated matrix pencil is regular (i.e. the kernel is singular only for finite number of values of $s \in \mathbb{C}$). These singularities can be found from the solution of a generalized eigenvalue problem as $\mathbf{G}\vec{\mathbf{X}} = \lambda\mathbf{C}\vec{\mathbf{X}}$, $\vec{\mathbf{X}} \neq \mathbf{0}$ [117]. It is well-known

that a macromodel is asymptotically stable if and only if all the finite eigenvalues of the associated matrix pencil lie in the open left half-plane [10].

Theorem 5.3. *If the chosen model-order reduction scheme applied to each subsystem preserves the stability of each subsystem reduced model, the diagonalized reduced-model for the overall system resulting from the proposed methodology (5.16) is asymptotically stable.*

Proof. To establish the stability property of the proposed method, the block diagonal structure of the matrix pencil $(\hat{\mathbf{G}} + s\hat{\mathbf{C}})$ for the resulting model (5.16) is considered. Using the block diagonal matrices $\hat{\mathbf{G}}$ in (5.20) and $\hat{\mathbf{C}}$ in (5.21), the block diagonal structure of the pencil will be obtained as

$$(\hat{\mathbf{G}} + s\hat{\mathbf{C}}) = \text{blkdiag} \left[\hat{\mathbf{G}}_1 + s\hat{\mathbf{C}}_1, \dots, \hat{\mathbf{G}}_i + s\hat{\mathbf{C}}_i, \dots, \hat{\mathbf{G}}_K + s\hat{\mathbf{C}}_K \right], \quad (5.28)$$

whose diagonally located blocks are the regular matrix pencil for the stable submodels. The spectrum of the diagonal pencil is the union of the spectra of blocks on diagonal as shown below

$$\lambda(\hat{\mathbf{G}} + s\hat{\mathbf{C}}) = \bigcup_{i=1}^K \lambda(\hat{\mathbf{G}}_i + s\hat{\mathbf{C}}_i). \quad (5.29)$$

Hence, the spectrum of the matrix pencil for the reduced system consists of the union of the complex numbers with negative real parts ($\in \mathbb{C}^-$). This explicitly proves, the reduced model in (5.16) is asymptotic stable. ■

5.4.3 Passivity

Another property that reduced-order models should inherit is passivity. It is important because, stable but non-passive models may lead to unstable systems when connected to other passive components. On the other hand, a passive macromodel, when terminated with any arbitrary passive load, always guarantees the stability of the overall network.

According to the positive-real lemma [10, 72, 240], a linear network is passive if its transfer-function matrix (in admittance or impedance form) is positive real. Strictly speaking, this requires that, for ensuring passivity of $\hat{\mathbf{H}}(s)$, the following conditions for passivity be fulfilled:

$$\left\{ \begin{array}{l} \hat{\mathbf{H}}(s) \text{ is defined and analytic in } \Re\{s\} > 0, \end{array} \right. \quad (5.30a)$$

$$\left\{ \begin{array}{l} \hat{\mathbf{H}}^*(s) = \hat{\mathbf{H}}(s^*), \end{array} \right. \quad (5.30b)$$

$$\left\{ \begin{array}{l} \Phi(s) = \left(\hat{\mathbf{H}}(s) + \hat{\mathbf{H}}^H(s) \right) \geq 0 \quad \forall s \in \mathbb{C} : \Re\{s\} > 0. \end{array} \right. \quad (5.30c)$$

Being an asymptotically stable reduced model (proved in Theorem 5.3), the entire spectrum of the regular matrix pencil $\{\hat{\mathbf{C}}, \hat{\mathbf{G}}\}$ is confined to the left-half in complex plane (LHP). Locating the singularities (poles) of the transfer function in LHP ensures $\hat{\mathbf{H}}(s)$ to be analytical at right-half plane (RHP) and holds (5.30a). The criterion (5.30b) equivalently states that, $\hat{\mathbf{H}}(s)$ should be a real-valued matrix for any real $s > 0$. This condition trivially establish for the transfer-function in (5.27). To investigate the positive semidefinite-ness of $\Phi(s)$ in (5.30c), it is to be noted that, $\hat{\mathbf{C}}$ and $(\hat{\mathbf{G}} + \hat{\mathbf{G}}^T)$ are symmetric non-negative definite matrices. However, considering $\hat{\mathbf{L}}$ defined in (5.19) and $\hat{\mathbf{B}}$ in (5.22), it is noted that, $\hat{\mathbf{L}} \neq \hat{\mathbf{B}}^T$. Due to the latter fact, the associated $\Phi(s)$ may not always be ensured as positive semidefinite at all frequencies throughout the frequency band (of interest). Strictly speaking, the proposed method generates accurate and stable reduced macromodels, nevertheless the passivity of the resulting macromodel is not always guaranteed.

We outline an approach to overcome this issue based on exploiting the relatively weak coupling between the clusters of transmission lines. For this purpose, it is necessary that partitioning the multiconductor transmission line system is done in such a way that every group of (α_i) strongly coupled lines are grouped as active lines in an i -th cluster as illustrated in Fig. 5.5, where the number of the lines in a group α_i can even be as low as one.

* Aggregation Index (A.I.) for the bundle of strongly coupled lines

Figure 5.5: Illustration of strongly coupled lines bundled together as active lines in the clusters.

Following a clustering scheme shown in Fig. 5.5, the admittance parameter matrix $\hat{\mathbf{H}}(s)$ from the proposed reduction algorithm is block-partitioned into $K \times K$ ($1 \leq K \leq N$) submatrices $\hat{\mathbf{H}}_{ij}(s)$, as shown in (5.31):

$$\hat{\mathbf{H}}(s) = \begin{bmatrix} \hat{\mathbf{H}}_{11} & \hat{\mathbf{H}}_{12} & \cdots & \hat{\mathbf{H}}_{1K} \\ \hat{\mathbf{H}}_{21} & \hat{\mathbf{H}}_{22} & \cdots & \hat{\mathbf{H}}_{2K} \\ \vdots & \vdots & & \vdots \\ \hat{\mathbf{H}}_{K1} & \hat{\mathbf{H}}_{K2} & \cdots & \hat{\mathbf{H}}_{KK} \end{bmatrix}. \quad (5.31)$$

In (5.31), each one of the diagonally located blocks $\hat{\mathbf{H}}_{ii}(s)$ is a $(2\alpha_i \times 2\alpha_i)$ reduced transfer-function submatrix characterizes the behavior of i -th cluster of (α_i) lines in the subsystem $\hat{\Psi}_i$ (5.13) at its $2\alpha_i$ ports. Each off-diagonal block $\hat{\mathbf{H}}_{ij}(s)$ ($i \neq j$) represents the coupling effect from the cluster i (including α_i lines) to the cluster j (including α_j other lines), and so does its counterpart $\hat{\mathbf{H}}_{ji}(s)$ in reverse direction.

The Hermitian matrix $\Phi(s)$ in (5.30c) can also be considered with the same block structure as $\hat{\mathbf{H}}(s)$ in (5.31):

$$\Phi(s) = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1K} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{K1} & \Phi_{K2} & \cdots & \Phi_{KK} \end{bmatrix}. \quad (5.32)$$

In (5.31), the relative weak inter-coupling among different clusters (which contain strongly coupled lines) ensures that,

$$\|\hat{\mathbf{H}}_{ii}\| \gg \|\hat{\mathbf{H}}_{ij}\| \quad \forall i, j \in \{1, 2, \dots, K\}, \quad (5.33)$$

and

$$\|\Phi_{ii}\| \gg \|\Phi_{ij}\| \quad \forall i, j \in \{1, 2, \dots, K\}, \quad (5.34)$$

where $\|\cdot\|$ is some consistent matrix norm (such as the 2-norm).

Definition 5.1. For the $2N \times 2N$ matrix $\Phi(s)$ partitioned as in (5.32) with nonsingular

diagonal submatrices $\Phi_{ii}(s)$ if

$$\sum_{j=1, j \neq i}^K \|\Phi_{ii}^{-1}(s) \Phi_{ij}(s)\| \leq 1, \quad (5.35)$$

at any frequency $s \in \mathbb{C}, \Re\{s\} > 0$, inequality holds strictly for at least one $1 \leq i \leq K$; then $\Phi(s)$ is “block-digonally dominant”, relative to the partitioning in (5.32).

A proper clustering, as explained above, attributes the properties of the block-diagonal dominance to the Hermitian matrix Φ (5.32).

Proposition 5.1. *If a passive reduced-order macromodeling scheme is applied to obtain subsystem reduced models, any diagonal block $\hat{\mathbf{H}}_{ii} \in \mathbb{C}^{\alpha_i \times \alpha_i}$ for $i = 1, \dots, K$ in the overall system transfer function (5.31) is positive real.*

Proof of the above proposition is straight-forward considering that, $\hat{\mathbf{H}}_{ii}$ defines the inputs to outputs transfer function at the terminals of the α_i active lines in the i -th cluster Ψ_i . Hence, it is $\mathbf{L}_{ii} = \mathbf{B}_{ii}^T$, where $\mathbf{B}_{ii} = \mathbf{B}_i$.

According to this proposition, the square blocks located on the diagonal of Φ in (5.32), defined as

$$\Phi_{ii}(s) = \mathbf{H}_{ii}(s) + \mathbf{H}_{ii}^H(s), \quad \forall i \in \{1, 2, \dots, K\}, \quad (5.36)$$

are positive definite.

The following property will be later used in the proof of the Theorem 5.4.

Proposition 5.2. *For any diagonal block in (5.32) as $\Phi_{ii} \in \mathbb{C}^{2\alpha_i \times 2\alpha_i}$ being Hermitian positive semidefnite and any arbitrary real number $\lambda < 0$, we have*

$$\|(\lambda \mathbf{I} - \Phi_{ii})^{-1} \Phi_{ij}\| \leq \|\Phi_{ii}^{-1} \Phi_{ij}\|, \quad \text{for } j = 1, \dots, K, \quad j \neq i \quad (5.37)$$

where $\|\cdot\|$ is the two-norm (spectral norm) of the matrix.

This can be proved by following the similar steps of the proof for lemma 3.9 in [241]. Considering proposition 5.1, the following Theorem 5.4 is defined.

Theorem 5.4. *Let the block-partitioned Hermitian matrix in (5.32) $\Phi = [\Phi_{ij}] \in \mathbb{C}^{2N \times 2N}$ for $i, j \in \{1, 2, \dots, K\}$ be block-diagonally dominant. Then for any eigenvalue λ of Φ , we have $\lambda > 0$.*

Proof. To proof by contradiction, assume that, at any given complex frequency s , $\Phi(s)$ has an eigenvalue $\lambda < 0$. Then, the matrix $(\lambda \mathbf{I} - \Phi(s))$ is singular and hence, $\det(\lambda \mathbf{I} - \Phi(s)) = 0$, where $\det(\cdot)$ is determinant of a matrix. Thus $\lambda \mathbf{I} - \Phi_{ii}(s)$ is non-Hermitian. Any matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is called non-Hermitian negative definite if $\Re e(\mathbf{x}^H \mathbf{A} \mathbf{x}) < 0$, for any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq \mathbf{0}$. From the definition, it is

$$\begin{aligned} \Re e(\mathbf{x}^H (\lambda \mathbf{I} - \Phi_{ii}(s)) \mathbf{x}) &= \Re e(\lambda \mathbf{x}^H \mathbf{x} - \mathbf{x}^H \Phi_{ii}(s) \mathbf{x}) = \\ \Re e(\lambda \|\mathbf{x}\|^2 - \mathbf{x}^H \Phi_{ii}(s) \mathbf{x}) &= \lambda \|\mathbf{x}\|^2 - \Re e(\mathbf{x}^H \Phi_{ii}(s) \mathbf{x}) < 0. \end{aligned} \quad (5.38)$$

The inequality in (5.38) is established considering that, $\lambda \|\mathbf{x}\|^2 < 0$ and $\Re e(\mathbf{x}^H \Phi_{ii}(s) \mathbf{x}) \geq 0$. This verifies $\lambda \mathbf{I} - \Phi_{ii}(s)$ as non-Hermitian negative definite matrix for all $i = 1, \dots, K$. As a result, $\mathcal{D}(s, \lambda) = \text{blkdiag}((\lambda \mathbf{I} - \Phi_{11}(s)), \dots, (\lambda \mathbf{I} - \Phi_{KK}(s)))$ is a (non-Hermitian) negative definite and hence, a nonsingular matrix. considering the nonsingularity of $\mathcal{D}(s, \lambda)$ a matrix $\mathcal{A}(s, \lambda)$ can be defined as

$$\mathcal{A}(s, \lambda) \triangleq \mathcal{D}^{-1}(s, \lambda) (\lambda \mathbf{I} - \Phi(s)) = [\mathcal{A}_{ij}], \quad i, j = 1, 2, \dots, K \in \mathbb{C}^{2N \times 2N}, \quad (5.39)$$

where

$$\mathcal{A}_{ij} = \begin{cases} \mathbf{I}_{2\alpha_i}, & i = j, \\ \left(\lambda \mathbf{I} - \Phi_{ii}(s) \right)^{-1} \Phi_{ij}(s), & i \neq j, \end{cases} \quad (5.40a)$$

$$(5.40b)$$

and $\mathbf{I}_{2\alpha_i}$ denotes $(2\alpha_i \times 2\alpha_i)$ identity matrix. Since $\Phi(s)$ is also block-diagonally dominant, from the definition in 5.1, it is

$$\sum_{\substack{j=1 \\ i \neq j}}^K \|\Phi_{ii}^{-1} \Phi_{ij}\| \leq 1, \quad \text{for } i = 1, \dots, K \quad (5.41)$$

where, inequality strictly holds at least for one i . From (5.41) and (5.37) in proposition 5.2, it is concluded

$$\sum_{\substack{j=1 \\ i \neq j}}^K \|(\lambda \mathbf{I} - \Phi_{ii})^{-1} \Phi_{ij}\| \leq 1, \quad \text{for } i = 1, \dots, K \quad (5.42)$$

that the inequality strictly holds at least for one i . Equation (5.42) explicitly implies that matrix $\mathcal{A}(s, \lambda)$ is also block-diagonally dominant (based on the assumption of $\lambda < 0$). Using Theorem 2 in [242] (or equivalently [243, corollary 3.1], that states “A block-diagonally dominant matrix is non-singular”; it is said that $\mathcal{A}(s, \lambda)$ is non-singular. Using Proposition 2.7.3 and corollary 2.7.4 in [74], it is

$$\det(\mathcal{D}^{-1}(s, \lambda)(\lambda \mathbf{I} - \Phi(s))) = \det(\mathcal{D}^{-1}(s, \lambda)) \times \det(\lambda \mathbf{I} - \Phi(s)) \neq 0, \quad (5.43)$$

that implies $\det(\lambda \mathbf{I} - \Phi(s)) \neq 0$ (a non-singular matrix). This contradicts the singularity of $(\lambda \mathbf{I} - \Phi(s))$ and implies that negative λ can not be an eigenvalue for $\Phi(s)$. Thus, for any eigenvalue λ of a Hermitian $\Phi(s)$ (5.32), we have $\lambda \geq 0$. ■

Theorem 5.4 establishes the block-diagonal dominance as the sufficient condition to ensure Φ to be positive definite, and hence, passivity of $\hat{\mathbf{H}}(s)$.

It is to be noted that, the definition 5.1, we used to develop Theorem 5.4 is based on the

most relaxed condition for block-diagonal dominance that defines a more general class of matrices. It is sometimes referred to as “weak” block-diagonal dominant [243] in the linear algebra context.

Block-diagonal dominance is more relaxed criterion compared to *diagonal dominance*. A matrix can be block-diagonally dominant without being diagonally dominant. As an example, please see [244]; Eq. (2.6).

5.4.4 Guideline for Clustering to Improve Passivity

In Sec. 5.3 the only constraint for the clustering was stated that, the sequel of input matrices as $[B_1, \dots, B_K]$ needs to have uncorrelated columns. This flexible clustering scheme allows the passivity preservation to be considered as the primary criterion when grouping the active lines and deciding the clusters.

Additionally, Theorem 5.4 is utilized as a guideline for proper clustering, according to which the bundling of strongly coupled lines as active lines in each cluster improves the passivity of the resulting macromodel. Strong coupling between the transmission lines can be decided by investigating the per-unit-length (PUL) parameter matrices and by comparing the norm of the off-diagonal block matrices in (5.31) with the norm of diagonal block matrices. We use the PUL matrices for initial partitioning followed by a second stage where we check the relative norm of the off-diagonal blocks of the admittance matrix (at the highest frequency of interest).

To illustrate this, Fig. 5.6 demonstrates the minimum (smallest) eigenvalue of the $\Phi(s)$ in (5.30c) ($\lambda_{\min}(s_j) = \min_{i=1}^{2N} \lambda_i(\Phi(s_j))$) for a structure of 32 coupled interconnect when each one of the 32 clusters has only one active line (details of the interconnect structure can be found in the Sec. 5.5). Fig. 5.7 depicts that, there are eigenvalues that extend to the negative region, indicating passivity violation. Following the proposed approach,

Figure 5.6: The frequency-spectrum of the minimum eigenvalue of $\Phi(s)$ containing 32 clusters.

interconnects were clustered into 16 subnetworks each including two active adjacent lines which are closely coupled. As shown in Figures 5.8 and 5.9, all eigenvalues of $\Phi(s)$ in (5.30c) are nonnegative and the passivity criterion is satisfied. Grouping the strongly coupled lines trades the sparsity (hence, efficiency) of the reduced model for the passivity. Proceeding with this approach will inevitably lead to passivity preservation. However, once the clusters of active lines reach a certain size, the advantage of using the proposed method may be undermined. In such extremes, the reduced model will be promisingly passive while efficiency will be reduced, lower bounded to the efficiency expected from a conventional reduction technique (such as PRIMA).

It is to be noted that, the passivity preservation by clustering can not be prescribed in all practical cases, when a certain level of efficiency needs to be insured. However, passivity

Figure 5.7: The enlarged region near the x-axis of Fig. 5.6 (illustrating eigenvalues extending to the negative region, indicating passivity violation).

adherence of the model, obtained through the proposed clustering scheme, will make it a good candidate for passivity enforcement process using any of the well-known enforcement techniques [245–247] without sacrificing much of the accuracy. Thereby, we can optimally conserve efficiency and accuracy of the reduced model, beside passivity assurance.

5.5 Numerical Examples

In this section, numerical results are presented to demonstrate the validity and accuracy of the proposed methodology. The CPU times reported here correspond to a PC platform with 4GB RAM and 2GHz Intel processor, executed in the Matlab 7.11.0 (R2010b) environment.

Figure 5.8: Spectrum of $\Phi(s)$ versus frequency with proper clustering to improve passivity (no passivity violations observed).

5.5.1 Example I

In this example, we consider a circuit containing a 32-coupled transmission line bus with the length of 10cm. The extracted line parameters are based on the data obtained from [248].

The multiconductor transmission line (MTL) subcircuit shown in Fig. 5.10 has 64 terminals through which it is connected to the rest of the circuit. Hence, the dimension of the matrix $\hat{\mathbf{H}}(s)$ in (5.31) is 64×64 . The five input voltage sources, connected to the near-ends (left side) of the lines 1, 8, 16, 24, and 32 are trapezoidal pulses with rise/fall times of

Figure 5.9: The frequency-spectrum of the minimum eigenvalue of $\Phi(s)$ with clustering to improve passivity behavior (no passivity violations observed).

0.2ns, delay of 2ns and pulse width of 5ns. The transmission lines in the original coupled network are discretized using conventional uniform lumped segmentation [249]. The size of the original network constructed for the MTL structure (excluding the peripheral components) is 29195×29195 . Using conventional PRIMA, matching the 40 first block Arnoldi moments of the original 64-ports MTL structure leads to a dense reduced matrix of size 2560×2560 , whose sparsity pattern is shown in Fig. 5.11. On the other hand, the clustering scheme in the proposed algorithm results in thirty-two decoupled matrices of size 80×80 each. By combining these submatrices associated with the subsystems, a block diagonal matrix realization for the entire network is obtained. The resulting MNA system matrix is 97% sparse (see Fig. 5.12) and consists of thirty-two block matrices along

Figure 5.10: 32 conductor coupled transmission line network with terminations considered in the example.

the diagonal each of size 80×80 . This represents significant sparsity advantage compared to using the conventional PRIMA algorithm. To demonstrate the accuracy of the proposed method, Figures 5.13–5.15 show sample comparisons of time-domain responses. In the graphs, the time-domain results obtained from applying the proposed method are compared to the responses from the original network as well as the conventional PRIMA reduced model. As seen from the plots, all these responses are in excellent agreement. Table 5.1 compares the CPU time expense for the transient simulation of the original system versus the proposed and conventional PRIMA based reduced macromodels. As the table depicts, applying a conventional MOR technique to this multiport system leads to a macromodel which is prohibitively expensive; even when compared to the unreduced circuit. In contrast, using the proposed algorithm, a speed-up of 15.5 compared to PRIMA was achieved. It was also observed that the speed-up ratio increases with increasing the

Figure 5.11: Sparsity pattern of reduced MNA equations using conventional PRIMA (dense).

Figure 5.12: Sparsity pattern of reduced MNA equations using the proposed method.

number of the lines.

Table 5.1: CPU-cost comparison between original system, PRIMA and proposed method.

	Original	PRIMA	Proposed
Total CPU-time (sec.)	645.9	1730	111.7

Figure 5.13: Transient responses at victim line near-end of line#2.

5.5.2 Example II

The idea of passivity preservation using the proposed flexible clustering is further investigated in this example. For the purpose of illustration, we consider an interconnection structure consisting of nine coupled lines of length $d = 2.54 \text{ cm}$ (see Fig 5.16). The RLGC parameters of the lines were calculated using the field solver in HSPICE [250].

First, the interconnect structure was clustered into nine subsystems with one active line in each as shown in Fig. 5.17. The minimum (smallest) eigenvalue of Hermitian matrix $\Phi(s)$ in (5.30c) as function of frequency ($\lambda_{\min}(s_j) = \min_{i=1}^{2N} \lambda_i(\Phi(s_j))$) is shown in Fig. 5.18 which depicts the presence of negative eigenvalues indicating passivity violation. This is also illustrated in Fig. 5.19 which shows all the negative eigenvalues of $\Phi(s)$ within the frequency spectrum of interest. Following the proposed approach, the interconnect structure was clustered into three subnetworks each including three active lines as shown

Figure 5.14: Transient responses at victim line near-end of line#12.

in Fig. 5.20. This clustering was decided by examining the physical geometry of the structure (Fig. 5.16) and the numerical values of the PUL parameters. Also, it was verified through examining the norm of the mutual admittances. For this clustering arrangement, all the eigenvalues of $\Phi(s)$ are shown in Fig. 5.21. Fig. 5.22 shows that the minimum eigenvalues (and hence, all other eigenvalues) are nonnegative (i.e., satisfying the passivity criterion).

Figure 5.15: Transient responses at victim line far-end of line#31.

Figure 5.16: Cross sectional geometry (Example 2).

Figure 5.17: Interconnect structure with nine clusters (Example 2).

Figure 5.18: Minimum eigenvalue of $\Phi(s)$ while using 9 clusters (each cluster with nine lines while one of them acting as an active line).

Figure 5.19: Negative eigenvalue of $\Phi(s)$ (using the 9-cluster approach).

Figure 5.20: Illustration of the interconnect structure grouped as three clusters (each cluster with nine lines while the three of the strongly coupled lines in each of them acting as active lines [shown in red color]).

Figure 5.21: Eigenvalue of $\Phi(s)$ (using 3 clusters based on the proposed flexible clustering approach).

Figure 5.22: Minimum eigenvalues of $\Phi(s)$ (using 3 clusters based on the proposed flexible clustering approach).

Chapter 6

Optimum Order Estimation of Reduced Linear Macromodels

In Chapter 3, some of the well-known linear model reduction methods were reviewed. Also, it was stated that, presently, a rich body of literature is available covering the linear MOR techniques. However, for all of these methods, the selection of order is an important issue. This chapter explains the details of a novel algorithm for optimal-order determination for the reduced linear macromodels.

6.1 Introduction

An important and practical common problem in prominently used order-reduction techniques is that of “selection of order” for the reduced model. The proper choice of order for a macromodel based approximation is important in terms of achieving the pre-defined accuracy, while not over-estimating the order, which otherwise can lead to inefficient transient simulations. This (an optimum order) becomes even more important, if the reduced macromodel is going to be used repeatedly as part of a larger simulation task such as in

the case of statistical analysis, optimization, design centering, etc. In this case, the unnecessary computational cost during repetitive simulations/optimization due to overestimating the order of the reduced-model can significantly exceed the computational cost of optimally pre-estimating the order. Current techniques for predicting an optimum order for an approximation *a-priori* is generally heuristic in nature.

This chapter presents a novel algorithm to obtain an optimally minimum order for a reduced model under consideration. The proposed methodology is based on the idea of monitoring the behavior of the projected trajectory in the reduced space [251, 251]. To serve this purpose, a mathematical algorithm is devised to observe the behavior of near neighboring points, lying on the projected trajectory, when increasing the dimension of a reduced-space. The order is determined such that the projected trajectory is unfolded properly in the reduced space, while monitoring the count of the "False Nearest Neighbor (FNN)" points on the projected trajectory. The reduced model in this optimally reduced subspace preserves the major dynamical properties of the original system.

6.2 Development of the Proposed Algorithm

6.2.1 Preliminaries

A set of differential algebraic equations can be used to represent the dynamical behavior of the system states [2, 21–23]. For electrical networks these equations are directly obtained using the modified nodal analysis (MNA) matrix formulation [37–39, 167] in the form:

$$\mathbf{C} \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t) \quad (6.1)$$

$$\mathbf{i}(t) = \mathbf{L} \mathbf{x}(t), \quad (6.2)$$

where $\mathbf{C}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are susceptance and conductance matrices, respectively, $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of MNA variables (the nodal voltages and some branch currents) of the circuit. $\mathbf{B}$ and $\mathbf{L}$ represent the input and output matrices, respectively.

The key idea in subspace projection-based model order reduction techniques is to project the original n -dimensional state space to a m -th order (e.g.: Krylov) subspaces, where practically $m \ll n$. This reduction process requires creation of a projection operator $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m] \in \mathbb{R}^{n \times m}$ such that the trajectory in the original space can be properly projected to a reduced subspace as $\mathbf{z}(t) \triangleq \mathbf{Q}^T \mathbf{x}(t)$. As a result, a linear system which is of much smaller order is obtained by a variable change as $\mathbf{x} = \mathbf{Q} \mathbf{z}$ [64, 89, 90].

The objective of the proposed method is to determine the optimum dimension for the reduced subspace while preserving desired accuracy. To serve this purpose, the “false nearest neighbors (FNN)” concept [252–255] is adopted. From a geometrical perspective, the variables set $\{\mathbf{x}_i(t) : \text{for } i = 1, 2, \dots, n\}$ is used as a coordinate system to define an n -dimensional space. Therefore, the response at each time instant t_j , represented by $\mathbf{x}(t_j) = \{x_1(t_j), x_2(t_j), \dots, x_n(t_j)\}$ ($t_j \in \Lambda_t$), defines a point in this response space. Consider the illustrative Fig. 6.1; starting from a given initial condition $\mathbf{x}(t_0)$, as the circuit’s response evolves with time, the point moving through the response space traces out a curve. Mathematically, the solution curve is a real-valued continuously-differentiable function (taken to be $\mathcal{C}^n$) [22] from an open interval $\Lambda_t \subset \mathbb{R}_+$ into the response space $\subseteq \mathbb{R}^n$. Such a curve as a flow of the states for all subsequent time is the key notion in the description of the behavior of dynamical circuits. We consider this trajectory curve whose definition is given below as a geometric model to study the dynamic behavior of the circuit.

Definition 6.1. A time-parameterized path in the multidimensional response space of a system, defined by $\mathbf{x}(t)$ for $t \geq t_0$ is referred to as trajectory (curve) of the system.

Using a projection operator $\mathbf{Q}_{n \times m}$, a reduced subspace $\mathbb{R}^m$ is defined with coordinates

Figure 6.1: Any state corresponding to a certain time instant can be represented by a point (e.g. A, N, E and F) on the trajectory curve (T) in the variable space.

that are linear combinations of the original coordinates; i.e. $z_i(t) = \sum_{j=1}^n q_{ji} x_j(t)$, for $i = 1, 2 \dots m$, where $m \ll n$. An image of the trajectory curve in the low dimensional subspace is obtained through the point-wise projection of the original trajectory onto the target subspace as

$$\mathbf{z}(\cdot) = \mathbf{Q}^T \mathbf{x}(\cdot). \quad (6.3)$$

In such a projection from the original n -dimensional space to its subspace $\mathbf{Q}^T \mathbf{x} : \mathbb{R}^n \mapsto \mathbb{R}^m$, the trajectory curve is contracted to reside in the reduced subspace (*cf.* lemma 6.1).

It is to be noted that, the application of the proposed techniques is not limited to a specific projection based model order reduction and the projection operators from any of the projection based methods such as: Krylov-subspace methods [82–85, 90, 91], TBR [88, 104, 112], and POD [87, 127, 210, 256] can be used.

The key idea in the proposed optimal order estimation algorithm is to topologically observe the behavior of near neighboring points, that are lying on the projected trajectory

in the reduced subspace and is described in the following sections.

6.2.2 Geometrical Framework for the Projection

In the proposed approach, we consider the pairwise closeness of the states on the trajectories as a measure to characterize the local geometrical structure of the trajectories. This mathematically requires endowing the multidimensional (original and target) spaces with a measure to compute the "distance" between any two points within a small multidimensional neighborhood around every state. Hence, we regard these spaces as metric spaces [9] with the metric

$$d_n(t_i, t_j) = \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| = \sqrt{\sum_{\rho=1}^n (x_\rho(t_i) - x_\rho(t_j))^2}, \quad (6.4)$$

where $\mathbf{x}_i = \mathbf{x}(t_i)$ and $\mathbf{x}_j = \mathbf{x}(t_j) (\in \mathbb{R}^n)$ are two states on the original trajectory. The distance function for the points in reduced space is also defined in a similar manner.

Theoretically, any open set $U_i \ni \mathbf{x}_i (\subset \mathbb{R}^n)$ can be considered as a "neighborhood" of $\mathbf{x}_i$. The specific U_i we use in the proposed approach is geometrically visualized as a n -dimensional open ball centered at $\mathbf{x}_i$ with a radius of ε_n which is referred to as ε_n -neighborhood of $\mathbf{x}_i$. Any point within this ball is considered a neighboring point to $\mathbf{x}_i$. For our case, where ε_n is small in a certain sense, it is referred to as "nearest neighborhood" of $\mathbf{x}_i$ and neighbors are defined as "nearest neighboring points". These concepts are pictorially explained in Fig. 6.2. Mapping the trajectory curve to a m -dimensional subspace ($\subseteq \mathbb{R}^m$), when m is too small, results in that, the projected curve passes a particular point more than once (self-intersections) due to the contraction in the geometrical structure. However, the n -dimensional trajectory curve in original space cannot have self-intersection or fold-over sections (existence and uniqueness theorem [6, 168]). Fig. 6.3 illustrates this fact, from a

Figure 6.2: Illustration of a multidimensional adjacency ball centered at $\mathbf{x}(t_i)$, accommodating its four nearest neighboring points.

geometrical perspective. It depicts a self-intersection point ($\hat{A}$ and $\hat{E}$) in the projected curve $\hat{\mathbf{T}}$, while the corresponding original states (A and E in Fig. 6.1) were not even neighbors, this occurs since the m -dimensional subspace is too small that the projected curve to be safely accommodated without over-contracting it. In such conditions, not all points that lie close to one another (e.g. $\hat{A}$, $\hat{F}$ and $\hat{E}$) are neighbors because of the original dynamics. There is a new neighbor point (e.g. $\hat{F}$) on the projected trajectory that is close to a candidate point ($\hat{A}$) solely because we are viewing the path $\hat{\mathbf{T}}$ in a dimension that is too small. In Fig. 6.4, the neighborhood geometry of the reference point $\mathbf{x}(t_i)$ in the state space is shown together with its projection $\mathbf{z}(t_i)$ and nearest neighbors in the m -dimensional target space. Two neighboring points $\mathbf{z}(t_i)$ and $\mathbf{z}(t_k)$ on the projected path are the images of $\mathbf{x}(t_i)$ and $\mathbf{x}(t_k)$, respectively; while they are not neighbors in the original space. Considering the aforementioned concepts, the following definitions are formalized.

Definition 6.2. The points $\mathbf{z}(t_i)$ and $\mathbf{z}(t_k)$ which are neighbors in the reduced space are

Figure 6.3: Illustration of false nearest neighbor (FNN), where $\hat{T}$ is the projection of T in Fig. 1.

defined as “false neighbors” if $\mathbf{x}(t_i)$ and $\mathbf{x}(t_k)$ are not neighbors in the original state space.

Definition 6.3. The neighboring points on the projected trajectory $\mathbf{z}(t_i)$ and $\mathbf{z}(t_j)$ are “true neighbors” when $\mathbf{x}(t_i)$ and $\mathbf{x}(t_k)$ are also neighbors in the original state space.

6.2.3 Neighborhood Preserving Property

In this section, it will be shown that in a projection to a subspace with a sufficient order, the projected trajectory curve inherits the same neighborhood structure of the original trajectory. This implies that, (a) the original nearest neighboring points remain neighbors in such a projection, (b) the near neighbors in that reduced subspace are true neighbors.

Figure 6.4: Illustration of the neighborhood structure of the state $\mathbf{x}_i$ and its projection $\mathbf{z}_i$ in the state space and reduced space, respectively.

Lemma 6.1. Contraction Property: *In projection using an orthogonal matrix $\mathbf{Q}_{n \times m}$ ($m \ll n$), the inner points of ϵ_n -neighborhood of any point on the original (n -dimensional) trajectory are preserved as inner points of ϵ_m -neighborhood in (m -dimensional) reduced subspace, where $\epsilon_m < \epsilon_n$.*

Proof. Let $\mathbf{x}(t_i)$ be any arbitrarily selected state on the original trajectory and $\mathbf{x}(t_j)$ be a neighboring state lying in the ϵ_n -neighborhood of $\mathbf{x}(t_i)$. Also, let $\mathbf{z}(t_i)$ and $\mathbf{z}(t_j)$ respectively, be the images of $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ in a reduced subspace of order m . The Euclidean distance between these points is

$$\begin{aligned} d_m(t_i, t_j) &= \|\mathbf{z}(t_i) - \mathbf{z}(t_j)\| = \|\mathbf{Q}^T \mathbf{x}(t_i) - \mathbf{Q}^T \mathbf{x}(t_j)\| \\ &= \|\mathbf{Q}^T (\mathbf{x}(t_i) - \mathbf{x}(t_j))\| \leq \|\mathbf{Q}^T\| \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|. \end{aligned} \quad (6.5)$$

Using the following properties of the matrix 2-norm [117]

$$\|\mathbf{Q}^T\| = \|\mathbf{Q}\|, \quad (6.6a)$$

$$\|\mathbf{Q}\| = \sqrt{\lambda_{\max}(\mathbf{Q}^T \mathbf{Q})} = \delta_{\max}(\mathbf{Q}), \quad (6.6b)$$

and for any orthogonal matrix $\mathbf{Q}$

$$\delta_{\max}(\mathbf{Q}) = 1, \quad (6.6c)$$

where $\lambda_{\max}$ and $\delta_{\max}$ denote the largest eigenvalue and the largest singular value, respectively. Considering the above properties (6.6a) - (6.6c) and using (6.5) we get

$$d_m(t_i, t_j) \leq \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|. \quad (6.7)$$

$\mathbf{x}(t_j)$ being in the open ε_n -neighborhood of $\mathbf{x}(t_i)$, we have $\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| < \varepsilon_n$. Hence, from (6.7),

$$\|\mathbf{z}(t_i) - \mathbf{z}(t_j)\| < \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| < \varepsilon_n. \quad (6.8)$$

Next, consider the surface of ε_m -neighborhood ball centered at any arbitrary point $\mathbf{z}(t_i) \in \hat{\mathbf{T}}$, defined as

$$\{\mathbf{z}(t) \mid \mathbf{z}(t) \in \mathbb{R}^m \text{ \& } d_m(t_i, t) = \varepsilon_m\}. \quad (6.9)$$

Using (6.8), it is straightforward to show that,

$$\varepsilon_m < \varepsilon_n. \quad (6.10)$$

This concludes the proof of the contraction property. ■

Lemma 6.2. *The accuracy for the macromodel is ensured if and only if all the nearest neighboring points on the projected trajectory are true neighbors.*

Proof. First, assuming a sufficient order for the reduced macromodel that can ensure an adequate accuracy, we prove that any two near neighboring points on the projected trajectory $\hat{\mathbf{T}}$ are true neighbors. Let $\hat{\mathbf{T}} = \{\widehat{\mathbf{z}}(t) : \text{for } t \in \Lambda_t\}$ denote the trajectory curve obtained from a reduced model. Consider $\mathbf{z}(t_j)$ to be a point in the ε_m -neighborhood of any arbitrary $\mathbf{z}(t_i)$ ($j \neq i$) on the projected trajectory $\hat{\mathbf{T}}$, we have

$$d_m(t_i, t_j) = \|\mathbf{z}(t_i) - \mathbf{z}(t_j)\| < \varepsilon_m, \quad (6.11)$$

where ε_m is a small neighborhood radius. Equation (6.11) can be equivalently rewritten as

$$\|\mathbf{z}(t_i) - \mathbf{z}(t_j)\| = \|\mathbf{Q}^T \mathbf{x}(t_i) - \mathbf{Q}^T \mathbf{x}(t_j)\| < \varepsilon_m. \quad (6.12)$$

From the direct solution of the reduced system $\widehat{\mathbf{z}}(t)$, the approximated responses $\mathbf{x}_a(t_i)$ and $\mathbf{x}_a(t_j)$ are obtained as

$$\begin{aligned} \mathbf{x}_a(t_i) &= \mathbf{Q} \widehat{\mathbf{z}}(t_i), \\ \mathbf{x}_a(t_j) &= \mathbf{Q} \widehat{\mathbf{z}}(t_j). \end{aligned} \quad (6.13)$$

Let the error vectors between the actual and approximated responses at time instants t_i and t_j be denoted as ζ_i and ζ_j ($\in \mathbb{R}^n$), respectively,

$$\begin{aligned} \mathbf{x}(t_i) - \mathbf{x}_a(t_i) &= \zeta_i, \\ \mathbf{x}(t_j) - \mathbf{x}_a(t_j) &= \zeta_j. \end{aligned} \quad (6.14)$$

Assuming that the order m_0 is sufficient to ensure accuracy of the reduced model,

$$\|\zeta_i\| = \|\mathbf{x}(t_i) - \mathbf{x}_a(t_i)\| < \bar{\xi}_{er},$$

$$\|\zeta_j\| = \|\mathbf{x}(t_j) - \mathbf{x}_a(t_j)\| < \bar{\xi}_{er}, \quad (6.15)$$

where $\bar{\xi}_{er}$ is a small positive value. By substituting $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ from (6.13) and (6.14) in (6.12), we get

$$\begin{aligned} \|\mathbf{Q}^T \mathbf{x}(t_i) - \mathbf{Q}^T \mathbf{x}(t_j)\| &= \|\mathbf{Q}^T (\mathbf{Q}\widehat{\mathbf{z}}(t_i) + \zeta_i) - \mathbf{Q}^T (\mathbf{Q}\widehat{\mathbf{z}}(t_j) + \zeta_j)\| = \\ &\|(\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)) - \mathbf{Q}^T (\zeta_j - \zeta_i)\| < \epsilon_m. \end{aligned} \quad (6.16)$$

The inverse triangle inequality [75] holds for any two vectors $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{R}^m$ holds as

$$\|\mathbf{V}_1\| - \|\mathbf{V}_2\| \leq \|\mathbf{V}_1 - \mathbf{V}_2\|. \quad (6.17)$$

Using this property, from (6.16) we get

$$\begin{aligned} \|\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)\| - \|\mathbf{Q}^T (\zeta_j - \zeta_i)\| &\leq \\ \|(\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)) - \mathbf{Q}^T (\zeta_j - \zeta_i)\| &< \epsilon_m. \end{aligned} \quad (6.18)$$

Applying (6.6) on (6.18), we get

$$\|\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)\| < \epsilon_m + \|\zeta_j - \zeta_i\| < \epsilon_m + \|\zeta_i\| + \|\zeta_j\| < \epsilon_m + 2\bar{\xi}_{er}. \quad (6.19)$$

This proves that $\widehat{\mathbf{z}}(t_i)$ and $\widehat{\mathbf{z}}(t_j)$ to be neighboring points on the solution trajectory obtained from the reduced macromodel. Multiplying both sides of (6.19) by $\|\mathbf{Q}\| = 1$ in (6.6), we get

$$\|\mathbf{Q}\| \|\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)\| < \epsilon_m + 2\bar{\xi}_{er}. \quad (6.20)$$

It is

$$\|\mathbf{Q}(\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j))\| \leq \|\mathbf{Q}\| \|\widehat{\mathbf{z}}(t_i) - \widehat{\mathbf{z}}(t_j)\| < \varepsilon_m + 2\bar{\xi}_{er}. \quad (6.21)$$

Combining (6.13) and (6.14) with (6.21) results in

$$\|\mathbf{x}(t_i) + \zeta_i - \mathbf{x}(t_j) - \zeta_j\| < \varepsilon_m + 2\bar{\xi}_{er} \quad (6.22)$$

and

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| < \varepsilon_m + 4\bar{\xi}_{er}. \quad (6.23)$$

Hence, $\mathbf{x}_j$ falls within a close neighborhood of $\mathbf{x}_i$. This verifies that $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ are neighboring points on the original trajectory and indicates $\mathbf{z}(t_i)$ and $\mathbf{z}(t_j)$ as true neighbors.

Second, we prove that, having all the nearest neighbors on the projected trajectory as true neighbors is a sufficient condition, to guarantee the accuracy of the reduced macro-model. For this, let $\mathbf{z}(t_j)$ be a neighboring point to an arbitrarily selected point on the projected trajectory $\mathbf{z}(t_i)$ ($j \neq i$). Being a true near neighbor, the corresponding state $\mathbf{x}(t_j)$ lies within a small neighborhood ball of $\mathbf{x}(t_i)$. This requires that the right hand side of the equation in (6.23) to be upper bounded to a small value, as

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| < \varepsilon_m + 4\bar{\xi}_{er} \leq \varepsilon_n. \quad (6.24)$$

From (6.24),

$$\bar{\xi}_{er} \leq \frac{\varepsilon_n - \varepsilon_m}{4}, \quad (6.25)$$

where both neighborhood radii $\varepsilon_n > \varepsilon_m > 0$ are small. According to (6.25), the small values of ε_n and ε_m guarantee the accuracy of the macromodel by upper-bounding the errors to a small value at all time instants throughout the projected trajectory. ■

Lemma 6.2 establishes that, when the reduced space is of a sufficient dimensionality

such that no false neighbors are present in the reduced space, a sufficient level of accuracy is ensured for the model. These facts form the underlying idea for the proposed method which is mainly based on successively reversing the trajectory folding process.

6.2.4 Unfolding the Projected Trajectory

Starting from a low-dimensional space, the order of the reduced space is consecutively increased. In each step, the projected trajectory is expanded into higher dimensions. Consequently, some neighboring points move far apart and reveal themselves as false nearest neighbors. This can be visualized as gradually unfolding the sections that have been folded over. The count of false nearest neighbors can be utilized to monitor this unfolding process. Ultimately, at some order (e.g. m_0), the count of false nearest neighbors drops to zero, such that, further increasing the order does not help the unfolding, and hence does not lead to revealing any new false nearest neighbors. Only then the points which are true neighbors on the original trajectory will stay neighbors on the projected trajectory in reduced space. This fact is illustrated in Fig. 6.5, where the changes of two nearest-neighbors (A and B) in a transition from order m to $m + 1$ is visualized. It shows a trajectory embedded in a subspace with insufficient order ($m < m_0$). By adding the $(m + 1)$ -th dimension, these two points move apart to their new locations $\hat{A}$ and $\hat{B}$, respectively. Their distance in the m dimension space; $AB = d_m(i, j)$ was changed to $\hat{A}\hat{B} = d_{m+1}(i, j)$ in the $m + 1$ dimensional subspace. In the unfolding process, such large change in separation between A and B indicates B as a false neighbor of A in m dimension. From Fig. 6.5 we have,

$$d_{m+1}^2(i, j) = d_m^2(i, j) + (z_{m+1}(t_i) - z_{m+1}(t_j))^2. \quad (6.26)$$

Figure 6.5: Displacement between two false nearest neighbors in the unfolding process.

In the proposed method, the component of the displacement vector between two neighbors on the new axis

$$\Delta_{z_{m+1}}(i, j) = |z_{m+1}(t_i) - z_{m+1}(t_j)|, \quad (6.27)$$

is used as a measure to monitor the behavior of the neighboring points in the unfolding process. If $\Delta_{z_{m+1}}(i, j)$ is not small compared to their Euclidean distance $d_m(i, j)$, they are false neighbors. This comparison is performed by checking the following ratio [253, 257]

$$R_{ij} = \left[\frac{d_{m+1}^2(i, j) - d_m^2(i, j)}{d_m^2(i, j)} \right]^{1/2} = \frac{|z_{m+1}(t_i) - z_{m+1}(t_j)|}{d_m(i, j)} = \frac{\Delta_{z_{m+1}}(i, j)}{d_m(i, j)}. \quad (6.28)$$

This central idea is summarized in the following corollary.

Corollary 6.1. *The order m_0 is an optimally minimum reduction order if increasing the order of the reduced subspace to m_1 , where $m_1 > m_0$ does not reveal any false nearest neighboring points on the reduced trajectory.*

6.3 Computational Steps of the Proposed Algorithm

The steps of the proposed algorithm are summarized as follows. *For the sake of simplicity in the notation, hereafter, we drop “t” in the equations (e.g. $\mathbf{z}(t_i)$ is referred to as $\mathbf{z}(i)$).*

(1) The proposed algorithm uses the time series data from the projected trajectory $\mathbf{z}(\cdot) \in \mathbb{R}^{m \times N}$,

$$\mathbf{z}(\cdot) = \{\mathbf{z}(\mathbf{i}) \in \mathbb{R}^{m \times 1} \mid \mathbf{z}(\mathbf{i}) = \mathbf{Q}^T \mathbf{x}(\mathbf{i}), \text{ for } \mathbf{i} = 1, \dots, N\}. \quad (6.29)$$

This requires one-time transient simulation of the circuit with any arbitrary inputs with a wide frequency spectrum up to the maximum frequency of interest to obtain the response $\mathbf{x}(\cdot)$. It is to be noted that, the typical goal of linear model reduction is to accurately represent a particular system output up to a certain maximum frequency as dependent on the specific application. Therefore, it is essential that the spectrum of the excitation signals adequately cover the frequency range of interest.

The projection matrix $\mathbf{Q}$ is formed with a small initial number (m) of orthonormal basis, and the time series data $\mathbf{z}(\cdot)$ is obtained using (6.3).

(2) A set of close points Π_i to each point on the projected trajectory $\mathbf{z}(i)$ is found using the following check:

$$d_{m_s}(i, j) < R, \text{ for } j = 1, \dots, N \text{ and } j \neq i, \quad (6.30)$$

where R is a search radius.

In general, any choice of R which leads to Π_i containing a few close points is adequate. An unnecessary large search radius, leads to a (unnecessary) large number of the neighbors. This does not improve the result, but may takes unnecessary CPU time and slow down the algorithm.

Efficient neighborhood searching algorithms have been extensively studied in computational geometry and image processing [258–260]. However, due to the relatively small size of the data set ($m \times N$) in this step of the proposed method, a simple and straightforward search algorithm as outlined in “Algorithm-5” is adequate.

(3) Next, the number of orthogonal basis is increased from m to $m + 1$ and the new dimension of the time-series data for the projected trajectory is computed

$$\mathbf{z}_{(m+1)}(\cdot) = \mathbf{q}_{(m+1)}^T \mathbf{x}(\cdot). \quad (6.31)$$

(4) All the close points $\mathbf{z}(j) \in \Pi_i$ to $\mathbf{z}(i)$ $i = 1, \dots, N$, that satisfies the following ratio (6.28) test are marked as false neighbors

$$R_{ij} = \frac{\Delta_{z_{m+1}}(i, j)}{d_m(i, j)} > \rho_t, \quad (6.32)$$

where ρ_t is a pre-specified threshold value.

This FNN search process is summarized in “Algorithm-6”.

(5) By repeating the steps (2)-(4), the projected trajectory is unfolded into higher dimensions. Ultimately, at some order m_0 , the count of false nearest neighbors in step (4) drops to zero, such that, further increasing the order does not lead to revealing any new false nearest neighbors. According to Corollary 6.1, m_0 is designated as the optimum minimal dimension for the reduced subspace.

The above computational steps are summarized in the pseudo codes depicted in "Algorithm-5", "Algorithm-6" and "Algorithm-7".

The following points are also worth mentioning:

1) The initial search for close neighbors in “Algorithm-5” is performed only once on the $N \times m$ time series from the projected trajectory, where m is a small (starting) order and

Algorithm 5: Neighborhood Search

Data: $\mathbf{z} \in \mathbb{R}^{m \times N}$ *(Data matrix for the projected trajectory)*
Result: Π *(Neighborhood information)*

```

1  $\mathbf{I} \leftarrow \{i \mid 1 \leq i \leq N\};$ 
2 foreach  $i \in \mathbf{I}$  do
3   foreach  $j \in \mathbf{I} - \{i\}$  do
4     Find  $d_m(i, j)$ ;
5     if  $d_m(i, j) < R$  (6.30) then
6        $\Pi_i \leftarrow (j, d_m(i, j))$ ;

```

Algorithm 6: False Nearest Neighbor (FNN)

Data: $\mathbf{z}_{m+1} \in \mathbb{R}^{m \times 1}$ *(new coordination)*
Result: False nearest neighbor count & Π *(updated)*

```

1 foreach  $i \in \mathbf{I}$  (If  $\Pi_i \neq 0$ ) do
2   foreach  $j \in \Pi_i$  do
3     Compute  $R_{ij}$  from (6.32);
4     if  $R_{ij} > \rho_t$  (6.32) then
5       False nearest neighbor count + 1;
6     Compute  $d_{m+1}(i, j)$  from (6.26);
7      $\Pi_i \leftarrow (j, d_{m+1}(i, j))$ ;

```

Algorithm 7: Proposed Order Estimation Algorithm

Data: $\mathbf{x} \in \mathbb{R}^{n \times N}$ *(Data matrix from trajectory)*
Result: Optimally minimum reduction order (m_0)

```

1  $m \leftarrow$  An arbitrary (small-starting) order;
2  $\mathbf{Q}_{n \times m} \leftarrow$  Projection matrix;
3 Find projected trajectory  $\mathbf{z}(\cdot)$  from (6.3);
4  $\Pi \leftarrow$  From Algorithm-5;
5 while False nearest neighbor count  $> 0$  do
6    $m \leftarrow m + 1$ ;
7   Find  $\mathbf{q}_{m+1}$ ;
8   Find  $\mathbf{z}_{m+1}$  from (6.31);
9   False nearest neighbor count,  $\Pi \leftarrow$  From Algorithm-6;
10 Optimally minimum reduction order  $(m_0) \leftarrow m$ 

```

N is the number of time points (generally, in the range of a few hundreds). Finding all neighbors in a set of N vectors of size m can be performed in $\mathcal{O}(N \log(m))$ or $\mathcal{O}(N)$ under mild assumptions [260, 261].

2) The computational complexity of the false nearest neighbors (FNN) search in “Algorithm-6” is $\mathcal{O}(Nn)$. In addition, the FNN algorithm is suitable for parallel implementation leading to additional reduction in the computational cost [262].

3) The proposed method does not require the formulation or simulation of the reduced macromodel.

4) Increasing the number of orthogonal basis from m to $m + 1$ keeps the initial m basis unchanged and thus requires the computation of only one new vector.

6.4 Numerical Examples

In this section, numerical results are presented to demonstrate the validity, performance and accuracy of the proposed methodology. It is to be noted that the proposed method is not limited to any specific projection method. For the purpose of illustration, the block Krylov subspace based projection method [64, 90] is used in the following numerical examples.

6.4.1 Example I

In this example, the transmission line shown in Fig. 6.6 is considered. The per-unit-length parameters are $C = 1.29 \text{ pF/cm}$, $L = 3.361 \text{ nH/cm}$ and $R = 5.421 \Omega/\text{cm}$. The transmission line is discretized using conventional lumped segmentation [249] to the form of 1500 lumped RLGC π -sections in cascade. The order of the subcircuit (excluding terminations) is 4500.

The input excitations at the two ports of the transmission line are set to be Gaussian

Figure 6.6: (a) A lossy transmission line as a 2-port network with the terminations; (b) Modeled by 1500 lumped RLGC π -sections in cascade.

voltage pulses with 60dB bandwidth at 5 GHz [263]. The terminal currents define the output vector.

Applying the proposed method, Fig. 6.7 shows the count of the false nearest neighbors on the projected trajectory, while the dimension of the model is changed from m to $m + 1$. Here, the vertical axis in Fig. 6.7 represents the percentage of the total count of FNN compared to the total count of neighbors. As seen from Fig. 6.7, $m \geq 65$ completely unfolds the projected trajectory with no false neighbors. Hence, according to Corollary-6.1, $m = 66$ is selected as the optimum order. The original subcircuit is reduced using the PRIMA algorithm with order 66. The equations of the reduced model are combined with the MNA equations of the rest of the circuit [264]. The input voltage source was a trapezoidal pulse with rise/fall times of 0.1ns, delay of 1ns and pulse width of 5ns. Comparison of the simulation results obtained from the original circuit of Fig. 6.6 and from the reduced circuit are shown in Fig. 6.8–6.9, which show excellent agreement.

To validate that $m = 66$ is the the optimum order, the voltage responses at the two ends of the subcircuit in Fig. 6.6 were recorded. The error in the response obtained from the

Figure 6.7: The percentage of the false nearest neighbors on the projected trajectory.

reduced circuit is defined as

$$e_k(\cdot) = S_k^{\text{Original}}(\cdot) - S_k^{\text{Reduced}}(\cdot),$$

$$RMS\ Error = \left(\frac{1}{PN_t} \sum_{k=1}^P \|e_k\|_2^2 \right)^{\frac{1}{2}}, \quad (6.33)$$

where S_k denotes the responses at P outputs of interest and N_t is the number of the time samples. Fig. 6.10 shows the error as function of the order of the reduced subcircuit.

Figure 6.8: Transient response of the current entering to the far-end of the line when the reduced model is of order $m = 66$.

6.4.2 Example II

In this example, we consider a RLC mesh shown in Fig. 6.11. The RLC subcircuit (in Fig. 6.11) is connected to the rest of the circuit through its 24 ports. The order of the subcircuit (excluding terminations) is 5800.

As explained for example 2, excitations directly at all the ports of the subnetwork are set to be Gaussian voltage pulses with 60dB bandwidth at the upper frequency limit of interest. The terminal currents define the output vector. The proposed method (Sec. 6.2) is applied to estimate the optimum order for the reduced macromodel for subcircuit. Fig. 6.12 shows the count of the false nearest neighbors on the projected trajectory, while the dimension of the model is changed from m to $m + 1$. Here, the vertical axis in Fig. 6.12 represents the percentage of the total count of FNN compared to the total count of neighbors. As

Figure 6.9: Transient response of the current at the far-end terminal of the line when the reduced model is of order $m = 66$.

seen from Fig. 6.12, $m \geq 290$ completely unfolds the projected trajectory with no false neighbors. Hence, according to Corollary-6.1, $m = 290$ is selected as the optimum order.

The original subcircuit is reduced using the PRIMA algorithm with order 290. The equations of the reduced model are combined with the MNA equations of the rest of the circuit. The three input voltage sources, connected to the near-ends (left side) of the horizontal traces 1, 6, and 12 are trapezoidal pulses with rise/fall times of 0.1ns, delay of 1ns and pulse width of 5ns. Comparison of the simulation results obtained from the original circuit of Fig. 6.11 and from the reduced circuit are shown in Fig. 6.13–6.14, which show excellent agreement.

To validate that $m = 290$ is the optimum order, Fig. 6.10 shows the error in the output voltages (6.33) as a function of the order of the reduced subcircuit.

Figure 6.10: Accuracy comparison in PRIMA models with different orders.

Figure 6.11: A RLC mesh as a 24-port subcircuit with the terminations.

Fig. 6.7 and Fig. 6.12 depict that the descending pace of the percentage of false nearest neighbors is not monotonic. Hence, dropping FNN(%) to zero for just the first time is not sufficient to decide the order; but it should also remain zero for several subsequent orders.

Figure 6.12: The percentage of the false nearest neighbors among 1000 data points on the projected trajectory.

Figure 6.13: Transient responses at near-end of horizontal trace#1.

Figure 6.14: Transient responses at near-end of horizontal trace#10.

Figure 6.15: Errors from using the reduced models with different orders in the frequency domain.

Chapter 7

Optimum Order Determination for Reduced Nonlinear Macromodels

Chapter 4 presented some of the well-known methods for the reduction of nonlinear systems such as Taylor series based methods, Trajectory PieceWise Linear (TPWL)-based methods, Proper Orthogonal Decomposition (POD), and Empirical Balanced Truncation. An important common problem in these nonlinear order-reduction techniques is the selection of order for the reduced model. In this chapter, the detail of a novel algorithm for optimal-order determination for the reduced nonlinear models is presented.

7.1 Introduction

In the prominently used nonlinear order-reduction techniques, “selection of order” is an important practical issue. The selection of an optimum order is important to achieve a pre-defined accuracy while not over-estimating the order, which otherwise can lead to inefficient transient simulations and hence, undermine the advantage from applying MOR. The reduced-order estimation issue for linear circuits has been recently addressed in [251,265] and explained in the previous chapter.

This chapter presents a novel algorithm to determine an optimally minimum order for a nonlinear circuit reduction, based on the geometric theory of nonlinear dynamical circuits (see e.g. [22, 27–30, 166, 266, 267]). The proposed methodology is founded on the idea of monitoring the local geometrical structure of the projected nonlinear trajectory in the reduced space [267]. To serve this purpose, a mathematical algorithm is devised to observe the behavior of near neighboring points, lying on the low-dimensional nonlinear trajectory, when increasing the dimension of a reduced-space. The order is determined such that the projected trajectory is unfolded properly in the reduced space, while monitoring the count of the "False Nearest Neighbor (FNN)" points on the projected trajectory. The reduced model in this optimally reduced subspace captures the major dynamical properties of the original system.

7.2 Background

The general class of systems occurring in a broad range of engineering problems is known as dynamical systems, whose behavior changes in time according to some deterministic rules. These rules specify, how a system's states evolve by time starting from an initial condition. A set of coupled Differential-Algebraic Equations (DAE) constitutes a mathematical model to characterize their dynamical behavior. Then, mapping from the space of input signals to the space of output signals is completed by an algebraic equation called output equation.

7.2.1 Formulation of Nonlinear Circuit Equations

Nonlinear electrical circuits can also be characterized in time-domain by a set of coupled nonlinear first-order DAE [21–23, 25, 27, 164–166]. In the context of circuit simulation,

these equations are directly obtained from the circuit netlist using the Modified Nodal Analysis (MNA) matrix formulation [37–39, 167] as follows:

$$\mathbf{C} \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) + \mathbf{f}(\mathbf{x}(t)) = \mathbf{B} \mathbf{u}(t) \quad (7.1)$$

$$\mathbf{y}(t) = \mathbf{L} \mathbf{x}(t) \quad (7.2)$$

where $\mathbf{C}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are susceptance and conductance matrices including the contribution of linear elements, respectively, $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of MNA variables (the nodal voltages, some branch currents and electrical charges) of the circuit. $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n$ is a vector of real-valued functions including the stamps of all nonlinear elements in the circuit. $\mathbf{B}$ and $\mathbf{L}$ are the input and output matrices, respectively.

7.2.2 Model Order Reduction of Nonlinear Systems

The basic idea of model order reduction of a circuit is to replace the original system by an approximated system with a reduced DAE realization of order m , which is significantly smaller than the original order n . Model reduction algorithms seek a proper order m for which the outputs from the reduced system and the original responses are approximately equal for inputs of interest $\mathbf{u}(t)$.

7.2.3 Projection Framework

In any projection based reduction process, an original n -dimensional state space is projected to a m -th order subspace ($m \ll n$). This requires the creation of some projection operators $\mathbf{W}$ and $\mathbf{Q} \in \mathbb{R}^{n \times m}$, where $\mathbf{W}^T \mathbf{Q} = \mathbf{I}_{m \times m}$. Assume that, there exists $\mathbf{z}(t) \in \mathbb{R}^m$ in a reduced subspace such that it satisfies $\mathbf{x}(t) = \mathbf{Q} \mathbf{z}(t)$. Due to the orthogonality of the projection matrices ($\mathbf{W}$ and $\mathbf{Q}$), $\mathbf{z}(t) = \mathbf{W}^T \mathbf{x}(t)$. The differential equations for the

reduced system are obtained through a variable change from (7.1) and (7.2) as

$$\mathbf{W}^T \frac{d}{dt} \mathbf{C}(\mathbf{Q}\mathbf{z}(t)) = \mathbf{W}^T \mathbf{f}(\mathbf{Q}\mathbf{z}(t)) + (\mathbf{W}^T \mathbf{B}) u(t) \quad (7.3)$$

$$\mathbf{y}(t) = (\mathbf{L}\mathbf{Q}) \mathbf{z}(t) \quad (7.4)$$

The approximate response $\tilde{\mathbf{x}}(t)$ is obtained by solving the reduced-order dynamical model in (7.3) as $\tilde{\mathbf{x}}(t) = \mathbf{Q}\mathbf{z}(t)$. The error between the original state variables and its approximation is $\boldsymbol{\zeta} = \mathbf{x} - \mathbf{Q}\mathbf{W}^T \mathbf{x}$. For a Galerkin projection scheme, i.e. $\mathbf{W} = \mathbf{Q}$.

In the rest of this chapter, we will use the classical POD [210, 256] to describe the proposed algorithm for order estimation. *However, it should be emphasized that, the proposed algorithm is not limited to a specific nonlinear model order reduction method and can be used in conjunction with any of the above mentioned methods.*

Using the POD method, for a given representative input $\mathbf{u}(t)$, the “time-snapshots” of the transient response are collected in a data matrix as $\mathbf{X} = [\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_N)] \in \mathbb{R}^{n \times N}$. The POD method seeks to find a projection basis $\mathbf{Q}$ to accurately approximate the original response with an approximate representation of data points by minimizing the overall projection error $\|\boldsymbol{\zeta}\|_2 = \|\mathbf{X} - \mathbf{Q}\mathbf{Q}^T \mathbf{X}\|_2$. The solution to this optimization problem is obtained by performing singular value decomposition (SVD) on the data matrix as $\mathbf{X} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^T$ [256]. The POD basis for a Galerkin projection is given by the first m columns in $\mathbf{V}$ ($\in \mathbb{R}^{n \times n}$) as $\mathbf{Q} = [\mathbf{v}_1, \dots, \mathbf{v}_m] \in \mathbb{R}^{n \times m}$.

POD constructs the matrix $\mathbf{Q}$ as shown in Algorithm 8.

POD is known as a promising method to provide efficient and accurate transient simulation (e.g. more accurate than the TPWL [256]). This is for the output responses corresponding to a family of excitation signals close to the one used to form the POD basis.

Algorithm 8: POD procedure for constructing $\mathbf{Q}$ of dimension m

-
- input** : Original trajectory $\mathbf{x}(\cdot)$, reduction order m
output: Projection Matrix $\mathbf{Q}$
- 1 Access the transient response data matrix, $\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_N)] \in \mathbb{R}^{n \times N}$;
 - 2 Perform the SVD of $\mathbf{X}$: $\mathbf{X} = \mathbf{V}\Sigma\mathbf{U}^T$ with
 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N] \in \mathbb{R}^{n \times n}$, $\Sigma \in \mathbb{R}^{n \times N}$ and $\mathbf{U} \in \mathbb{R}^{N \times N}$;
 - 3 Truncate the first m left singular vector to obtain the order reduction projection matrix $\mathbf{Q} = [\mathbf{v}_1, \dots, \mathbf{v}_m] (\in \mathbb{R}^{n \times m})$;
-

7.3 Order Estimation for Nonlinear Circuit Reduction

7.3.1 Differential Geometric Concept of Nonlinear Circuits

In this subsection, we define the concepts of the geometric theory of nonlinear dynamical circuits [22, 27, 29, 30, 166] that are related to this work, in a rather intuitive manner.

Let $\mathbf{x}(t) = \{\mathbf{x}_i(t) : \text{for } i = 1, 2, \dots, n\}$ denote the variables set in the state equation (7.1). The variables set $\mathbf{x}(t)$ is used as a coordinates system in an n -dimensional space, called state space. It is a geometric model for the set of all possible states in the dynamic (transient) behavior of a nonlinear system for any possible inputs. Consider the solution of (7.1) for a given input $\mathbf{u}(t)$ and an initial condition $\mathbf{x}_0$. This solution $\mathbf{x}(t)$ would be a set of time-dependent functions $x_i(t)$ ($1 \leq i \leq n$), which are practically obtained through transient simulation as a set of time-sequenced (time-series) data throughout a certain time span $t_i \in \mathcal{D} = [t_0, t_{max}]$. The response at each time instant t_i , represented by $\mathbf{x}(t_i) = [x_1(t_i), x_2(t_i), \dots, x_n(t_i)]^T$ defines a point $\mathbf{x}_i$ in the multidimensional state space ($\mathbf{x}_i \in \mathcal{S}$). The locus of states $\mathbf{x}_i$ in this space for all $t_i \in \mathcal{D} \subset \mathbb{R}_+$ is a time-parametrized directional path Φ_t that starts at the point $\mathbf{x}(t_0) = \mathbf{x}_0$, henceforth this curve is referred to as the "trajectory" of the system.

These concepts are illustrated in Figures 7.2 and 7.3 for the Chua's circuit, shown in Fig. 7.1. The electronic circuit realization and the typical values of the parameters G_a, G_b ,

Figure 7.1: Chua's circuit.

and B_P are given in [268]. The dynamics of the circuit and the characteristic of its nonlinear resistor (a.k.a. chua's diode) are described by (7.5) [268, 269]. Despite the deceptively simple appearance of chua's circuit, it demonstrates a surprisingly complex dynamic behavior, due to the high non-linearity. This makes it a popular representative example used to demonstrate and study complex nonlinear trajectories.

Figure 7.2: Trajectory of the Chua's circuit in the state-space (scaled time: $0 \leq t \leq 100$) for a given initial condition.

It is to be noted that, response (solution) of a dynamic circuit is a real-valued

Figure 7.3: The time-series plot of the system variables ($x_i(t)$) as coordinates of state space.

continuously-differentiable function [22] from an open interval $\mathcal{D}(\subset \mathbb{R}_+)$ into the state space. Under practical assumptions [270], (7.1) is guaranteed to have a unique analytical solution over any finite time interval, that passes through the initial state at $t = t_0$ [271]. This establishes a certain properties for the geometrical structure of state trajectories, such as, trajectories (a) do not intersect each other, (b) do not have self-crossing points, and (c) do not have over-folding sections [6, 7, 17, 168]. Accordingly, state space of a dynamical nonlinear system is considered as a subspace $\mathcal{S} (\subseteq \mathbb{R}^n)$ with a sufficient dimension to embed all the possible states trajectories of a dynamic system while ensuring the above properties in (a)-(c). To further elaborate, we consider an inverter circuit as shown in Fig. 7.4-(a), where the nonlinear dynamics of the inverter gate is described by its behavioral model shown in Fig. 7.4-(b) as proposed in [272]. For a set of logic pulses (with different timings) at the input of inverter circuit, the associated responses are plotted in Fig. 7.5. It depicts a geometrical structure constituted by the family of response trajectories in its 3-dimensional state space $\mathcal{S}$, termed manifold $\mathcal{M}$. The manifold of a nonlinear dynamic circuit attracting

Figure 7.4: (a) Digital inverter circuit; (b) The circuit model to characterize the dynamic behavior of digital inverter at its ports.

its major response trajectories is a bounded region of the state-space ($\mathcal{M} \subset \mathcal{S}$), where these trajectories exist. The n-dimensional manifold that we consider in this work is the observable state space of the system which is representative for rich dynamical behavior of a nonlinear system. To contrast the behavior of linear and nonlinear state-space models, it should be noted that, for linear systems trajectories often (provably) stay close to a linear subspace (vector space). Whereas, trajectories of a nonlinear system tend to stay on a nonlinear manifold (curved surface) which is interpreted as a differentiable (smooth) geometrical structure [22, 27, 273, 274]. If we then join the other ends, we get a Klein bot-

Figure 7.5: A geometric structure $\mathcal{M}$ attracting the trajectories of the circuit in Fig.7.4.

tle, which requires four dimensions to describe it. Fig. 7.6 pictorially exemplify the notion of differential manifolds by showing, (a) Möbus and (b) Torus as topological subspaces locally consisting of 2-D patches of Euclidean space, while they are globally 3-D objects with curved surfaces. For example, we can fold and attach the ends of the paper and get a Mobius strip, which requires three dimensions to describe it, but locally, the geometry is still two-dimensional (in the paper). Similarly, nonlinear manifolds which capture the major system responses are globally curved surfaces and can not be realized as a subspace (vector space) always sitting inside a fixed Euclidean space, but it looks locally like Euclidean (vector) space. The above idea is roughly outlined in the following definition.

Definition 7.1. An n -dimensional manifold $\mathcal{M}$ is a topological space so that $\mathcal{M}$ is locally Euclidean of dimension n , i.e. for every $\mathbf{x} \in \mathcal{M}$, there exist an open neighborhood of $\mathbf{x}$ that is the same as the open n -dimensional sphere in $\mathbb{R}^n$.

In order to compute the distances on the manifold, one needs to equip a distance metric to the manifold. Considering the locally Euclidean structure of manifold, the distance between any two states on the trajectory $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ when $\mathbf{x}(t_j)$ falls within a small multidimensional neighborhood around $\mathbf{x}(t_i)$ is trustfully measured using the Euclidean

Figure 7.6: (a) The Möbus strip and (b) Torus are visualizations of 2D manifolds in $\mathbb{R}^3$

norm as

$$d_n(i, j) \triangleq \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\| = \sqrt{\sum_{r=1}^n (x_r(t_i) - x_r(t_j))^2} \quad (7.6)$$

For the sake of simplicity in the notation, henceforth, we drop “(t)” in the equations (e.g. $\mathbf{x}(t_i)$ is referred to as $\mathbf{x}_i$).

It is also to be noted that, confining the neighboring search to the close adjacency of every state allows the intrinsic metric properties of vector space to be locally used. However, the global distance between two states on the manifold $d_S(\mathbf{x}_i, \mathbf{x}_j)$ is generally defined as the length of the shortest trajectory curve connecting them $\widetilde{\mathbf{x}_i \mathbf{x}_j}$ (Geodesic distance). Accordingly, the results from using (7.6) to measure the distance for far points can be deceptively inaccurate. In order to study the global properties of a curve, such as Geodesic distance, the number of times that a curve wraps around a point or convexity properties, the topological tools are needed. But the further explanation falls beyond the scope of this work.

The aforementioned facts imply, the trajectory curves possess certain geometric properties and structure. Properties of curves can be generically classified into "local properties" and "global properties". Local properties are the properties that hold in a small neighborhood of a point on a curve.

In this chapter, the parametrized curves (trajectories) is considered as a geometrical model to study the dynamic behavior of the nonlinear circuits and to develop the proposed order estimation algorithm for nonlinear circuit reduction. We study the local properties of trajectories namely the neighborhood structure that is the set of the neighbors for each points.

7.3.2 Nearest Neighbors

In the proposed approach, we consider the pairwise closeness of the states on the trajectories (in Euclidean sense) as a measure to characterize the local geometrical structure of the trajectories. For this purpose, we define the σ_n -neighborhood of $\mathbf{x}(i)$ as $\mathbf{U}(\mathbf{x}_i, \sigma_n) = \{\mathbf{x}(t) \in \mathbb{R}^n \mid d_n(\mathbf{x}_i, \mathbf{x}(t)) < \sigma_n\}$. It is geometrically visualized as an n -dimensional open ball centered at $\mathbf{x}_i$ with a radius of σ_n . To study the local geometry, σ_n needs to be small in a certain sense, hence, $\mathbf{U}(\mathbf{x}(i), \sigma_n)$ is referred to as “(local) nearest neighborhood” of $\mathbf{x}(i)$ and neighbors are defined as “nearest neighboring points”. These concepts are illustrated in Fig. 7.7.

Figure 7.7: Illustration of a multidimensional adjacency ball centered at $\mathbf{x}(t_i)$ (x), accommodating its two nearest neighboring points (v) on the trajectory of the Chua's circuit (for $0 \leq t \leq 2$).

7.3.3 Geometrical Framework for the Projection

Using a projection operator $\mathbf{Q}_{n \times m}$, an image of the trajectory is obtained through a point-wise projection of the original trajectory onto a low-dimensional subspace as $\mathbf{z}(\cdot) = \mathbf{Q}^T \mathbf{x}(\cdot)$. The coordinate system defining the reduced subspace are the functions $z_i(t)$ for $i = 1, \dots, m$ that are linear combinations of the original state functions; i.e. $z_i(t) = \sum_{j=1}^n q_{ji} x_j(t)$, for $i = 1, 2 \dots m$, where $m \ll n$.

- **Folding:** Under projection-based dimension reduction, the original nonlinear trajectories manifold is practically contracted to reside in reduced subspace [265]. In cases that, the reduced space lacks an adequate dimensionality, the nonlinear manifold is overly-condensed and hence its geometric structure crumples up. One may visualize this as "folding" an n -dimensional object to fit in a low-dimensional space. This fact is demonstrated in Fig. 7.8, where the POD basis of the Chua's circuit were used to project its trajectory in Fig. 7.3 to a planar subspace. Similar to the underlying nonlinear manifold, when m is too small, an over-contraction of the projected trajectories in the target subspace may happen. This means that, the adjacency relationship among the states on the original trajectories will not be preserved in the reduced subspace, i.e., that the projected curve passes a particular point more than once (self-intersections). This shows that, the projected subspace is lacking required dimensionality to ensure the uniqueness of points on the projected curve in the subspace, which is crucial to represent a dynamic system. Fig. 7.9 illustrates this fact, from a geometrical perspective. It depicts a self-intersection point ($\mathbf{X}$, $\bigcirc$) in the projected curve, while the corresponding original states ($\mathbf{X}$, $\bigcirc$) were not neighbors. Also, the points (e.g. $\bigcirc$) on the projected trajectory are close to a candidate point ($\mathbf{X}$), not because of the dynamics, but solely because the projected path is viewed in a dimension that is too small.

Figure 7.8: Illustration of Chua's trajectory in Fig.7.7 projected to a two-dimensional subspace, where its underlying manifold is over-contracted.

Figure 7.9: (left) Illustration of false nearest neighbor (FNN), where the 3-dimensional trajectory of the Chua's circuit in Fig.7.7 is projected; (right) A zoomed-in view of the projected trajectory.

7.3.4 Proposed Order Estimation for Nonlinear Reduced Models

To explain the proposed method, first, the concept of the "false nearest neighbor" (FNN) on a nonlinear manifold needs to be formally defined as the following.

Definition 7.2. The points which are neighbors in the reduced space are defined as "false neighbors" when the corresponding states are not neighbors in the original manifold, and are "true neighbors" when the corresponding original states are also neighbors in the original manifold.

- **Unfolding:** Based on a geometric intuition, unfolding can be inferred as reversing the folding process. increasing the dimension of the reduced subspace, e.g. from $m = 2$ in Fig. 7.8 to $m + 1 = 3$ in Fig. 7.7 can unfold the geometric structure of trajectory by distancing the false nearest neighbors such as (X) and (O).

Inspired by these observations, the underlying idea in the proposed algorithm is to geometrically observe the behavior of near neighboring points that are lying on the projected nonlinear trajectory in an unfolding process. Starting from a low-dimensional subspace, the order of the reduced space is consecutively increased. In each step, the projected trajectory is expanded into higher dimensions. Consequently, some neighboring points move far apart and reveal themselves as false nearest neighbors. This is illustrated in Fig. 7.10. It depicts how neighborhood relations may change by going from m to $m + 1$ (note that in Fig. 7.10 $m = 1$). The neighboring point (O) that is closely located to the reference point (□) in $\mathbb{R}^m$ is noticeably displaced by the transition to $\mathbb{R}^{m+1}$ and hence is revealed as false neighbor.

This process of expansion, ultimately leads to a minimal order m_o for which and also for other higher orders $m > m_o$, only neighbors on the projected trajectory in the reduced space are true neighbors. In this way, the projected trajectory in an m_o -dimensional reduced space is a one-to-one image of the system trajectory in the original manifold. Thus, the

Figure 7.10: Drastic displacement between two false nearest neighbors in the unfolding process.

neighbors of a given point are mapped onto neighbors in the reduced space. When the reduced space has an adequate dimensionality the local geometric structure of the response trajectory will remain invariant to the orthogonal projection in a neighborhood of each state. However, if an m -dimensional space ($m < m_o$) is considered, then the topological structures are not preserved and the points are projected onto neighborhoods of other points to which they would not belong in higher dimensions (false neighbors). After constructing a subspace of sufficient order in which an unfolded projected trajectory can be embedded, further increasing the order does not lead to revealing any new false nearest neighbors. This is illustrated in Fig. 7.11. It depicts that, after complete unfolding of the geometric structure of trajectory in an m_o -dimensional subspace, by going from m_0 to $m + 1$ (in Fig. 7.11, $m = 1$) any point ($\square$) and its near neighbors ($\circ$, $\blacktriangledown$) in $\mathbb{R}^m$ are only slightly displaced by the transition to $\mathbb{R}^{m+1}$. As a quantitative measure of these effect, we consider the ratio of Euclidean distances between a point $\mathbf{x}_i$ and its nearest neighbor $\mathbf{x}_j$, first on an

Figure 7.11: Small displacement between every two nearest neighbors by adding a new dimension ($m + 1$ or higher), when trajectory was fully unfolded in m dimensional space.

m -dimensional and then on an $(m + 1)$ -dimensional space, it is [265,275].

$$R_{ij} = \left[\frac{d_{m+1}^2(i, j) - d_m^2(i, j)}{d_m^2(i, j)} \right]^{\frac{1}{2}} = \frac{\Delta z_{m+1}(i, j)}{d_m(i, j)} \quad (7.7)$$

Using (7.7), the relative change in distance by adding one more dimension is evaluated as a mean to decide if the states were not truly close together due to the dynamics but as a result of projection from a higher state space to smaller space with an inadequate dimension. To deem $\mathbf{x}_j$ to be a false nearest neighbor of $\mathbf{x}_i$ in an m -dimensional subspace the following should hold.

$$R_{ij} = \frac{\Delta z_{m+1}(i, j)}{d_m(i, j)} > \rho_{FNN} \quad (7.8)$$

where $\Delta z_{m+1}(i, j) \triangleq |z_{m+1}(i) - z_{m+1}(j)|$ and ρ_{FNN} is a threshold value.

- Upper bound for the choice of the threshold ρ_{FNN} :

The one parameter that needs to be determined before performing the false nearest neighbors algorithm is the threshold constant ρ_{FNN} in (7.8). For the FNN algorithm

to correctly find that there are no false nearest neighbors in the reduced subspace with adequate order the threshold value should be chosen in a proper range. This subsection investigates the bounds for this range of selections.

Proposition 7.1. *The choice of proper threshold value in the ratio test for the FNN algorithm of the proposed method is bounded to $0 \leq \rho_{FNN} \leq 1$.*

Proof. In order to determine an upper bound, let ρ_{max} be a large enough selection such that all the near neighboring points on the projected trajectory of order m can hold the following ration test:

$$R_{ij} = \frac{\Delta z_{m+1}(i, j)}{d_m(i, j)} \leq \rho_{max}, \quad \forall \mathbf{z}_i \text{ and nearest neighbor } \mathbf{z}_j. \quad (7.9)$$

From (7.9), we get

$$\begin{aligned} \frac{\Delta z_{m+1}(i, j)}{d_m(i, j)} &= \frac{|z_{m+1}(i) - z_{m+1}(j)|}{\|\mathbf{z}_m(i) - \mathbf{z}_m(j)\|} = \\ &= \frac{|\mathbf{q}_{m+1}^T \mathbf{x}_i - \mathbf{q}_{m+1}^T \mathbf{x}_j|}{\|\mathbf{Q}_m^T \mathbf{x}_i - \mathbf{Q}_m^T \mathbf{x}_j\|} = \frac{|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)|}{\|\mathbf{Q}_m^T (\mathbf{x}_i - \mathbf{x}_j)\|} \leq \rho_{max} \end{aligned} \quad (7.10)$$

and hence

$$|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)| \leq \rho_{max} \|\mathbf{Q}_m^T (\mathbf{x}_i - \mathbf{x}_j)\|. \quad (7.11)$$

Considering the consistent matrix norm in (7.11), it is

$$|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)| \leq \rho_{max} \|\mathbf{Q}_m^T (\mathbf{x}_i - \mathbf{x}_j)\| \leq \rho_{max} \|\mathbf{Q}_m^T\| \|\mathbf{x}_i - \mathbf{x}_j\|. \quad (7.12)$$

Due to the property of the 2-norm of orthonormal matrices, for the projection matrix $\mathbf{Q}$

$$\|\mathbf{Q}^T\| = \|\mathbf{Q}\| = 1. \quad (7.13)$$

From (7.12) and (7.13);

$$|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)| \leq \rho_{max} \|\mathbf{x}_i - \mathbf{x}_j\|. \quad (7.14)$$

Applying Cauchy-Schwarz inequality [276] to the left hand side of (7.12), we get

$$|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)| = |\langle \mathbf{q}_{m+1}, (\mathbf{x}_i - \mathbf{x}_j) \rangle| < \|\mathbf{q}_{m+1}\| \|\mathbf{x}_i - \mathbf{x}_j\| \quad (7.15)$$

and considering that the orthogonal projection basis have unity norm $\|\mathbf{q}_{m+1}\| = 1$, from (7.15)

$$|\mathbf{q}_{m+1}^T (\mathbf{x}_i - \mathbf{x}_j)| < \|\mathbf{x}_i - \mathbf{x}_j\| \quad (7.16)$$

$\rho_{max} = 1$ can be trivially decided to ensure both (7.16) and (7.15) are hold for any selection of the (self excluded) neighboring points. Hence it is $0 < \rho_{FNN} < \rho_{max} = 1$ that concludes the proof. ■

We established the steps of the proposed nonlinear order estimation in this work based on the fact that, when the reduced space is of a sufficient dimensionality, such that no false neighbors are present in the reduced space, a sufficient level of accuracy is ensured for the nonlinear reduced model. The rigorous justification for this can be formally based on the followings.

Lemma 7.1. *In an orthogonal projection of a nonlinear trajectory the near-neighborhood of any state on the original manifold are projected to a near neighborhood with smaller neighborhood radius in the reduced manifold.*

This may be referred to as contraction property of projection.

Lemma 7.2. *Having all the nearest neighboring points on the projected trajectory as true neighbors is a necessary and sufficient to ensure the accuracy of reduced nonlinear macro-model.*

The proofs of these lemmas are possible in a similar fashion as the proofs for lemmas 1-2 in [265] by the authors and the references therein.

Hence, based on the above lemmas the following corollary is concluded.

Corollary 7.1. *For nonlinear systems, the order m_o is an optimally minimum reduction order if increasing the order of the reduced subspace to m_l , where $m_l > m_o$, does not reveal any false nearest neighboring (FNN) points on the nonlinear reduced trajectory.*

The objective of the proposed method is to determine this optimum dimension m_o for the reduced subspace while preserving desired accuracy. The general steps of the proposed order determination algorithm for nonlinear systems using the FNN are explained in the Algorithm-1.

7.4 Computational Steps of the Proposed Algorithm

The steps of the proposed algorithm are summarized as follows. *For the sake of simplicity in the notation, hereafter, we drop “t” in the equations (e.g. $\mathbf{z}(t_i)$ is referred to as $\mathbf{z}(i)$).*

Algorithm 2: Proposed Order Estimation Algorithm

Input: $\mathbf{X} \in \mathbb{R}^{n \times N}$ (data matrix from original trajectory)

output: Optimal minimum reduction order (m_o)

- 1 Using the POD algorithm, the projection matrix $\mathbf{Q}$ is formed with a small initial number (m) of orthonormal basis;
- 2 The time-series data from the projected trajectory is stored in the form $\mathbf{Z} = [\mathbf{z}(t_0), \dots, \mathbf{z}(t_N)] \in \mathbb{R}^{m \times N}$;
- 3 A set of close points Π_i to each point on the projected trajectory is found based on the following criteria $d_m(i, j) < R$, for $j = 1, \dots, N$ and $j \neq i$, where R is a search radius. In general, any choice of R which leads to Π_i containing a few close points is adequate;
- 4 The number of orthogonal basis is increased from m to $m + 1$ and the new dimension of the subspace is computed as $\mathbf{z}_{(m+1)}(\cdot) = \mathbf{q}_{(m+1)}^T \mathbf{x}(\cdot)$;
- 5 All the close points $\mathbf{z}(j) \in \Pi_i$ to $\mathbf{z}(i)$ $i = 1, \dots, N$, that satisfies the following ratio test are marked as false neighbors $R_{ij} = \frac{\Delta z_{m+1}(i, j)}{d_m(i, j)} > \rho_{FNN}$, where ρ_{FNN} is a pre-specified threshold value;
- 6 By repeating the steps (4)-(5), the projected trajectory is expanded into higher dimensions. Ultimately, at a particular order m_o , the count of false nearest neighbors in step (5) drops to zero, such that, further increasing the order does not lead to revealing any new false nearest neighbors and m_o is selected as the minimum acceptable order for the reduced model.

The computational steps are summarized in the the flowchart shown in Fig. 7.12. The flowchart also depicts the interaction between the proposed method and the classical process of nonlinear Model reduction, to ensure the parsimony of the model generation cost, as well as the optimum size for the model.

Figure 7.12: Flowchart of the proposed nonlinear order estimation strategy. The gray blocks are the steps of nonlinear MOR interacting with the proposed methods.

From an implementation perspective, the following elaborations are important.

(a) The number of the false nearest neighboring states revealed in the step-5 of Algorithm-1 are traced as a function of the reduced dimension FNN(m). For this purpose, a measure in a percentage scale is defined as the ratio between the total count of FNN to the total count of initial near neighbors, i.e.

$$\text{FNN(m)}(\%) = \frac{\text{numel}(\mathbf{F}_m)}{\text{numel}(\mathbf{\Pi})} \times 100 \quad (7.17)$$

where $\text{numel}(\cdot)$ returns the number of elements in an array.

(b) Initial Nearest Neighbors (INN) search: Given the transient response data matrix

$$\mathbf{X} = [\mathbf{x}_i \mid i \in \mathcal{T} = \{1, \dots, N\}] \in \mathbb{R}^{n \times N},$$

let

$$\mathbf{Z}_{ini} = [\mathbf{z}_i \mid i \in \mathcal{T}] \in \mathbb{R}^{m_s \times N}$$

be the *initially* projected trajectory, where m_s is the starting order for the unfolding process. For any point on the initially projected trajectory $\mathbf{z}_i \in \mathbf{Z}_{ini}$, an array containing its “*initial* near neighboring” points is given as

```

1: for  $i \leftarrow 1$  to  $N$  do
2:   for  $j \leftarrow i$  to  $N$  do
3:      $\Pi_i = \{j \mid j \in \mathcal{T} - \{i\} \text{ and } \|\mathbf{z}_i - \mathbf{z}_j\| < R\}$ 
4:   end for
5: end for

```

where $\mathbf{z}_j \in \mathbf{Z}_{ini}$.

The result for the "Initial Nearest Neighbors (NN) search" in the step-3 of Algorithm-2 is

$$\mathbf{\Pi} = \bigcup_{i \in \mathcal{T}} \Pi_i.$$

(c) False Nearest Neighbors (FNN) search: For any point on the projected trajectory of order m ($> m_s$), an array containing its “false near neighbors” is given as

```

1: for  $i \leftarrow 1$  to  $N$  do
2:   if  $\Pi_i \neq \emptyset$  then
3:     for  $k \leftarrow 1$  to  $\text{numel}(\Pi_i)$  do
4:        $\mathbf{F}_{m,i} = \left\{ j \mid j = \Pi_i(k) \text{ and } \mathbf{z}_i \xleftrightarrow{\text{FNN}} \mathbf{z}_j \in \mathbf{Z}_m, \right\}$ 
5:     end for
6:   end if
7: end for

```

Hence, the result from step-5 in Algorithm-2 is $\mathbf{F}_m = \bigcup_{i \in \mathcal{T}} \mathbf{F}_{m,i}$.

It is sensibly known that, $\mathbf{F}_{m,i} \subseteq \Pi_i$ for any reduced order m , which ensures $FNN(m) \leq 1$.

(d) The following points should also be noted:

d.1) The initial search for close neighbors in the step-3 of Algorithm-2” is performed only once on the $N \times m$ time series from the projected trajectory, where m is a small (starting) order and N is the number of time points (generally, in the range of a few hundreds). Finding all neighbors in a set of N vectors of size m can be performed in $\mathcal{O}(N \log(m))$ or $\mathcal{O}(N)$ under mild assumptions [260, 261].

d.2) The computational complexity of the false nearest neighbors (FNN) search in the step-5 of Algorithm-2 is $\mathcal{O}(Nn)$. In addition, the FNN algorithm is suitable for parallel implementation, leading to additional reduction in the computational cost [262].

d.3) The proposed method does not require the formulation or simulation of the reduced macromodel.

d.4) Increasing the number of orthogonal basis from m to $m + 1$ keeps the initial m basis unchanged and thus requires the computation of only one new vector.

7.5 Numerical Examples

In this section, numerical results are presented to demonstrate the validity and accuracy of the proposed methodology. To serve this purpose, we consider examples of nonlinear analog circuits, exhibiting highly nonlinear dynamical behaviors, described by equations (7.1). Through these examples, it is demonstrated that, by virtue of the estimated order, the optimally minimum order reduced models ensure efficient and accurate transient behavior for the resulting reduced model.

In following numerical examples, starting from a reduced space of order two, the count of false nearest neighbors on the projected trajectory revealed in each step of the unfolding process ($m \rightarrow m + 1$) is monitored. To this end, a measure on a percentage scale (0 to 100) is defined in (7.17).

For the purpose of illustration, Proper Orthogonal Decomposition (POD) (cf. Algorithm-8) is used as the method of choice in the following numerical examples. It is to be noted that the proposed method is not limited to any specific nonlinear projection based algorithm.

7.5.1 Example I

The first example considered is the diode chain network shown in Fig. 7.13-a. The circuit exhibits significantly nonlinear characteristics and has been considered earlier in [41, 212, 213, 277, 278]. The circuit consists of $N_s = 302$ sections. The values of the resistors and capacitors are $R = 10k\Omega$ and $C = 10pF$. The diodes are characterized by equation $I_d = I_s(e^{\frac{v_d}{V_T}} - 1)$, where the saturation current and thermal voltage are given by $I_s = 10^{-14}A$ and $V_T = 0.0256V$, respectively. The state vector is taken as $\mathbf{x} = [v_1, \dots, v_N]^T$, where v_i is the voltage at node i . A sample of representative input excitations $u(t)$ is shown in

Fig. 7.13-b. Applying the proposed method, Fig. 7.14 shows the percentage of the false

Figure 7.13: (a) Diode chain circuit, (b) Excitation waveform at input.

nearest neighbors on the projected trajectory as defined in (7.17), while the dimension of the model is changed from m to $m + 1$. As seen from Fig. 7.14, since reaching $m \geq 13$ the count of false nearest neighbors drops to zero, such that, further increasing the order does not help the unfolding, and hence does not lead to revealing any new false nearest neighbors for subsequent orders. Since $m = 13$ completely unfolds the projected trajectory with no false neighbors, according to Corollary-7.1, it is selected as the optimum order.

The error between the response (trajectory) obtained from the original system $\mathbf{x}^{(org)}(\cdot)$ and its approximation from reduced macromodel $\mathbf{x}^{(pod)}(\cdot) = \mathbf{Q}\mathbf{z}(\cdot)$ is defined as

$$\text{Error in Trajectories} \triangleq \sqrt{\frac{\sum_{j=1}^N \sum_{i=1}^n \left(x_i^{(org)}(j) - x_i^{(mor)}(j) \right)^2}{n \times N}} \quad (7.18)$$

Figure 7.14: The percentage of the false nearest neighbors on the projected nonlinear trajectory.

where n is the size of the original system and N is the number of time points.

To validate that $m = 13$ is the optimum order, macromodels of consequent orders are first generated, followed by variant macromodel-based simulations. The results are compared against the full simulations. Fig. 7.15 shows the error in the response (7.18) over all the states obtained from the reduced circuit in (7.18) as function of the order of the reduced circuit. The FNN graph from Fig. 7.14 is also plotted in the same graph in Fig. 7.15 against a separate y-axis on the right. The graph clearly depicts that $m = 13$ is a minimum order to ensure the accuracy in the model such that, further increasing the order does not noticeably improve the accuracy of the model. It is important to notice that, the proposed algorithm does not require any simulation of the reduced system. Fig. 7.15 is shown only for the elaboration and validation purposes.

To verify that $m = 13$ provides an accurate reduced model to properly reproduce the

Figure 7.15: Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).

response of the nonlinear system, the original circuit is reduced using the POD algorithm with order 13. For a sample test input in Fig. 7.16 (different from the one used for POD), Fig. 7.16 compares the simulation results from the reduced circuit with the original responses. It clearly depicts an excellent agreement between the corresponding responses.

7.5.2 Example II

In this example, a circuit model for the nonlinear transmission line shown in Fig.7.17-(a) is examined. Due to its strongly nonlinear behavior, the similar network is used as the test example in most papers about nonlinear MOR [188, 189, 279]. We set all linear resistors and capacitors to have unit values, i.e., $R = R_{in} = 1$ and $C = 1$. All diodes have the constitutive equation $I_d(v) = \exp(40V_d) - 1$. The input is the current source $J(t)$ entering

Figure 7.16: Excitation test waveform at input and comparison of the responses at nodes 3, 5 and 7, respectively.

node 1.

7.5.2.1 Test-case A

First, similar to [188,279] by considering the negligible inductive effects ($L_s = 0$), the simplified nonlinear transmission line example consisting of resistors, capacitors, and diodes is tested. A sample of representative input excitation is shown in Fig.7.17-(b). The state vector is taken as $\mathbf{x} = [v_1, \dots, v_N]^T$, where v_i is the voltage at node i . Initially, we considered a network consisting of $N_s = 800$ segments, with $N = 802$ nodes.

Applying the proposed method, Fig. 7.18 shows the count of the false nearest neighbors on the projected trajectory revealed in each step of the unfolding process, while the dimension of the subspace accommodating the projected trajectory is consecutively expanding from $m - \Delta$ to m (here, $\Delta = 1$). Similar to previous example and for further illustration,

Figure 7.17: (a) Nonlinear transmission line circuit model, (b) Excitation waveform at input.

in Fig. 7.19, the error in the trajectory (7.18) obtained from the reduced circuits as a function of the order is shown. The FNN graph from Fig. 7.18 is also illustrated in the same graph, but against the y-axis on the right. As evidenced by Fig. 7.19, the reduced model attain a commendable level of accuracy only when the projected trajectory is fully unfolded. This proves $m = 15$ as the minimum order to ensure the accuracy of the model. To further demonstrate that a reduced model of order 15 accurately represents the behavior of the original system, the reduced (POD) model is tested with the different input in Fig. 7.20-(a) and the results are shown in Fig. 7.20-(b). The figure clearly depicts an excellent agreement between the corresponding responses.

Figure 7.18: The percentage of the false nearest neighbors on the projected nonlinear trajectory.

7.5.2.2 Test-case B

Let us now consider a circuit model of a nonlinear transmission line (*cf.* Fig. 7.17) comprised of $N_s = 1500$ segments where the inductors connected in series with the resistors in the segments are $L = 10$, similar to [189]. I apply the MNA formulation in order to obtain a dynamical system in form (7.1) of order $N = 4502$ with voltages at the nodes and currents in the inductors (branches) of the circuit as circuit variables. A large input signal is chosen as shown in Fig. 7.21 to ensure that, a rich nonlinear behavior is captured in the transient response. Fig. 7.22 shows the count of the false nearest neighbors on the projected trajectory while moving to increasingly larger subspaces. In Fig. 7.23 the error in transient result of the reduced macromodels with different model sizes is shown along with the FNN graph, again serving the illustrative purpose here. The two latter figures show $m = 29$ as

Figure 7.19: Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).

the minimum order for the reduced model to guaranty accuracy, signified by consistently drooping the count of the FNN to zero in the unfolding process. As a second step of verification, the transient response for the reduced model of order $m = 29$ is obtained for the input current $J(t) = 2.5(1 - \cos 5\pi t)$, which is significantly different from the POD "training" input. To further illustrate, Fig. 7.24 shows an excellent agreement between the responses from the reduced and the original systems.

Figures 7.14 and 7.18 depict that the descending pace of the percentage of false nearest neighbors is not monotonic. Hence, dropping FNN(%) to zero for just the first time is not sufficient to decide the order; but it should also remain zero for several subsequent orders.

(a)

(b)

Figure 7.20: (a) Excitation test waveform at input, (b) Comparison of the responses at nodes 5, 50, 70, and 200, respectively.

Figure 7.21: Excitation waveform at input.

Figure 7.22: The percentage of the false nearest neighbors on the projected nonlinear trajectory.

Figure 7.23: Accuracy comparison in the reduced models with different orders (left y-axis) along with the FNN (%) on the projected nonlinear trajectories (right y-axis).

Figure 7.24: Comparison of the responses at output nodes for the segments 30, 60 and 70 respectively.

Chapter 8

Conclusions and Future Work

This chapter contains a summary of the work presented in this thesis. In addition, the possible directions for future work are discussed.

8.1 Conclusions

In this thesis, several new algorithms are presented to address the important issues in the field of model order reduction for linear and nonlinear systems. The presented algorithms can be classified into two categories. The first category of algorithms address the issue of multiport reduction for linear systems. It also deals with the emerging issue of passivity preservation in the macromodelling of massively coupled multiconductor interconnect networks. The second category of algorithms provides novel ways for optimal-order determination for the reduced linear macromodels as well as the reduced nonlinear models. These methodologies ensure the accuracy and efficiency of the resulting macromodels when they are incorporated into an overall circuit simulation and undergo transient analysis.

I. Multiport reduction and clustering:

- The algorithms under this category address various challenging issues which arise in

the model order reduction of networks with large number of input/output terminals. Direct application of the conventional MOR on a multiport network often leads to an inefficient transient simulation due to the large and dense reduced models. This can easily undermine the advantage of using MOR. To address this prohibitive issue, a new, robust, and practical algorithm was presented. The proposed approach is based on the superposition paradigm for linear systems, and thereby it does not degrade the level of accuracy expected from the reduction technique of choice. The proposed algorithm results in reduced models that are sparse and block-diagonal in nature, leading to faster transient simulations. It is not limited to any specific model order reduction technique and can work in association with any existing reduction method of choice. It does not assume any correlation between the responses at ports; and thereby the algorithm overcomes the accuracy degradation that is normally associated with the existing terminal reduction techniques. An immediate application for the proposed algorithms is creating efficient reduced order macromodels for massively coupled (multiconductor) interconnect networks, such as on-chip data/address buses. For the latter application, an efficient scheme of clustering was also introduced to improve the passivity violations that may occur in macromodels.

II. Optimum order determination algorithms for reduced macromodels:

- An algorithm was devised to properly choose the order in the reduction process for linear networks, which is very important for achieving both efficiency and accuracy. Guided by geometrical considerations, the new and efficient algorithm was presented to obtain the optimal order for reduced linear models. It also identifies the redundant states from the first-level reduction techniques such as PRIMA and thus provides vital information for a second-level reduction. The application of the proposed method is not limited to a specific order reduction algorithm and can be used along with any
-

intended projection based methods for linear MOR such as: Krylov-subspace methods and TBR.

- Estimating an optimal order for the reduced nonlinear model is also of crucial importance. To this end, a novel and efficient algorithm has been presented to obtain the smallest order that ensures the accuracy and efficiency of the reduced nonlinear model. The proposed method, by deciding a proper order for the projected subspace, ensures that the reduced model can inherit the dominant dynamical characteristics of the original nonlinear system. In the proposed method, the False Nearest Neighbors (FNN) approach has been adapted to the case of nonlinear reduction to trace the deformation of nonlinear manifold in the unfolding process. The proposed method is incorporated into the projection basis generation algorithm to avoid the computational cost associated with the extra basis. The proposed nonlinear method works in conjunction with any intended nonlinear reduced modeling scheme such as: TPWL with a global reduced subspace, TBR, or POD, etc.

8.2 Future Research

1) Passivity preservation scheme:

The strict diagonal dominance of the transfer function matrix is a sufficient (but not a necessary) condition for passivity (see, Chapter 5). In practical cases, ensuring the diagonal dominance for the “Hermitian part” of hybrid transfer function matrices ($\mathbf{Z}(s)$ or $\mathbf{Y}(s)$) is too restrictive. In Chapter 5, we relaxed this condition to the block-diagonally strictly dominance. This work can be extended by investigating less restricted criterion for the enforcement of positive non-negativeness of ($\mathbf{Z}/\mathbf{Y}$) multiport transfer functions.

2) Developing an order estimation algorithm in the frequency-domain:

The proposed work can be extended to use the frequency-domain response data to determine an optimum order for the systems. To achieve this goal a methodology should be developed to analyze the coherence between neighboring samples through the state space in the frequency domain.

3) Developing an algorithm that does not need a pre-chosen neighborhood range:

The proposed work can be extended by developing a novel algorithm to reveal the false nearing neighbors in going from order m to $m + 1$. The new algorithm will be independent of any preselected threshold ρ_{FNN} . This approach is expected to decide the optimum reduced order for the resulting model more efficiently.

List of References

- [1] N. B. Tufillaro, T. Abbot, and J. Reilly, *An Experimental Approach to Nonlinear Dynamics and Chaos*. New York: Perseus Books, 1992.
- [2] E. Beltrami, *Mathematics for Dynamic Modeling*, 2nd ed. San Diego, CA: Academic Press, 1997.
- [3] S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering*. Boulder, CO: Westview Press, 1994.
- [4] T. Kailath, *Linear Systems*. New Jersey: Prentice-Hall, 1980.
- [5] C.-T. Chen, *Linear System Theory and Design*, ser. Oxford Series in Electrical and Computer Engineering, Third, Ed. New York: Oxford University Press, 1998.
- [6] L. Perko, *Differential Equations and Dynamical Systems*, 3rd ed., J. E. Marsden, L. Sirovich, and M. Golubitsky, Eds. New York: Springer, 2006.
- [7] F. Verhulst, *Nonlinear Differential Equations and Dynamical Systems*. New York: Springer, 1990.
- [8] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*. Malabar, FL: Robert E. Krieger, 1984.
- [9] W. Walter, *Ordinary Differential Equations*. New York: Springer, 1998.
- [10] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*. Englewood Cliffs, NJ: Prentice Hall, 1973.
- [11] D. G. Luenberger, "Dynamic equations in descriptor form," *IEEE Transactions on Automatic Control*, vol. 22, no. 3, pp. 312–321, Jun. 1977.
- [12] S. L. Campbell, *Singular Systems of Differential Equations*, ser. Research Notes in Mathematics. London, UK: Pitman Publishing, 1980, vol. 40.

-
- [13] —, *Singular Systems of Differential Equations II*, ser. Research Notes in Mathematics. London, UK: Pitman Publishing, 1982, vol. 61.
 - [14] J. R. Phillips, “Projection-based approaches for model reduction of weakly nonlinear, time-varying systems,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 2, pp. 171–187, Feb. 2003.
 - [15] R. Johansson, *System Modeling and Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
 - [16] G.-R. Duan, *Analysis and Design of Descriptor Linear Systems*, ser. Advances in Mechanics and Mathematics. New York: Springer, 2010, vol. 23.
 - [17] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
 - [18] D. G. Luenberger, “Nonlinear descriptor systems,” *Journal of Economic dynamics control 1*, no. 3, pp. 219–242, Apr. 1979.
 - [19] N. Tufillaro, “A dynamical systems approach to behavioral modeling,” Hewlett-Packard, Integrated Solutions Laboratory, HP Laboratories Palo Alto, Tech. Rep. HPL-1999-22, Feb. 1999.
 - [20] J. K. Hedrick and A. Girard, “Control of nonlinear dynamic systems: Theory and applications,” Berkeley University of California, Berkeley, CA, Course notes (ME-237), Spring 2010.
 - [21] R. A. Rohrer, *Circuit Theory: an introduction to the State Variable Approach*. New York: McGraw-Hill, 1970.
 - [22] S. Smale, “On the mathematical foundation of electrical circuit theory,” *Journal of Differential Geometry*, vol. 7, no. 1-2, pp. 193–210, 1972.
 - [23] E. S. Kuh and R. A. Rohrer, “The state-variable approach to network analysis,” *Proc. IEEE*, vol. 53, no. 7, pp. 672–686, Jul. 1965.
 - [24] L. O. Chua, “State variable formulation of nonlinear rlc networks in explicit normal form,” *Proc. IEEE*, vol. 53, no. 2, pp. 206–207, Feb. 1965.
 - [25] L. Chua and R. Rohrer, “On the dynamic equations of a class of nonlinear RLC networks,” *IEEE Transactions on Circuit Theory*, vol. 12, no. 4, pp. 475–489, Dec. 1965.
-

-
- [26] L. O. Chua, "On the choice of state variables for the potential function in nonlinear networks," *Proc. IEEE*, vol. 53, no. 12, p. 2110, Dec. 1965.
 - [27] R. K. Brayton and J. K. Moser, "A theory of nonlinear networks Parts I & II," *Quarterly of Applied Mathematics*, vol. 22, no. 1,2, pp. 1–33, 81–104, Apr./Jul. 1964.
 - [28] T. Matsumoto, "On several geometric aspects of nonlinear networks," *Journal of the Franklin Institute*, vol. 301, no. 1-2, pp. 203–225, Jan.-Feb. 1976.
 - [29] S. Ichiraku, "Connecting electrical circuits: Transversality and well-posedness," *Yokohama mathematical Journal*, vol. 27, no. 5, pp. 111–126, 1979.
 - [30] T. Matsumoto, L. Chua, H. Kawakami, and S. Ichiraku, "Geometric properties of dynamic nonlinear networks: Transversality, local-solvability and eventual passivity," *IEEE Transactions on Circuits and Systems*, vol. CAS-28, no. 5, pp. 406–428, May 1981.
 - [31] A. Szatkowski, "On the dynamics of non-linear RLC networks from a geometric point of view," *International Journal of Circuit Theory and Applications*, vol. 11, no. 2, pp. 117–129, Apr. 1983.
 - [32] F. H. Branin, "Computer methods of network analysis," *Proc. IEEE*, vol. 55, no. 11, pp. 1787–1801, Nov. 1967.
 - [33] S. R. Sedore, "SCEPTRE: a second generation transient analysis program," in *Proc. Computer-Aided Circuit Design Seminar*, Cambridge, MA, 1967, pp. 55–61.
 - [34] C. W. Gear, "The automatic integration of stiff ordinary differential equations," in *Proc. the IFIPS Congress*, 1968, pp. A81–A85.
 - [35] I. W. Sandberg and H. Shichman, "Numerical integration of stiff nonlinear differential equation," *Bell System Technical Journal*, vol. 47, no. MR38 #1836, pp. 511–527, Apr. 1968.
 - [36] L. R. Petzold, "Differential/algebraic equations are not ODEs," *SIAM Journal on Scientific and Statistical Computing*, vol. 3, no. 3, pp. 367–384, Sep. 1982.
 - [37] J. Vlach and K. Singhal, *Computer Methods for Circuit Analysis and Design*, 2nd ed. Boston, MA: Kluwer Academic Publishers, 2003.
 - [38] C.-W. Ho, A. Ruehli, and P. Brennan, "The modified nodal approach to network analysis," *IEEE Transactions on Circuits and Systems*, vol. 22, no. 6, pp. 504–509, Jun. 1975.
-

-
- [39] L. T. Pillage, R. A. Rohrer, and C. Visweswariah, *Electronic Circuit and System Simulation Methods*. New York: McGraw-Hill, 1994.
- [40] M. Striebel and J. Rommes, “Model order reduction of nonlinear systems: status, open issues, and applications,” Technischen Universitat Chemnitz, Chemnitz, Germany, Tech. Rep. CSC/08-07, Nov. 2008.
- [41] A. Verhoeven, M. Striebel, J. Rommes, E. Maten, and T. Bechtold, “Model order reduction for nonlinear IC models with POD,” in *Progress in Industrial Mathematics at ECMI 2008*, ser. Mathematics in Industry, A. D. Fitt, J. Norbury, H. Ockendon, and E. Wilson, Eds. Berlin, Germany: Springer, 2010, pp. 441–446.
- [42] Y. Chen, “Model order reduction for nonlinear systems,” Master thesis, Massachusetts Institute of Technology, Cambridge, MA, Sep. 1999.
- [43] A. Verhoeven, E. J. ter Maten, M. Striebel, and R. Mattheij, “Model order reduction for nonlinear IC models,” in *System Modeling and Optimization*, ser. IFIP Advances in Information and Communication Technology, A. Korytowski, K. Malanowski, W. Mitkowski, and M. Szymkat, Eds. Berlin, Germany: Springer, 2009, vol. 312, pp. 476–491.
- [44] W. H. A. Schilders, H. A. van der Vorst, and J. Rommes, Eds., *Model Order Reduction: Theory, Research Aspects and Applications*. Berlin, Germany: Springer, 2008.
- [45] N. Gershenfeld, *The nature of mathematical modeling*. Cambridge, UK: Cambridge University Press, 1999.
- [46] E. A. Bender, *An Introduction to Mathematical Modeling*. New York: Dover, 2000.
- [47] R. Achar and M. Nakhla, “Simulation of high-speed interconnects,” *IEEE Proc.*, vol. 89, no. 5, pp. 693–728, May 2001.
- [48] D. Vasilyev, “Theoretical and practical aspects of linear and nonlinear model order reduction techniques,” Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2008.
- [49] D. M. Walker, R. Brown, and N. B. Tufillaro, “A constructing transportable behavioral models for nonlinear electronic devices,” *Physics Letters A*, vol. 255, no. 4-6, pp. 236–242, Feb. 1999.
-

- [50] A. L. Walker, "Behavioral modeling and characterization of nonlinear operation in rf and microwave systems," Ph.D. dissertation, North Carolina State University, Raleigh, NC, 2005.
 - [51] D. E. R. John Wood, *Fundamentals of Nonlinear Behavioral Modeling for RF and Microwave Design*. Boston, MA: Artech House, 2005.
 - [52] D. E. Root, J. Wood, and N. Tufillaro, "New techniques for non-linear behavioral modeling of microwave/RF ICs from simulation and nonlinear microwave measurements," in *Proc. 40th annual Design Automation Conference*, Anaheim, CA, 2003, pp. 85–90.
 - [53] J. Wood, D. Root, and N. Tufillaro, "A behavioral modeling approach to nonlinear model-order reduction for RF/microwave ICs and systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, pp. 2274–2284, Sep. 2004.
 - [54] J. Wood and D. Root, "The behavioral modeling of microwave/rf ics using non-linear time series analysis," in *Proc. IEEE/MTT-S International Microwave Symposium Digest*, vol. 2, Jun. 2003, pp. 791–794.
 - [55] S. Sastry, *Nonlinear Systems: Analysis, Stability, and Control*. New York: Springer, 1999.
 - [56] P. G. Drazin, *Nonlinear Systems*. Cambridge, MA: Cambridge University Press, 1992.
 - [57] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Boston, MA: Birkhäuser, 1997.
 - [58] W. J. Rugh, *Linear System Theory*, ser. second. Upper Saddle River, NJ: Prentice Hall, 1996.
 - [59] Y.-L. Jiang and H.-B. Chen, "Time domain model order reduction of general orthogonal polynomials for linear input-output systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 330–343, Feb. 2012.
 - [60] J. Cong, L. He, C.-K. Koh, and P. H. Madden, "Performance optimization of VLSI interconnect layout," *Integration, the VLSI Journal*, vol. 21, no. 1&2, pp. 1–94, Nov. 1996.
 - [61] C.-K. Cheng, J. Lillis, S. Lin, and N. Chang, *Interconnect analysis and synthesis*. Hoboken, NJ: John Wiley, 1999.
-

-
- [62] S. Lee, P. Roblin, and O. Lopez, "Modeling of distributed parasitics in power FETs," *IEEE Transactions on Electron Devices*, vol. 49, no. 10, pp. 1799–1806, Oct. 2002.
- [63] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA: SIAM, 2005.
- [64] M. Celik, L. Pileggi, and A. Odabasioglu, *IC Interconnect Analysis*. Norwell, MA: Kluwer, 2002.
- [65] S. X.-D. Tan and L. He, *Advanced Model Order Reduction Techniques in VLSI Design*. Cambridge, MA: Cambridge University Press, 2007.
- [66] P. Benner, M. Hinze, and E. J. W. ter Maten, Eds., *Model Reduction for Circuit Simulation*. Berlin, Germany: Springer, 2011.
- [67] R. W. Freund, "Reduced-order modeling techniques based on Krylov subspaces and their use in circuit simulation," in *Applied and Computational Control, Signals, and Circuits*, B. N. Datta, Ed. Boston, MA: Birkhäuser Boston, 1999, vol. 1, ch. 9, pp. 435–498.
- [68] B. Men, Q. Zhang, X. Li, C. Yang, and Y. Chen, "The stability of linear descriptor systems," *International Journal of information and systems sciences*, vol. 2, no. 3, pp. 362–374, Nov. 2006.
- [69] T. Stykel, "Analysis and numerical solution of generalized lyapunov equations," Ph.D. dissertation, Technische universität Berlin, Berlin, Germany, Jan. 2002.
- [70] D. Debeljković, N. Višnjić, and M. R. Pješćić, "The stability of linear continuous singular systems in the sense of lyapunov: An overview," *Scientific Technical Reviews*, vol. LVII, no. 1, pp. 51–64, 2007.
- [71] M. R. Wohlers, *Lumped and Distributed Passive Networks*, H. G. Booker and N. Declaris, Eds. New York: Academic Press, 1969.
- [72] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.
- [73] P. Triverio, S. Grivet-Talocia, M. Nakhla, F. Canavero, and R. Achar, "Stability, causality, and passivity in electrical interconnect models," *IEEE Transactions on Advanced Packaging*, vol. 30, no. 4, pp. 795–808, Nov. 2007.
-

-
- [74] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. New Jersey: Princeton University Press, 2009.
- [75] T. Lyche, “Numerical linear algebra,” University of Oslo, Course Notes Inf-Mat 4350, Aug. 2012.
- [76] L. Pillage and R. Rohrer, “Asymptotic waveform evaluation for timing analysis,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 9, no. 4, pp. 352–366, Apr. 1990.
- [77] V. Raghavan, R. Rohrer, L. Pillage, J. Lee, J. Bracken, and M. Alaybeyi, “AWE-inspired,” in *Proc. IEEE Custom Integrated Circuits Conference*, May 1993, pp. 18.1.1–18.1.8.
- [78] E. Chiprout and M. S. Nakhla, *Asymptotic Waveform Evaluation*. Boston, MA: Kluwer Academic Publishers, 1994.
- [79] W. C. Elmore, “The transient analysis of damped linear networks with particular regard to wideband amplifiers,” *Journal of Applied Physics*, vol. 9, pp. 55–63, Jan. 1948.
- [80] E. Chiprout and M. S. Nakhla, “Analysis of interconnect networks using complex frequency hopping,” *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 2, pp. 186–200, Feb. 1995.
- [81] P. Feldmann and R. W. Freund, “Efficient linear circuit analysis by Padé approximation via the Lanczos process,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 5, pp. 639–649, May 1995.
- [82] R. W. Freund, “Krylov-subspace methods for reduced-order modeling in circuit simulation,” *Journal of Computational and Applied Mathematics*, vol. 123, no. 1-2, pp. 395–421, Nov. 2000.
- [83] ———, “SPRIM: structure-preserving reduced-order interconnect macromodeling,” in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2004, pp. 80–87.
- [84] L. M. Silveira, M. Kamon, and J. White, “Efficient reduced-order modeling of frequency-dependent coupling inductances associated with 3-D interconnect structures,” *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B*, vol. 19, no. 2, pp. 283–288, May 1996.
-

-
- [85] A. Odabasioglu, M. Celik, and L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 8, pp. 645–654, Aug. 1998.
- [86] N. L. Schryer, "A tutorial on galerkin's method, using b-splines, for solving differential equations," Bell Laboratories, Murray Hill, NJ, Tech. Rep. 52, Sep. 1976.
- [87] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," *The American Institute of Aeronautics and Astronautics Journal*, vol. 40, no. 11, pp. 2323–2330, Nov. 2002.
- [88] C. B. Moore, "Principal component analysis in linear systems: Controllability, Observability, and Model Reduction," *IEEE Transaction on Automatic Control*, vol. AC-26, pp. 17–32, Feb. 1981.
- [89] C. de Villemagne and R. E. Skelton, "Model reductions using a projection formulation," in *Proc. 26th IEEE Conference on Decision and Control*, Los Angeles, CA, 1987, pp. 461–466.
- [90] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, University of Illinois, Urbana and Champaign, IL, 1997.
- [91] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Applied Numerical Mathematics*, vol. 43, no. 1-2, pp. 9–44, 2002.
- [92] W. E. Arnoldi, "The principle of minimized iteration in the solution of the matrix eigenvalue problem," *The Quarterly of Applied Mathematics*, vol. 9, no. 17, pp. 17–29, 1951.
- [93] I. M. Elfadel and D. D. Ling, "A block rational Arnoldi algorithm for multipoint passive model-order reduction of multiport RLC networks," in *Proc. IEEE/ACM international conference on Computer-aided design*, San Jose, CA, 1997, pp. 66–71.
- [94] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003.
- [95] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," *Journal of Research of the National Bureau of Standards*, vol. 45, no. 4, pp. 255–282, Oct. 1950.
-

- [96] E. J. Grimme, D. C. Sorensen, and P. V. Dooren, "Model reduction of state space systems via an implicitly restarted lanczos method," *Numerical Algorithms*, vol. 12, pp. 1–31, 1996.
 - [97] P. Feldmann and R. W. Freund, "Reduced-order modeling of large linear subcircuits via a block lanczos algorithm," in *Proc. 32nd annual IEEE/ACM Design Automation Conference*, San Francisco, CA, Jun. 1995, pp. 474–479.
 - [98] J. I. Aliaga, D. L. Boley, R. W. Freund, and V. Hernández, "A lanczos-type method for multiple starting vectors," *Mathematics of Computation*, vol. 69, no. 232, pp. 1577–1601, May 2000.
 - [99] R. W. Freund and P. Feldmann, "Reduced-order modeling of large passive linear circuits by means of the SyPVL algorithm," in *Proc. IEEE/ACM international conference on Computer-aided design*, San Jose, CA, Nov. 1996, pp. 280–287.
 - [100] —, "The SyMPVL algorithm and its applications to interconnect simulation," in *Proc. International Conference on Simulation of Semiconductor Processes and Devices*, Cambridge, MA, Sep. 1997, pp. 113–116.
 - [101] K. Glover, "All optimal hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_\infty$ -error bounds," *International Journal of Control*, vol. 36, no. 6, pp. 1115–1193, Jun. 1984.
 - [102] D. F. Enns, "Model reduction with balanced realizations: An error bound and a frequency weighted generalization," in *Proc. 23rd IEEE Conference on Decision and Control*, vol. 23, Dec. 1984, pp. 127–132.
 - [103] A. Laub, M. Heath, C. Paige, and R. Ward, "Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms," *IEEE Transactions on Automatic Control*, vol. 32, no. 2, pp. 115–122, Feb. 1987.
 - [104] M. G. Safonov and R. Y. Chiang, "A schur method for balanced-truncation model reduction," *IEEE Transaction Automatic Control*, vol. 34, pp. 729–733, Jul. 1989.
 - [105] C.-A. Lin and T.-Y. Chiu, "Model reduction via frequency weighted balanced realization," in *Proc. American Control Conference*, May 1990, pp. 2069–2070.
 - [106] V. Sreeram and B. Anderson, "Frequency weighted balanced reduction technique: a generalization and an error bound," in *Proc. 34th IEEE Conference on Decision and Control*, vol. 4, Dec. 1995, pp. 3576–3581.
-

-
- [107] G. Wang, V. Sreeram, and W. Liu, "A new frequency-weighted balanced truncation method and an error bound," in *Proc. 36th IEEE Conference on Decision and Control*, vol. 4, Dec. 1997, pp. 3329–3334.
- [108] J.-R. Li and J. White, "Efficient model reduction of interconnect via approximate system gramians," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, 1999, pp. 380–383.
- [109] J.-R. Li, F. Wang, and J. White, "An efficient lyapunov equation-based approach for generating reduced-order models of interconnect," in *Proc. 36th Design Automation Conference*, 1999, pp. 1–6.
- [110] V. Sreeram, "On the properties of frequency weighted balanced truncation techniques," in *Proc. American Control Conference*, vol. 3, no. 1753–1754, 2002.
- [111] J. Phillips, L. Daniel, and L. M. Silveira, "Guaranteed passive balancing transformations for model order reduction," in *Proc. 39th Annual Design Automation Conference*, New Orleans, LA, 2002, pp. 52–57.
- [112] J. Phillips, L. Daniel, and L. Silveira, "Guaranteed passive balancing transformations for model order reduction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 8, pp. 1027–1041, Aug. 2003.
- [113] J. Phillips and L. M. Silveira, "Poor man's TBR" a simple model reduction scheme," in *Proc. European Design and Test Conference*, 2004, pp. 938–943.
- [114] J. Phillips and L. Silveira, "Poor man's TBR: a simple model reduction scheme," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 1, pp. 43–55, Jan. 2005.
- [115] N. Wong and V. Balakrishnan, "Fast balanced stochastic truncation via a quadratic extension of the alternating direction implicit iteration," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, 2005, pp. 801–805.
- [116] P. Heydari and M. Pedram, "Model-order reduction using variational balanced truncation with spectral shaping," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 4, pp. 879–891, Apr. 2006.
- [117] G. H. Golub and C. F. van Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.
-

-
- [118] J.-R. Li and J. White, “Low rank solution of lyapunov equations,” *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 1, pp. 260–280, Jan. 2002.
- [119] P. Benner, J.-R. Li, and T. Penzl, “Numerical solution of large-scale lyapunov equations, riccati equations, and linear-quadratic optimal control problems,” *Numerical Linear Algebra with Applications*, vol. 15, no. 9, pp. 755–777, 2008.
- [120] A. Lu and E. Wachspress, “Solution of lyapunov equations by alternating direction implicit iteration,” *Computers and Mathematics with Applications*, vol. 21, no. 9, pp. 43–58, 1991.
- [121] N. S. Ellner and E. L. Wachspress, “Alternating direction implicit iteration for systems with complex spectra,” *SIAM Journal on Numerical Analysis*, vol. 28, no. 3, pp. 859–870, 1991.
- [122] S. V. Brian D. O. Anderson, *Network Analysis and Synthesis: a Modern Systems Theory Approach*, 2nd ed. New York: Dover, 2006.
- [123] Arnold, W.F., III and A. Laub, “Generalized eigen problem algorithms and software for algebraic riccati equations,” *Proc. IEEE*, vol. 72, no. 12, pp. 1746–1754, Dec. 1984.
- [124] G. Wang, V. Sreeram, and W. Liu, “A new frequency-weighted balanced truncation method and an error bound,” *IEEE Transactions on Automatic Control*, vol. 44, no. 9, pp. 1734–1737, Sep. 1999.
- [125] M. Loève, *Probability Theory I*, 4th ed. New York: Springer, 1977.
- [126] H. Hotelling, “Simplified calculation of principal components,” *Psychometrika*, vol. 1, no. 1, pp. 27–35, 1936.
- [127] G. Berkooz, P. Holmes, and J. L. Lumley, “The proper orthogonal decomposition in the analysis of turbulent flows,” *Annual Review of Fluid Mechanics*, vol. 25, pp. 539–575, Jan. 1993.
- [128] T. Kim, “Frequency-domain karhunen-loève method and its application to linear dynamic systems,” *American Institute of Aeronautics and Astronautics Journal*, vol. 36, no. 11, pp. 2117–2123, 1998.
- [129] K. E. Willcox, J. Paduano, J. Peraire, and K. Hall, “Low order aerodynamic models for aeroelastic control of turbomachines,” in *Proc. on the American Institute of*
-

Aeronautics and Astronautics, Structural Dynamics and Materials (SDM) Conference, vol. 3, no. 2204–2214, St Louis, MO, Apr. 1999.

- [130] K. C. Hall, E. H. Dowell, and J. P. Thomas, “Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows,” *American Institute of Aeronautics and Astronautics Journal*, vol. 38, no. 10, pp. 1853–1862, 2000.
 - [131] K. Willcox, “Reduced-order aerodynamic models for aeroelastic control of turbomachines,” Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2000.
 - [132] A. George, J. R. Gilbert, and J. W. H. Liu, *Graph Theory and Sparse Matrix Computation*. New York: Springer, 1993.
 - [133] A. Antoulas, D. C. Sorensen, and S. Gugercin, “A survey of model reduction methods for large-scale systems,” in *Structured Matrices in Mathematics, Computer Science, and Engineering I*, V. Olshevsky, Ed. American Mathematical Society, 2001, ch. III, pp. 193–219.
 - [134] A. J. Sasane, “Hankel norm approximation for infinite-dimensional systems,” Ph.D. dissertation, University of Groningen, Netherlands, Sep. 2001.
 - [135] A. Megretski, “Model order reduction,” Massachusetts Institute of Technology, Cambridge, MA, Course Notes (6.242), Fall 2004.
 - [136] Y. Liu, L. Pileggi, and A. Strojwas, “Model order-reduction of RC(L) interconnect including variational analysis,” in *Proc. 36th Design Automation Conference*, 1999, pp. 201–206.
 - [137] R. Colgren, “Efficient model reduction for the control of large-scale systems,” in *Efficient Modeling and Control of Large-Scale Systems*, J. Mohammadpour and K. M. Grigoriadis, Eds. New York: Springer, 2010, ch. 2, pp. 59–72.
 - [138] C. Coelho, J. Phillips, and L. Silveira, “Robust rational function approximation algorithm for model generation,” in *Proc. 36th Design Automation Conference*, 1999, pp. 207–212.
 - [139] B. Gustavsen and A. Semlyen, “Rational approximation of frequency domain responses by vector fitting,” *IEEE Transactions on Power Delivery*, vol. 14, no. 3, pp. 1052–1061, Jul. 1999.
-

- [140] D. Deschrijver, B. Haegeman, and T. Dhaene, "Orthonormal vector fitting: A robust macromodeling tool for rational approximation of frequency domain responses," *IEEE Transactions on Advanced Packaging*, vol. 30, no. 2, pp. 216–225, May 2007.
 - [141] S. Grivet-Talocia, "The time-domain vector fitting algorithm for linear macromodeling," *AEU-International Journal of Electronics and Communications*, vol. 58, no. 4, pp. 293–295, 2004.
 - [142] S. Grivet-Talocia, F. G. Canavero, I. S. Stievano, and I. A. Maio, "Circuit extraction via time-domain vector fitting," in *Proc. International Symposium on Electromagnetic Compatibility*, vol. 3, Aug. 2004, pp. 1005–1010.
 - [143] S. Grivet-Talocia, "Package macromodeling via time-domain vector fitting," *IEEE Microwave and Wireless Components Letters*, vol. 13, no. 11, pp. 472–474, Nov. 2003.
 - [144] Y. Mekonnen and J. Schutt-Aine, "Broadband macromodeling of sampled frequency data using z-domain vector-fitting method," in *Proc. IEEE Workshop on Signal Propagation on Interconnects*, Genova, Italy, May 2007, pp. 45–48.
 - [145] B. Nouri, R. Achar, and M. S. Nakhla, "z-Domain orthonormal basis functions for physical system identifications," *IEEE Transactions on Advanced Packaging*, vol. 33, no. 1, pp. 293–307, Feb. 2010.
 - [146] B. Gustavsen and A. Semlyen, "Simulation of transmission line transients using vector fitting and modal decomposition," *IEEE Transactions on Power Delivery*, vol. 13, no. 2, pp. 605–614, Apr. 1998.
 - [147] B. Gustavsen, "Relaxed vector fitting algorithm for rational approximation of frequency domain responses," in *Proc. 10th IEEE Workshop Signal Propagation Interconnects*, Berlin, Germany, May 2006, p. 97–100.
 - [148] —, "Improving the pole relocating properties of vector fitting," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1587–1592, Jul. 2006.
 - [149] D. Deschrijver, M. Mrozowski, T. Dhaene, and D. De Zutter, "Macromodeling of multiport systems using a fast implementation of the vector fitting method," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 6, pp. 383–385, Jun. 2008.
 - [150] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme, "Parametric identification of transfer functions in the frequency domain—a survey," *IEEE Transactions on Automatic Control*, vol. 39, no. 11, pp. 2245–2260, Nov. 1994.
-

-
- [151] S.-B. Nouri, "Advanced macromodeling algorithm for sampled time/frequency domain measured/tabulated data," Master's thesis, Carleton University, Ottawa, Canada, Feb. 2008.
- [152] E. C. Levi, "Complex curvefitting," *IEEE Transactions on Automatic Control*, vol. AC-4, no. 1, pp. 37–43, Jan. 1959.
- [153] C. K. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Transactions on Automatic Control*, vol. 8, no. 1, pp. 56–58, Jan. 1963.
- [154] B. Nouri, R. Achar, M. S. Nakhla, and D. Saraswat, "z-Domain orthonormal vector fitting for macromodeling high-speed modules characterized by tabulated data," in *Proc. 12th IEEE Workshop on Signal Propagation on Interconnects*, Avignon, France, May 2008, pp. 1–4.
- [155] K. C. Sou, A. Megretski, and L. Daniel, "A quasi-convex optimization approach to parameterized model order reduction," in *Proc. 42nd Design Automation Conference*, Jun. 2005, pp. 933–938.
- [156] —, "A quasi-convex optimization approach to parameterized model order reduction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 3, pp. 456–469, Mar. 2008.
- [157] C. Beattie and S. Gugercin, "Interpolatory projection methods for structure-preserving model reduction," *Systems & Control Letters*, vol. 58, no. 3, pp. 225–232, 2009.
- [158] S. Lefteriu and A. C. Antoulas, "A new approach to modeling multiport systems from frequency-domain data," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 1, pp. 14–27, Jan. 2010.
- [159] M. Kamon, F. Wang, and J. White, "Generating nearly optimally compact models from krylov-subspace based reduced-order models," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 4, pp. 239–248, Apr. 2000.
- [160] A. Antoulas and D. C. Sorensen, "Approximation of large-scale dynamical system: An overview," *International Journal of Applied Mathematics and Computer Science*, vol. 11, no. 5, pp. 1093–1121, 2001.
-

-
- [161] J. A. Martínez, “Model order reduction of nonlinear dynamic systems using multiple projection bases and optimization state-space sampling,” Ph.D. dissertation, University of Pittsburgh, Pittsburgh, PA, Jun. 2009.
- [162] S. Gugercin and A. C. Antoulas, “Model reduction of large-scale systems by least squares,” *Linear Algebra and its Applications*, vol. 415, no. 2, pp. 290–321, 2006.
- [163] M. F. Selivanov and Y. A. Chernoiyan, “A combined approach of the laplace transform and padé approximation solving viscoelasticity,” *International Journal of Solids and Structures problems*, vol. 44, no. 1, pp. 66–76, Mar. 2007.
- [164] Y. Kawamura, “IEEE transactions on circuits and systems,” *A formulation of nonlinear dynamic networks*, vol. 25, no. 2, pp. 88–98, Feb. 1978.
- [165] T. Matsumoto, “On the dynamics of electrical networks,” *Journal of Differential Equations*, vol. 21, no. 1, pp. 179–196, May 1976.
- [166] L. O. Chua, “Dynamic nonlinear networks: State-of-the-art,” *IEEE Transactions on Circuits and Systems*, vol. CAS-27, no. 11, pp. 1059–1087, Nov. 1980.
- [167] F. N. Najm, *Circuit Simulation*. Hoboken, NJ: Wiley-IEEE Press, 2010.
- [168] J.-M. Ginoux, *Differential Geometry Applied to Dynamical Systems*, L. O. Chua, Ed. Hackensack, NJ: World Scientific, 2009.
- [169] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems and an Introduction to Chaos*, 2nd ed. San Diego, CA: Academic Press, 2004.
- [170] B. N. Bond, “Stability-preserving model reduction for linear and nonlinear systems arising in analog circuit applications,” Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2010.
- [171] B. N. Bond and L. Daniel, “Stable reduced models for nonlinear descriptor systems through piecewise-linear approximation and projection,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 10, pp. 1467–1480, Oct. 2009.
- [172] —, “Stabilizing schemes for piecewise-linear reduced order models via projection and weighting functions,” in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, Nov. 2007, pp. 860–867.
-

- [173] D. Amsallem and C. Farhat, “Stabilization of projection-based reduced-order models,” *International Journal for Numerical Methods in Engineering*, vol. 91, no. 4, pp. 358–377, Jul. 2012.
 - [174] —, “On the stability of reduced-order linearized computational fluid dynamics models based on POD and galerkin projection: descriptor versus non-descriptor forms,” in *Reduced Order Methods for Modeling and Computational Reduction*, ser. MS&A Series, A. Quarteroni and G. Rozza, Eds. New York: Springer, 2014, vol. 9, pp. 215–233.
 - [175] M. Vidyagar, *Nonlinear Systems Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1993.
 - [176] J. R. Phillips, “Projection frameworks for model reduction of weakly nonlinear systems,” in *Proc. Design Automation Conference*, Los Angeles, CA, 2000, pp. 184–189.
 - [177] A. Mohammad and J. A. De Abreu-Garcia, “A transformation approach for model order reduction of nonlinear systems,” in *Proc. 16th Annual Conference of IEEE Industrial Electronics Society*, vol. 1, Pacific Grove, CA, Nov. 1990, pp. 380–383.
 - [178] E. Hung, Y.-J. Yang, and S. Senturia, “Low-order models for fast dynamical simulation of MEMS microstructures,” in *Proc. International Conference on Solid State Sensors and Actuators*, vol. 2, Chicago, IL, Jun. 1997, pp. 1101–1104.
 - [179] J. Chen and S.-M. Kang, “An algorithm for automatic model-order reduction of nonlinear mems devices,” in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 2, 2000, pp. 445–448.
 - [180] Y. Chen and J. White, “A quadratic method for nonlinear model order reduction,” in *Proc. International Conference on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators*, San Diego, CA, Mar. 2000, pp. 477–480.
 - [181] Y. Su, J. Wang, X. Zeng, Z. Bai, C. Chiang, and D. Zhou, “SAPOR: second-order arnoldi method for passive order reduction of RCS circuits,” in *Proc. IEEE/ACM International Conference on Computer Aided Design*, Washington, DC, Nov. 2004, pp. 74–79.
 - [182] B. Liu, X. Zeng, Y. Su, J. Tao, Z. Bai, C. Chiang, and D. Zhou, “Block SAPOR:
-

-
- block second-order arnoldi method for passive order reduction of multi-input multi-output RCS interconnect circuits,” in *Proc. Asia and South Pacific Design Automation Conference*, vol. 1, Jan. 2005, pp. 244–249.
- [183] Y. Shi, H. Yu, and L. He, “SAMSON: a generalized second-order arnoldi method for reducing multiple source linear network with susceptance,” in *Proc. international symposium on Physical design*, San Jose, CA, Apr. 2006, pp. 25–32.
- [184] L. Feng, “Review of model order reduction methods for numerical simulation of nonlinear circuits,” *Applied Mathematics and Computation*, vol. 167, no. 1, pp. 576–591, Aug. 2005.
- [185] W. J. Rugh, *Nonlinear System Theory, The Volterra/Wiener Approach*. Baltimore, MD: Johns Hopkins University Press, 1981.
- [186] M. Condon and R. Ivanov, “Krylov subspaces from bilinear representations of nonlinear systems,” *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 26, no. 2, pp. 399–406, 2007.
- [187] M. J. Rewieński and J. White, “A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices,” in *Proc. IEEE/ACM International Conference on Conference on Computer-Aided Design*, San Jose, CA, Nov. 2001, pp. 252–257.
- [188] —, “A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 2, pp. 155–170, Feb. 2003.
- [189] D. Vasilyev, M. J. Rewieński, and J. White, “A TBR-based trajectory piecewise-linear algorithm for generating accurate low-order models for nonlinear analog circuits and MEMS,” in *Proc. Design Automation Conference*, Jun. 2003, pp. 490–495.
- [190] —, “Macromodel generation for BioMEMS components using a stabilized balanced truncation plus trajectory piecewise-linear approach,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 2, pp. 285–293, Feb. 2006.
- [191] S. Tiwary and R. Rutenbar, “Scalable trajectory methods for on-demand analog macromodel extraction,” in *Proc. 42nd Design Automation Conference*, Jun. 2005, pp. 403–408.
-

- [192] —, “Faster, parametric trajectory-based macromodels via localized linear reductions,” in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2006, pp. 876–883.
 - [193] —, “On-the-fly fidelity assessment for trajectory-based circuit macromodels,” in *Proc. IEEE Custom Integrated Circuits Conference*, San Jose, CA, Sep. 2006, pp. 185–188.
 - [194] T. Bechtold, M. Striebel, K. Mohaghegh, and E. J. W. ter Maten, “Nonlinear model order reduction in nanoelectronics: Combination of POD and TPWL,” *Proc. Applied Mathematics and Mechanics*, vol. 8, pp. 10 057–10 060, May 2008.
 - [195] S. Dabas, N. Dong, and J. Roychowdhury, “Automated extraction of accurate delay/timing macromodels of digital gates and latches using trajectory piecewise methods,” in *Proc. Design Automation Conference, Asia and South Pacific*, Yokohama, Japan, Jan. 2007, p. 361–366.
 - [196] N. Dong and J. Roychowdhury, “General-purpose nonlinear model-order reduction using piecewise-polynomial representations,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 249–264, Feb. 2008.
 - [197] Y. Zhang, N. Fong, and N. Wong, “Piecewise-polynomial associated transform macromodeling algorithm for fast nonlinear circuit simulation,” in *Proc. 18th Asia and South Pacific Design Automation Conference*, Yokohama, Japan, Jan. 2013, pp. 515–520.
 - [198] M. Farooq, L. Xia, F. Hussin, and A. Malik, “Automated model generation of analog circuits through modified trajectory piecewise linear approach with chebyshev newton interpolating polynomials,” in *Proc. 4th International Conference on Intelligent Systems Modelling Simulation*, Jan. 2013, pp. 605–609.
 - [199] M. J. Rewieński, “A trajectory piecewise-linear approach to model order reduction of nonlinear dynamical systems,” Ph.D. dissertation, Massachusetts Institute of Technology MIT, Cambridge, MA, Jun. 2003.
 - [200] T. Voß, “Model reduction for nonlinear differential algebraic equations,” Masters thesis, University of Wuppertal, Wuppertal, Germany, 2005.
-

- [201] A. Vollebregt, T. Bechtold, A. Verhoeven, and E. ter Maten, "Model order reduction of large scale ODE systems: MOR for ANSYS versus ROM workbench," in *Scientific Computing in Electrical Engineering*, ser. Mathematics in Industry, G. Ciuprina and D. Ioan, Eds. Berlin, Heidelberg: Springer, 2007, vol. 11, pp. 175–182.
 - [202] R. Ionutiu, S. Lefteriu, and A. C. Antoulas, "Comparison of model reduction methods with applications to circuit simulation," in *Scientific Computing in Electrical Engineering*, ser. Mathematics in Industry, G. Ciuprina and D. Ioan, Eds. Berlin Heidelberg: Springer, 2007, vol. 11, pp. 3–24.
 - [203] T. Voß, A. Verhoeven, T. Bechtold, and J. Maten, "Model order reduction for nonlinear differential algebraic equations in circuit simulation," in *Progress in Industrial Mathematics at ECMI 2006*, ser. Mathematics in Industry, L. Bonilla, M. Moscoso, G. Platero, and J. Vega, Eds. Berlin, Heidelberg: Springer, 2008, vol. 12, pp. 518–523.
 - [204] K. Mohaghegh, M. Striebel, E. ter Maten, and R. Pulch, "Nonlinear model order reduction based on trajectory piecewise linear approach: Comparing different linear cores," in *Scientific Computing in Electrical Engineering SCEE 2008*, ser. Mathematics in Industry, J. Roos and L. R. Costa, Eds. Berlin, Heidelberg: Springer, 2010, pp. 563–570.
 - [205] J. Roychowdhury, "Reduced-order modelling of time-varying systems," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 10, pp. 1273–1288, Oct. 1999.
 - [206] J. R. Phillips, "Automated extraction of nonlinear circuit macromodels," in *Proc. the IEEE Custom Integrated Circuits Conference*, Orlando, FL, 2000, pp. 451–454.
 - [207] P. Li and L. Pileggi, "NORM: compact model order reduction of weakly nonlinear systems," in *Proc. Design Automation Conference*, Jun. 2003, pp. 472–477.
 - [208] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *Philosophical Magazine, Series 6*, vol. 2, no. 11, pp. 559–572, 1901.
 - [209] L. Sirovich, "Turbulence and the dynamics of coherent structures, parts: I-III," *Quarterly of Applied Mathematics*, vol. 45, no. 3, pp. 561–590, Oct. 1987.
 - [210] P. Astrid, "Reduction of process simulation models: a proper orthogonal decomposition approach," Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, Netherlands, Nov. 2004.
-

- [211] P. Astrid and A. Verhoeven, "Application of least squares mpe technique in the reduced order modeling of electrical circuits," in *Proc. 17th International Symposium Mathematics Theory of Networks and Systems*, Kyoto, Japan, Jul. 2006, pp. 1980–1986.
 - [212] A. Verhoeven, M. Striebel, and E. Maten, "Model order reduction for nonlinear IC models with POD," in *Scientific Computing in Electrical Engineering SCEE 2008*, ser. Mathematics in Industry, L. R. Costa and J. Roos, Eds. Berlin/Heidelberg, Germany: Springer, 2008, pp. 571–578.
 - [213] A. Verhoeven, M. Striebel, J. Rommes, E. Maten, and T. Bechtold, "Proper orthogonal decomposition model order reduction of nonlinear IC models," in *Progress in Industrial Mathematics at ECMI 2008*, ser. Mathematics in Industry, A. D. Fitt, J. Norbury, H. Ockendon, and E. Wilson, Eds. Berlin, Germany: Springer, 2010, pp. 441–446.
 - [214] M. Striebel and J. Rommes, "Model order reduction of nonlinear systems in circuit simulation: Status and applications," in *Model Reduction for Circuit Simulation*, ser. Lecture Notes in Electrical Engineering, P. Benner, M. Hinze, and E. J. W. ter Maten, Eds. Netherlands: Springer, 2011, vol. 74, ch. 17, pp. 289–301.
 - [215] J. M. Scherpen, "Balancing for nonlinear systems," *Systems and Control Letters*, vol. 21, pp. 143–153, 1993.
 - [216] J. M. Aleida, "Balancing for nonlinear systems," Ph.D. dissertation, University of Twente, Enschede, Netherlands, 1994.
 - [217] J. Hahn and T. Edgar, "Reduction of nonlinear models using balancing of empirical gramians and galerkin projections," in *Proc. American Control Conference*, vol. 4, 2000, pp. 2864–2868.
 - [218] K. Fujimoto and J. M. A. Scherpen, "Nonlinear balanced realization based on singular value analysis of hankel operators," in *Proc. 42nd IEEE Conference on Decision and Control*, vol. 6, Maui, HI, Dec. 2003, pp. 6072–6077.
 - [219] K. Fujimoto, "On subspace balanced realization and model order reduction for nonlinear interconnected systems," in *Proc. IEEE 51st Annual Conference on Decision and Control*, Dec. 2012, pp. 4314–4319.
-

-
- [220] S. Lall, J. E. Marsden, and S. Glavaški, “Empirical model reduction of controlled nonlinear systems,” in *Proc. IFAC World Congress*, Beijing, China, Jul. 1999, pp. 473–478.
- [221] J. Hahn and T. F. Edgar, “An improved method for nonlinear model reduction using balancing of empirical gramians,” *Computers & chemical engineering*, vol. 26, no. 10, pp. 1379–1397, Oct. 2002.
- [222] M. Condon and R. Ivanov, “Empirical balanced truncation of nonlinear systems,” *Journal of Nonlinear Science*, vol. 14, no. 5, pp. 405–414, 2004.
- [223] N. Nakhla, A. Ruehli, M. Nakhla, and R. Achar, “Simulation of coupled interconnects using waveform relaxation and transverse partitioning,” *IEEE Transactions on Advanced Packaging*, vol. 29, no. 1, pp. 78–87, Feb. 2006.
- [224] N. Nakhla, A. Ruehli, M. Nakhla, R. Achar, and C. Chen, “Waveform relaxation techniques for simulation of coupled interconnects with frequency-dependent parameters,” *IEEE Transactions on Advanced Packaging*, vol. 30, no. 2, pp. 257–269, May 2007.
- [225] M. Farhan, N. Nakhla, M. Nakhla, R. Achar, and A. Ruehli, “Overlapping partitioning techniques for simulation of strongly coupled distributed interconnects,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 7, pp. 1193–1201, Jul. 2012.
- [226] A. Dounavis, X. Li, M. Nakhla, and R. Achar, “Passive closed-form transmission-line model for general-purpose circuit simulators,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2450–2459, Dec. 1999.
- [227] A. Dounavis, R. Achar, and M. Nakhla, “Efficient passive circuit models for distributed networks with frequency-dependent parameters,” *IEEE Transactions on Advanced Packaging*, vol. 23, no. 3, pp. 382–392, Aug. 2000.
- [228] A. Dounavis, E. Gad, R. Achar, and M. Nakhla, “Passive model reduction of multiport distributed interconnects,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 12, pp. 2325–2334, Dec. 2000.
- [229] A. Dounavis, R. Achar, and M. Nakhla, “A general class of passive macromodels for lossy multiconductor transmission lines,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 10, pp. 1686–1696, Oct. 2001.
-

-
- [230] H. Yu, C. Chu, Y. Shi, D. Smart, L. He, and S.-D. Tan, "Fast analysis of a large-scale inductive interconnect by block-structure-preserved macromodeling," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 10, pp. 1399–1411, Oct. 2010.
- [231] P. Feldmann, "Model order reduction techniques for linear systems with large numbers of terminals," in *Proc. Design, Automation and Test in Europe Conference and Exhibition*, vol. 2, Paris, France, Feb. 2004, pp. 944–947.
- [232] P. Feldmann and F. Liu, "Sparse and efficient reduced order modeling of linear sub-circuits with large number of terminals," in *Proc. IEEE/ACM International Conference on Computer Aided Design*, Nov. 2004, pp. 88–92.
- [233] P. Liu, S.-D. Tan, B. Yan, and B. McGaughy, "An extended SVD-based terminal and model order reduction algorithm," in *Proc. IEEE International Behavioral Modeling and Simulation Workshop*, San Jose, CA, Sep. 2006, pp. 44–49.
- [234] P. Liu, S.-D. Tan, H. Li, Z. Qi, J. Kong, B. McGaughy, and L. He, "An efficient method for terminal reduction of interconnect circuits considering delay variations," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2005, pp. 821–826.
- [235] P. Liu, S.-D. Tan, B. McGaughy, and L. Wu, "Compact reduced order modeling for multiple-port interconnects," in *Proc. 7th International Symposium on Quality Electronic Design*, San Jose, CA, Mar. 2006, pp. 413–418.
- [236] P. Liu, S.-D. Tan, B. McGaughy, L. Wu, and L. He, "TermMerg: an efficient terminal-reduction method for interconnect circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 8, pp. 1382–1392, Aug. 2007.
- [237] B. Nouri, M. S. Nakhla, and R. Achar, "A novel clustering scheme for reduced-order macromodeling of massively coupled interconnect structures," in *Proc. IEEE 19th Conference on Electrical Performance of Electronic Packaging and Systems*, Austin, TX, Oct. 2010, pp. 77–80.
- [238] Z. Zhang, X. Hu, C.-K. Cheng, and N. Wong, "A block-diagonal structured model reduction scheme for power grid networks," in *Proc. Design, Automation Test in Europe Conference and Exhibition*, Grenoble, France, Mar. 2011, pp. 1–6.
- [239] R. E. Scott, *Linear Circuits*. New York: Addison-Wesley, 1960.
-

-
- [240] R. W. Newcomb, *Linear Multiport Synthesis*. New York: McGraw-Hill, 1966.
- [241] C.-Y. Zhang, S. Luo, A. Huang, and C. Xu, "The eigenvalue distribution of block diagonally dominant matrices and block h-matrices," *Electronic Journal of Linear Algebra*, vol. 20, pp. 621–639, Sep. 2010.
- [242] K. Urahama, "Local convergence of waveform relaxation method," *Electronics and Communications in Japan, Part I*, vol. 71, no. 2, pp. 52–60, 1988.
- [243] S.-H. Xiang and Z.-Y. You, "Weak block diagonally dominant matrices, weak block H-matrix and their applications," *Linear Algebra and Its Applications*, vol. 282, pp. 263–274, Oct. 1998.
- [244] D. G. Feingold and R. S. Varga, "Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem," *Pacific Journal of Mathematics*, vol. 12, no. 4, pp. 1241–1250, 1962.
- [245] D. Saraswat, R. Achar, and M. Nakhla, "Global passivity enforcement algorithm for macromodels of interconnect subnetworks characterized by tabulated data," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 13, no. 7, pp. 819–832, Jul. 2005.
- [246] B. Gustavsen, "Passivity enforcement of rational models via modal perturbation," *IEEE Transactions on Power Delivery*, vol. 23, no. 2, pp. 768–775, Apr. 2008.
- [247] S. Grivet-Talocia and A. Ubolli, "A comparative study of passivity enforcement schemes for linear lumped macromodels," *IEEE Transactions on Advanced Packaging*, vol. 31, no. 4, pp. 673–683, Nov. 2008.
- [248] A. R. Djordjevic, M. B. Bazdar, T. K. Sarkar, and R. F. Harrington, *LINPAR for Windows Matrix Parameters for Multiconductor Transmission Lines, Software and Users' Manual, Version 2.0*. Massachusetts, MA: Artech House Publishers, 1999.
- [249] C. R. Paul, *Analysis of Multiconductor Transmission Lines*, 2nd ed. Hoboken, NJ: John Wiley, 2008.
- [250] *HSPICE*, Synopsis Inc., Mountain View, CA, 2009, ver. C-2009.03.
- [251] B. Nouri, M. S. Nakhla, and R. Achar, "A novel algorithm for optimum order estimation of reduced order macromodels," in *Proc. 15th IEEE Workshop on Signal and Power Integrity*, Naples, Italy, May 2011, pp. 33–36.
-

-
- [252] W. Liebert¹, K. Pawelzik, and H. G. Schuster, “Optimal embeddings of chaotic attractors from topological considerations,” *Europhysics Letter*, vol. 14, no. 6, pp. 521–526, Mar. 1991.
- [253] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phys. Rev. A*, vol. 45, no. 6, pp. 3403–3411, Mar. 1992.
- [254] C. Rhodes and M. Morari, “The false nearest neighbors algorithm: An overview,” *Computers and Chemical Engineering*, vol. 21, Supplement 1, pp. 1149–1154, 1997.
- [255] ———, “False-nearest-neighbors algorithm and noise-corrupted time series,” *Physical Review E*, vol. 55, no. 5, pp. 6162–6170, May 1997.
- [256] M. Rathinam and L. Petzold, “A new look at proper orthogonal decomposition,” *SIAM Journal on Numerical Analysis*, vol. 41, no. 5, pp. 1893–1925, 2003.
- [257] T. Aittokallio, M. Gyllenberg, J. Hietarinta, T. Kuusela, and T. Multamäki, “Improving the false nearest neighbors method with graphical analysis,” *Physical Review E*, vol. 60, no. 1, pp. 416–421, Jul. 1999.
- [258] K. Mehlhorn, *Multidimensional Searching and Computational Geometry*, ser. Data Structures and Algorithms 3, W. Brauer, G. Rozenberg, and A. Salomaa, Eds. Berlin, Germany: Springer, 1984.
- [259] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching,” *Journal of the ACM*, vol. 45, pp. 891–923, Nov. 1998.
- [260] T. Schreiber, “Efficient neighbor searching in nonlinear time series analysis,” *International Journal of Bifurcation and Chaos*, vol. 5, no. 2, pp. 349–358, Apr. 1995.
- [261] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge, MA: Cambridge University Press, 1997.
- [262] I. M. Carrión, E. A. Antúnez, M. M. A. Castillo, and J. J. M. Canals, “A distributed memory architecture implementation of the False Nearest Neighbors method based on distribution of dimensions,” *The Journal of Supercomputing*, vol. 59, no. 3, p. 15961618, Mar. 2012.
-

-
- [263] S.-J. Moon and A. Cangellaris, "Order estimation for time-domain vector fitting," in *Proc. IEEE 18th Conference on Electrical Performance of Electronic Packaging and Systems*, Portland, OR, Oct. 2009, pp. 69–72.
- [264] B. Nouri, M. S. Nakhla, and R. Achar, "Efficient reduced-order macromodels of massively coupled interconnect structures via clustering," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 3, no. 5, pp. 826–840, May 2013.
- [265] —, "Optimum order estimation of reduced macromodels based on a geometric approach for projection-based MOR methods," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 3, no. 7, pp. 1218–1227, Jul. 2013.
- [266] C. Desoer and F. Wu, "Trajectories of nonlinear rlc networks: A geometric approach," *IEEE Transactions on Circuit Theory*, vol. 19, no. 6, pp. 562–571, Nov. 1972.
- [267] B. Nouri, M. S. Nakhla, and R. Achar, "A novel algorithm for optimum order estimation of nonlinear reduced macromodels," in *Proc. 22nd Electrical Performance of Electronic Packaging and Systems*, San Jose, CA, Oct. 2013, pp. 137–140.
- [268] T. Matsumoto, L. Chua, and M. Komuro, "The double scroll," *IEEE Transactions on Circuits and Systems*, vol. 32, no. 8, pp. 797–818, Aug. 1985.
- [269] R. N. Madan, Ed., *Chuas's Circuit: A Paradigm for Chaos*. Singapore: World Scientific, 1993.
- [270] L. Chua, "Nonlinear circuits," *IEEE Transactions on Circuits and Systems*, vol. 31, no. 1, pp. 69–87, Jan. 1984.
- [271] T. Roska, "On the uniqueness of solutions of nonlinear dynamic networks and systems," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 3, pp. 161–169, Mar. 1978.
- [272] C. Amin, C. Kashyap, N. Menezes, K. Killpack, and E. Chiprout, "A multi-port current source model for multiple-input switching effects in CMOS library cells," in *Proc. 43rd ACM/IEEE Design Automation Conference*, San Francisco, CA, 2006, pp. 247–252.
- [273] M. Spivak, *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*, ser. Mathematics Monograph, R. Gunning and H. Rossi, Eds. New York: Addison-Wesley, 1971.
-

-
- [274] L. Conlon, *Differentiable Manifolds*, ser. Modern Birkhäuser Classics. Boston, MA: Birkhäuser, 2008.
- [275] C. Rhodes and M. Morari, “Determining the model order of nonlinear input/output systems,” *AIChE Journal*, vol. 44, no. 1, pp. 151–163, 1998.
- [276] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1997.
- [277] A. Verhoeven, T. Voss, P. Astrid, E. ter Maten, and T. Bechtold, “Model order reduction for nonlinear problems in circuit simulation,” *Proc. Applied Mathematics and Mechanics*, vol. 7, no. 1, pp. 1 021 603–1 021 604, Dec. 2007.
- [278] T. Voss, A. Verhoeven, T. Bechtold, and E. J. ter Maten, “Model order reduction for nonlinear differential algebraic equations in circuit simulation,” in *Progress in Industrial Mathematics at ECMI 2006*, L. L. Bonilla, M. Moscoso, G. Platero, and J. M. Vega, Eds. Berlin/Heidelberg, Germany: Springer, 2008, pp. 518–523.
- [279] B. Bond and L. Daniel, “Parameterized model order reduction of nonlinear dynamical systems,” in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2005, pp. 487–494.
- [280] E. Kreyszig, *Introductory Functional Analysis With Applications*. New Jersey: John Wiley & Sons Inc, 1978.
- [281] W. Rudin, *Functional Analysis*. New York: McGraw-Hill, 1991.
- [282] K. Yosida, *Functional Analysis*, 6th ed. Berlin, Germany: Springer, 1996.
- [283] J. B. Conway, *A course in functional analysis*. Berlin, Germany: Springer, 1985.
- [284] A. Megretski and J. Wyatt, “Linear algebra and functional analysis for signals and systems,” Massachusetts Institute of Technology, Cambridge, MA, Lecture Notes, Sep. 2009.
- [285] R. Melrose, “Introduction to functional analysis,” Massachusetts Institute of Technology (MIT-Open Course Ware, Cambridge, MA, Course Notes 18.102, 2009.
- [286] M. P. Sanchez. (2013, Apr.) Surjective (version-4). [Online]. Available: <http://planetmath.org/Surjective>
-

Appendix A

Properties of Nonlinear Systems in Compare to Linear

To contrast the behavior of generic linear and nonlinear systems, examples of essentially nonlinear phenomena [17, 55] are highlighted below:

- ***Multiple isolated equilibria.*** A linear system can have only one isolated equilibrium point; thus, it can have only one steady-state operating point that attracts the state of the system irrespective of the initial state. A nonlinear system can have more than one isolated equilibrium point. The state may converge to one of several steady-state operating points, depending on the initial state of the system.
- ***Finite escape time.*** The state of an unstable linear system goes to infinity as time approaches infinity; a nonlinear system's state, however, can go to infinity in finite time.
- ***Limit cycles.*** For a linear time-invariant system to oscillate, it must have a pair of eigenvalues on the imaginary axis, which is a non-robust condition that is almost impossible to maintain in the presence of perturbations. Even if we do, the amplitude of oscillation will be dependent on the initial state. In real life, stable oscillation must

be produced by nonlinear systems. There are nonlinear systems that can go into an oscillation of fixed amplitude and frequency, irrespective of the initial state. This type of oscillation is known as a limit cycle.

- ***Subharmonic, harmonic, or almost-periodic oscillations.*** A stable linear system under a periodic input produces an output of the same frequency. A nonlinear system under periodic excitation can oscillate with frequencies that are submultiples or multiples of the input frequency. It may even generate an almost-periodic oscillation, an example is the sum of periodic oscillations with frequencies that are not multiples of each other.
 - ***Chaos.*** A nonlinear system can have a more complicated steady-state behavior that is not equilibrium, periodic oscillation, or almost-periodic oscillation. Such behavior is usually referred to as chaos. Some of these chaotic motions exhibit randomness, despite the deterministic nature of the system.
 - ***Multiple modes of behavior.*** It is not unusual for two or more modes of behavior to be exhibited by the same nonlinear system. An unforced system may have more than one limit cycle. A forced system with periodic excitation may exhibit harmonic, subharmonic, or more complicated steady-state behavior, depending upon the amplitude and frequency of the input. It may even exhibit a discontinuous jump in the mode of behavior as the amplitude or frequency of the excitation is smoothly changed.
-

Appendix B

Model Order Reduction Related Concepts

This appendix presents the concepts and techniques that are needed to study the theory of dynamical systems. It is included to enhance the thoroughness of this thesis in its focus on the subject of MOR. For more details, interested reader can refer to the given references.

B.1 Tools From Linear Algebra and Functional Analysis

An overview on the mathematical concepts and definitions from “linear algebra” and “functional analysis” [280–285], relevant to the work in this report, is presented.

B.1.1 Review of Vector Space and Normed Space

B.1.1.1 Real and Complex Vector (Linear) Space

Definition B.1 (*Vector Space*). Consider a nonempty set V of elements of the same type v_i . Elements v_i may be vectors in an n -dimensional space, or sequences of numbers, or functions [280], however it is generally called *vectors*. The set $V = \{v_1, v_2, v_3, \dots\}$. V is called a *vector space* or *linear space* if with respect to two algebraic operations, "addition"

and "scalar multiplication" the followings properties are satisfied:

(i) **V is a closed set under addition:** $\forall \mathbf{v}_i, \mathbf{v}_j \in \mathbf{V}$, there is $(\mathbf{v}_i + \mathbf{v}_j) \in \mathbf{V}$.

Informally it is said, every two vectors (elements) in the set can be added (one to another) to produce another vector (or element of the same type) in the set.

(ii) **V is a closed set under scalar multiplication** (scalars are real or complex numbers): $\forall \mathbf{v}_i \in \mathbf{V}$ and each scalar $\lambda \in \mathbb{F}$, there is $\lambda \mathbf{v}_i \in \mathbf{V}$.

Informally it is said, each vector (element) in the set can be scaled with a (real/complex) number and results is an element of the same type, that is in the set.

(iii) **Associative:** $(\mathbf{v}_i + \mathbf{v}_j) + \mathbf{v}_k = \mathbf{v}_i + (\mathbf{v}_j + \mathbf{v}_k)$, for all $\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k \in \mathbf{V}$.

(iv) **Cumulative:** $\mathbf{v}_i + \mathbf{v}_j = \mathbf{v}_j + \mathbf{v}_i$, $\forall \mathbf{v}_i, \mathbf{v}_j \in \mathbf{V}$.

(v) **Distributivity:**

(a) $\lambda (\mathbf{v}_i + \mathbf{v}_j) = (\lambda \mathbf{v}_i) + (\lambda \mathbf{v}_j)$, $\forall \mathbf{v}_i, \mathbf{v}_j \in \mathbf{V}$ and $\lambda \in \mathbb{F}$,

(b) $(\lambda + \mu) \mathbf{v}_i = (\lambda \mathbf{v}_i) + (\mu \mathbf{v}_i)$, $\forall \mathbf{v}_i \in \mathbf{V}$ and $\lambda, \mu \in \mathbb{F}$,

(c) $\lambda (\mu \mathbf{v}_i) = (\lambda \mu) \mathbf{v}_i$, $\forall \mathbf{v}_i \in \mathbf{V}$ and $\lambda, \mu \in \mathbb{F}$.

(vi) **zero vector:** $\exists! \mathbf{0} \in \mathbf{V} \mid \mathbf{v}_i + \mathbf{0} = \mathbf{v}_i$, $\forall \mathbf{v}_i \in \mathbf{V}$

(vii) **additive inverse:** $\forall \mathbf{v}_i \in \mathbf{V} \exists (-\mathbf{v}_i) \in \mathbf{V} \mid \mathbf{v}_i + (-\mathbf{v}_i) = \mathbf{0}$

(viii) **Multiplicative identity:** $\exists! \mathbf{0} \in \mathbf{V} \mid \mathbf{v}_i + \mathbf{0} = \mathbf{v}_i$, $\forall \mathbf{v}_i \in \mathbf{V}$

- **V is called a *real vector space***, if the scalars come from the field of real numbers ($\lambda, \mu \in \mathbb{R}$).
- **V is called a *complex vector space***, if the scalars come from the field of complex numbers ($\lambda, \mu \in \mathbb{C}$).

A nonempty subset of **V** that is a linear space too is called a (linear) subspace of **V**.

B.1.1.2 Normed and Metric Spaces

Definition B.2 (Normed Space). Let V be a real or complex linear space. A real-valued function $\|v\|$, defined for $v \in V$, is called a **norm** if it has the properties [9, 284]:

(i) **definiteness:**

$$(a) \quad \|v\| \geq 0,$$

$$(b) \quad \|v\| = 0 \Leftrightarrow v = 0,$$

(ii) **homogeneity:** $\|\lambda v\| = |\lambda| \|v\|$,

(iii) **triangle inequality:** $\|v_i + v_j\| \leq \|v_i\| + \|v_j\|$.

The space V is said to be "*normed*" by $\|\cdot\|$ [9, 284], and shortly referred to as "*normed space*"

Two simple consequences of the triangle inequality are

$$\|v_i + \dots + v_j\| \leq \|v_i\| + \dots + \|v_j\|, \quad (B.1)$$

$$0 \leq \left| \|v_i\| - \|v_j\| \right| \leq \|v_i - v_j\|. \quad (B.2)$$

from (B.2), it can explicitly be concluded that,

$$\|v_i\| - \|v_j\| \leq \|v_i - v_j\|. \quad (B.3)$$

Definition B.3 (Metric Space). A set V together with a real-valued function $d : V \times V \rightarrow \mathbb{R}$ as $d(v_i - v_j) = \|v_i - v_j\|$ is called a *metric space* and the function d a *metric* or *distance* function, if the following holds:

For all v_i, v_j , and $v_k \in \mathbb{R}$,

(i) **positivity:** $d(v_i, v_j) \geq 0$. for $v_i \neq v_j$, and $d = (v_i, v_i)$

(ii) **symmetry**: $d(\mathbf{v}_i, \mathbf{v}_j) = d(\mathbf{v}_j, \mathbf{v}_i),$

(iii) **triangle inequality**: $d(\mathbf{v}_i, \mathbf{v}_j) \leq d(\mathbf{v}_i, \mathbf{v}_k) + d(\mathbf{v}_k, \mathbf{v}_j).$

It is to be noted that, a "norm" defines a "distance function" (or metric), thus a normed space is a metric space.

Using this "canonical" distance function and proceeding in a natural manner, we can extend the definition of familiar mathematical objects from the Euclidean space $\mathbb{R}^n$ to any normed space $\mathbf{L}$. The concepts such as balls, ε -neighborhoods, neighborhoods, interior points, boundary points, open and closed sets, etc. [9].

B.1.2 Review of the Different Norms

B.1.2.1 Norms for Real Vector Space

An n -dimensional real space $\mathbb{R}^n$ which is the set of all n -tuples of real numbers $\mathbf{a} = (a_1, \dots, a_n) = (a_i)$ where $a_i \in \mathbb{R}$ can be normed in many ways, *e.g.*, by any one of the following:

$$\|\mathbf{a}\|_2 = \sqrt{a_1^2 + \dots + a_n^2} \quad \text{2-norm,} \quad (\text{B.4})$$

$$\|\mathbf{a}\|_1 = |a_1| + \dots + |a_n| \quad \text{unity-norm,} \quad (\text{B.5})$$

$$\|\mathbf{a}\|_\infty = \max_i |a_i| \quad \text{maximum-norm.} \quad (\text{B.6})$$

B.1.2.2 Norms for Complex Vector Space

The n -dimensional complex space $\mathbb{C}^n$ is normed in the same manner as in previous item, except in the definition of the the norms as shown in (B.4), it is necessary to use absolute value as the magnitude of the complex number $|a| = |\alpha + j\beta| = +\sqrt{\alpha^2 + \beta^2}$. For example, similar to 2-norm in (B.4), Euclidean-norm for the complex vector space is defined

as

$$\|\mathbf{a}\|_e = \sqrt{|a_1|^2 + \cdots + |a_n|^2}, \quad \text{Euclidean norm.} \quad (\text{B.7})$$

It is easy to see that, for the case of real vector spaces Euclidean-norm and 2-norm are equivalent.

B.1.2.3 Norms for Vector Space with Function Elements

It was mentioned that, the elements of a "vector space" may be functions [280]. Hence, the norm operator for spaces of function should be defined.

Let $\mathbf{V} = \{\mathbf{x}(t)\}$ be a set of continuous function defined on the closed interval $\mathbf{I}$ ($t \in \mathbf{I}$). The space $\mathbf{V}$ can be normed in the following ways:

$$\|\mathbf{x}\|_{\mathbf{I},p} = \left(\int_{t \in \mathbf{I}} |\mathbf{f}(t)|^p dt \right)^{\frac{1}{p}} \quad p\text{-norm,} \quad (\text{B.8})$$

$$\|\mathbf{x}\|_{\mathbf{I},\infty} = \sup_{t \in \mathbf{I}} |\mathbf{x}(t)| \quad \text{maximum norm} \quad (\text{B.9})$$

B.2 Mappings Concepts

The concept of "mapping" is fundamentally important in many areas of mathematics, such as functional analysis and differential geometry [280].

Definition B.4 (Mapping [9,280,281]). Let $\mathbf{X}$, $\mathbf{Y}$ be two sets of points in an n -dimensional space and $\mathbf{A} \subseteq \mathbf{X}$ and $\mathbf{B} \subseteq \mathbf{Y}$ be two subsets of them, respectively.

$$\mathbf{T} : \mathbf{X} \rightarrow \mathbf{Y}$$

is a *mapping* (or *transformation*) $\mathbf{T}$ from $\mathbf{A}$ into $\mathbf{B}$, that associates with each $x \in \mathbf{A}$ a single $y \in \mathbf{B}$ as illustrated in Fig. B.1.

If $A \subset X$ and $B \subset Y$, $T(A)$ the *image* of A and $T^{-1}(B)$ the *inverse image* or *preimage* of B are defined by

$$T(A) = \{y = T(x) \mid x \in A\}, \quad T^{-1}(B) = \{x \mid T(x) \in B\}.$$

The set $A := \mathcal{D}_T$ is called the “*domain of definition*” of T or, more briefly *domain* of T . The set $B := \mathcal{R}_T$ is also called the *range* of T . The transformation $T(x)$ may be shorten as Tx .

Figure B.1: Visualization of a mapping

Definition B.5 (Injective or “one-to-one” Mapping [280]). A mapping T is *injective*, or *one-to-one* if every element of the range $\mathcal{R}_T$ is mapped to by at most one element of the domain $\mathcal{D}_T$.

Notationally, $\forall x_1, x_2 \in \mathcal{D}_T \mid x_1 \neq x_2$, implies $Tx_1 \neq Tx_2$. As illustrated in Fig. B.2, it is said that, different points in $\mathcal{D}_T$ have different images, so that the inverse image of any point in $\mathcal{R}_T$ is a single point in $\mathcal{R}(T)$. More mathematically, T is an **injective** mapping (iff):

$$\forall x_1, x_2 \in \mathcal{D}_T, \quad Tx_1 = Tx_2 \iff x_1 = x_2,$$

or equivalently, $\forall x_1, x_2 \in \mathcal{D}_T \quad x_1 \neq x_2 \iff Tx_1 \neq Tx_2$.

Figure B.2: Visualization of an injective mapping

Definition B.6 (Surjective or “onto” Mapping [280, 286]). A mapping T is called *surjective* or a mapping of $\mathcal{D}_T$ **onto** $\mathbf{Y}$ if $\mathcal{R}_T = \mathbf{Y}$. This states that, for every $y \in \mathbf{Y}$, there is at least one $x \in \mathcal{D}_T$ such that $y = Tx$. This is illustrated in Fig B.3.

For example: $T: \mathcal{D}_T \longrightarrow \mathcal{R}_T$ is always surjective.

Figure B.3: Visualization of a surjective mapping

Definition B.7 (Bijective Mapping [280]). T is *bijective* if T is both injective (one-to-one) and surjective (onto).

It is said that there is a one-to-one correspondence between elements in domain and elements in range. If every element of the range is mapped to exactly one element of the

domain as shown in Fig. B.4.

Notationally, $\forall y \in \mathbf{Y} \exists! x \in \mathcal{D}_T \mid y = Tx$ and $\forall x \in \mathcal{D}_T \exists Tx \in \mathbf{Y}$

Figure B.4: Inverse mapping $T^{-1}: \mathbf{Y} \longrightarrow \mathcal{D}(T) \subseteq \mathbf{X}$ of a bijective mapping T

Definition B.8 (Linear Mappings [9, 281]). An operator $T: \mathbf{D} \rightarrow \mathbf{F}$ is called *linear* if $\mathbf{D}$ is a "linear subspace" of $\mathbf{E}$ and $T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$ holds for $x, y \in \mathbf{D}$ and $\lambda, \mu \in \mathbb{R}$ or $\mathbb{C}$.

Definition B.9 (Continuous Operator [9]). An operator $T: \mathbf{D} \rightarrow \mathbf{F}$ is said to be continuous at a point $x_0 \in \mathbf{D}$ if $x_n \in \mathbf{D}$, $x_n \longrightarrow x_0$ implies that $Tx_n \longrightarrow Tx_0$.

Remark B.1. For every $\varepsilon > 0$, there exists $\delta > 0$ such that from $x \in \mathbf{D}$, $\|x - x_0\| < \delta$, it follows that $\|Tx - Tx_0\| < \varepsilon$ [9].

Remark B.2 (Lipschitz Condition for Operators [9]). An operator T satisfies a Lipschitz condition in $\mathbf{D}$ (with Lipschitz constant q) if

$$\|Tx - Ty\| \leq q \|x - y\|, \quad \text{for } x, y \in \mathbf{D} \quad (\text{B.10})$$

It is easy to check that such an operator is continuous in $\mathbf{D}$.

Example B.1. Consider $Tf = \int_0^1 f(x) dx$, where $f(x)$ is a continuous function in $\mathbf{D} = \mathcal{C}([0, 1])$, and $\mathbf{F} = \mathbb{R}$. This operator $T : \mathbf{D} \rightarrow \mathbb{R}$ is continuous.

Solution:

for $f_1 = f(x_1)$ and $f_2 = f(x_2) \in \mathbf{D}$, we have

$$\|Tf_1 - Tf_2\| = \underbrace{\left\| \int_0^1 f_1 dx - \int_0^1 f_2 dx \right\|}_{\text{norm on } \mathbf{F}} \leq \int_0^1 \overbrace{\|f_1 - f_2\|}^{\text{norm on } \mathbf{D}} dx = \|f_1 - f_2\|,$$

Lipschitz constant can be picked (e.g.) as $q = 1$, hence, satisfying Lipschitz condition T is continuous in $\mathbf{D}$. ■

Definition B.10 (Contractive Mapping). Given a mapping $T : \mathbf{D} \rightarrow \mathbf{F}$ as

$$T(\mathbf{x}) = \mathbf{x}, \quad (\text{B.11})$$

where $\mathbf{F}$ is a suitably chosen Banach space and $\mathbf{D} \subset \mathbf{F}$. This mapping is called *contractive* (or a *contraction*) if it satisfies Lipschitz condition (cf. Remark-B.2) on $\mathbf{D}$ with Lipschitz constant $q < 1$, i.e.:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbf{D} \exists q < 1 : \|T\mathbf{x} - T\mathbf{y}\| \leq q \|\mathbf{x} - \mathbf{y}\|. \quad (\text{B.12})$$

From (B.12) one may say that the distance between the image points $T\mathbf{x}$ and $T\mathbf{y}$, under the mapping is smaller by a factor q than the distance between the two original points $\mathbf{x}, \mathbf{y}$ and hence T “contracts” distances between points.

Example B.2. Consider mapping $T : [0, 1] \rightarrow [0, 1]$, with the mapping operator $Tx = 1 - \frac{x}{2}$.

Solution:

For any $x, y \in [0, 1]$, we have $|Tx - Ty| = \left| \frac{x}{2} - \frac{y}{2} \right| = \frac{1}{2} |x - y|$, so, $q = \frac{1}{2}$, having the Lipschitz constant $q = \frac{1}{2} < 1$ then it is Lipschitz and contractive. ■

The solution of (B.11) is called *fixed point* of mapping T . It is a point which remains “fixed” under the map T . Fixed points can be found using an iteration procedure called the method of *successive approximation*. To find $\bar{x}$ satisfying $\bar{x} = T(\bar{x})$, it starts from an element $x_0 \in D$ and successively forms

$$x_1 = T(x_0), \quad x_2 = T(x_1), \quad \dots, x_{n+1} = T(x_n), \quad \dots \quad (\text{B.13})$$

The convergence of the sequence in (B.13) (to the fixed point $\bar{x}$), is intimately connected to the contracting of mapping.

Example B.3. Does the mapping $T(x) = 1 - \frac{x}{2}$ on $[0, 1]$ (in example-B.2) converge?

Solution:

Let $x_0 = \frac{1}{4} \rightarrow x_1 = \frac{7}{8} \rightarrow x_2 = \frac{9}{16} \rightarrow \dots$, i.e. $\lim_{x \rightarrow \infty} x_n = \frac{2}{3} = \bar{X}$ (fixed point). ■

Theorem B.1 (Fixed Point Theorem). Let D be a nonempty, closed set in a Banach space F . Let the operator $T : D \rightarrow F$ map D into itself, $T(D) \subset D$. Given $x_0 \in D$ the sequence $x_n = T^n x_0$ converges to a unique value $\bar{x}$ in D for the fixed point of mapping such that $T\bar{x} = \bar{x}$.

Proof. For the proof [9, pp.59-60] can be referred to. ■

Appendix C

Proof of Theorem-5.1 in Section 5.4

The subsystems in the proposed methodology (5.12) share the same $\mathbf{G}$ and $\mathbf{C}$ matrices with the original system, i.e. $\mathbf{G}_i = \mathbf{G}$ and $\mathbf{C}_i = \mathbf{C}$, (for $i = 1, \dots, K$). However, for the sake of the clarity when following the proof; hereafter, matrices for every i -th subsystem will be signified with its index i .

For the system in (5.12), let $s_0 \in \mathbb{C}$ be a properly selected expansion point such that the matrix pencil $(\mathbf{G}_i + s_0 \mathbf{C}_i)$ is nonsingular. The corresponding input-to-state transfer function is obtained by applying Laplace transformation on (5.12a) as

$$\underline{\mathbf{H}}_i(s) = (\mathbf{G}_i + s \mathbf{C}_i)^{-1} \mathbf{B}_i. \quad (\text{C.1})$$

From (C.1) and following the similar steps shown in (5.4)-(5.6), the following matrices are defined

$$\mathbf{A}_i \triangleq (\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{C}_i \quad (\text{C.2})$$

and

$$\mathbf{R}_i \triangleq (\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{B}_i. \quad (\text{C.3})$$

Consider the complex-valued matrix function in (C.1) be a smooth (continuously derivable)

with invertible $(\mathbf{G}_i + s_0 \mathbf{C}_i)$. Its Taylor series approximation $\underline{\mathbf{H}}_i(s)$, in the proximity of s_0 , can be represented by the expansion

$$\underline{\mathbf{H}}_i(s) = \sum_{j=0}^{\infty} (-1)^j \underline{\mathbf{M}}_{i,j}(s_0) (s - s_0)^j \quad (\text{C.4})$$

where the j -th moment of the function at s_0 is

$$\underline{\mathbf{M}}_{i,j}(s_0) = \mathbf{A}_i^j \mathbf{R}_i, \quad (\text{for all } j). \quad (\text{C.5})$$

Similarly, expanding the approximant transfer function (5.26) about s_0 gives

$$\hat{\underline{\mathbf{H}}}_i(s) = \sum_{j=0}^{\infty} (-1)^j \hat{\underline{\mathbf{M}}}_{i,j}(s_0) (s - s_0)^j, \quad (\text{C.6})$$

where

$$\hat{\underline{\mathbf{M}}}_{i,j}(s_0) = \mathbf{Q}_i \hat{\mathbf{A}}_i^j \hat{\mathbf{R}}_i, \quad (\text{for all } j) \quad (\text{C.7})$$

and matrices are given by

$$\hat{\mathbf{A}}_i \triangleq \left(\hat{\mathbf{G}}_i + s_0 \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{C}}_i \quad (\text{C.8})$$

and

$$\hat{\mathbf{R}}_i \triangleq \left(\hat{\mathbf{G}}_i + s_0 \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{B}}_i. \quad (\text{C.9})$$

Next, we show that the first $\mathcal{M}_i = \lfloor q_i/m_i \rfloor$ coefficient matrices (block moments) in the expansions (C.4) and (C.6) are identical. To serve this purpose, we first define the following proposition C.1 and lemma C.1, which are used later in the proof of theorem 5.1.

Proposition C.1. *Let $\mathbf{Q}_i$ be a $N \times q_i$ full column-rank projection matrix, whose columns*

span the q_i dimensional Krylov subspace, as

$$\begin{aligned} \text{colspan} \{ \mathbf{Q}_i \} &= \mathcal{K}_{\mathcal{M}_i} (\mathbf{A}_i, \mathbf{R}_i) \\ &= \text{span} \{ \mathbf{R}_i, \mathbf{A}_i \mathbf{R}_i, \dots, \mathbf{A}_i^{\mathcal{M}_i-1} \mathbf{R}_i \}, \end{aligned} \quad (\text{C.10})$$

where $\mathbf{R}_i$ has m_i columns. There exists a $N \times m_i$ matrix $\mathbf{E}_{i,j}$, such that

$$\mathbf{A}_i^j \mathbf{R}_i = \mathbf{Q}_i \mathbf{E}_{i,j}, \quad 0 \leq j < \mathcal{M}_i. \quad (\text{C.11})$$

Proof. From the definition of Krylov subspace (C.10) it is deduced that, associated with every (j -th) moment, there exists a projection of it in the Krylov subspace. Let a $N \times m_i$ matrix $\mathbf{E}_{i,j}$ be the projection of the block moment $\underline{\mathbf{M}}_{i,j}$ into Krylov subspace, induced by $\mathbf{A}_i$ and $\mathbf{R}_i$. We thus have

$$\underline{\mathbf{M}}_{i,j} = \mathbf{Q}_i \mathbf{E}_{i,j}. \quad (\text{C.12})$$

Substituting (C.5) in (C.12) proves the relationship in (C.11). ■

Lemma C.1. *Given the projection matrix $\mathbf{Q}_i$ as specified in proposition C.1, a $N \times N$ matrix as*

$$\mathbf{F}_i \triangleq \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{C}_i \quad (\text{C.13})$$

satisfies the relation

$$\mathbf{F}_i^j \mathbf{R}_i = \mathbf{A}_i^j \mathbf{R}_i, \quad 0 \leq j < \mathcal{M}_i. \quad (\text{C.14})$$

Proof. The proof is possible by induction on j . First, it is trivial to prove (C.14) for $j = 0$. Next, assume that (C.14) is true for any $j - 1$ when $0 < j < \mathcal{M}_i$, that is

$$\mathbf{F}_i^{j-1} \mathbf{R}_i = \mathbf{A}_i^{j-1} \mathbf{R}_i. \quad (\text{C.15})$$

Multiplying both sides of (C.15) by $\mathbf{A}_i$ in (C.2) yields

$$(\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{C}_i \times \mathbf{F}_i^{j-1} \mathbf{R}_i = \mathbf{A}_i^j \mathbf{R}_i. \quad (\text{C.16})$$

Substituting (C.11) from proposition C.1 in (C.15), we have

$$(\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{C}_i \times \mathbf{F}_i^{j-1} \mathbf{R}_i = \mathbf{Q}_i \mathbf{E}_{i,j}. \quad (\text{C.17})$$

Multiplying both sides of (C.17) with a $N \times N$ matrix $\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i)$, where superscript $*$ denotes the conjugate (Hermitian) transpose, we obtain

$$\begin{aligned} & \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \times (\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{C}_i \times \mathbf{F}_i^{j-1} \mathbf{R}_i = \\ & \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \mathbf{Q}_i \mathbf{E}_{i,j}, \end{aligned} \quad (\text{C.18})$$

then,

$$\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{C}_i \times \mathbf{F}_i^{j-1} \mathbf{R}_i = \mathbf{Q}_i \mathbf{E}_{i,j}. \quad (\text{C.19})$$

Using (C.11) and (C.13), (C.19) leads to $\mathbf{F}_i^j \mathbf{R}_i = \mathbf{A}_i^j \mathbf{R}_i$ which is the desired relation in (C.14). ■

Following the proposition C.1 and lemma C.1 established above, theorem 5.1 is proved.

Proof of theorem 5.1. In this proof we establish that

$$\underline{\mathbf{M}}_{i,j}(s_0) = \hat{\underline{\mathbf{M}}}_{i,j}(s_0), \quad 0 \leq j < \mathcal{M}_i. \quad (\text{C.20})$$

I) For $j = 0$: To show that the first two block moments are equal, for $j = 0$, from

proposition C.1, we recall that

$$\mathbf{R}_i = \mathbf{Q}_i \mathbf{E}_{i,0}. \quad (\text{C.21})$$

Multiplying both sides of (C.21) by the matrix $\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i)$, used in the proof of lemma C.1, we have

$$\begin{aligned} & \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \times \mathbf{R}_i \\ &= \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \mathbf{Q}_i \mathbf{E}_{i,0}. \end{aligned} \quad (\text{C.22})$$

Substituting $\mathbf{R}_i$ from (C.3),

$$\begin{aligned} & \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \times (\mathbf{G}_i + s_0 \mathbf{C}_i)^{-1} \mathbf{B}_i = \\ & \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* (\mathbf{G}_i + s_0 \mathbf{C}_i) \mathbf{Q}_i \mathbf{E}_{i,0}, \end{aligned} \quad (\text{C.23})$$

it will be

$$\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{B}_i = \mathbf{Q}_i \mathbf{E}_{i,0}. \quad (\text{C.24})$$

Next, substituting (C.21) in (C.24), we get

$$\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{B}_i = \mathbf{R}_i. \quad (\text{C.25})$$

Also, combining (5.14), (C.9) and (C.25), we get

$$\mathbf{Q}_i \hat{\mathbf{R}}_i = \mathbf{R}_i. \quad (\text{C.26})$$

According to the definition of the moments for the original and reduced system in (C.5) and (C.7), respectively, (C.26) can be rewritten as $\hat{\underline{\mathbf{M}}}_{i,0} = \underline{\mathbf{M}}_{i,0}$, that establishes the matching of the first block moments.

II) For $0 < j < \mathcal{M}_i$: From (C.7), (C.8) and (C.9), we have

$$\hat{\mathbf{M}}_{i,j} = \mathbf{Q}_i \times \left(\left(\hat{\mathbf{G}}_i + s_0 \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{C}}_i \right)^j \times \left(\hat{\mathbf{G}}_i + s_0 \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{B}}_i. \quad (\text{C.27})$$

Using the definitions of the reduced matrices in (5.14), (C.27) can be rewritten as

$$\begin{aligned} \hat{\mathbf{M}}_{i,j} = \mathbf{Q}_i \left((\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i \right)^j \times \\ (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{B}_i, \end{aligned} \quad (\text{C.28})$$

which equivalently is

$$\begin{aligned} \hat{\mathbf{M}}_{i,j} = \left(\mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{C}_i \right)^j \times \\ \mathbf{Q}_i (\mathbf{Q}_i^* \mathbf{G}_i \mathbf{Q}_i + s_0 \mathbf{Q}_i^* \mathbf{C}_i \mathbf{Q}_i)^{-1} \mathbf{Q}_i^* \mathbf{B}_i. \end{aligned} \quad (\text{C.29})$$

Next, substituting (C.13) and (C.25) in (C.29), we get

$$\underline{\hat{\mathbf{M}}}_{i,j} = \mathbf{F}_i^j \mathbf{R}_i. \quad (\text{C.30})$$

The result from substituting (C.14) from lemma C.1 in (C.30) and using (C.5) establishes (C.20). ■

Appendix D

Proof of Theorem-5.2 in Section 5.4

Based on the superposition paradigm, the output(s) of the original system can be constructed by superposing the responses from the subsystems

$$\mathbf{I}(s) = \sum_{i=1}^K \mathbf{I}_i(s). \quad (\text{D.1})$$

From (D.1) and definition of transfer function, we get

$$\mathbf{H}(s)\mathbf{U}(s) = \mathbf{H}_1(s)\mathbf{U}_1(s) + \mathbf{H}_2(s)\mathbf{U}_2(s) + \cdots + \mathbf{H}_K(s)\mathbf{U}_K(s), \quad (\text{D.2})$$

which equivalently can be written as

$$\mathbf{H}(s)\mathbf{U}(s) = [\mathbf{H}_1(s), \mathbf{H}_2(s), \cdots, \mathbf{H}_K(s)] \begin{bmatrix} \mathbf{U}_1(s) \\ \mathbf{U}_2(s) \\ \vdots \\ \mathbf{U}_K(s) \end{bmatrix}$$

$$= [\mathbf{H}_1(s), \mathbf{H}_2(s), \dots, \mathbf{H}_K(s)] \mathbf{U}(s). \quad (\text{D.3})$$

Applying Laplace transformation to (5.12), the (input-to-output) transfer function matrix for any i -th submodel is

$$\mathbf{H}_i(s) = \mathbf{L} (\mathbf{G}_i + s\mathbf{C}_i)^{-1} \mathbf{B}_i. \quad (\text{D.4})$$

Considering the relation in (C.1), from (D.4) we have

$$\mathbf{H}_i(s) = \mathbf{L} \underline{\mathbf{H}}_i(s). \quad (\text{D.5})$$

From (D.3) and (D.5), it is

$$\begin{aligned} \mathbf{H}(s) &= [\mathbf{H}_1(s), \mathbf{H}_2(s), \dots, \mathbf{H}_K(s)] \\ &= \mathbf{L} [\underline{\mathbf{H}}_1(s), \underline{\mathbf{H}}_2(s), \dots, \underline{\mathbf{H}}_K(s)]. \end{aligned} \quad (\text{D.6})$$

By expanding the matrix transfer function $\mathbf{H}(s)$ in (D.6) in the form of Taylor series in proximity of the complex frequency s_0 and considering (5.7), we get

$$\begin{aligned} \mathbf{H}(s) &= \mathbf{L} \sum_j (-1)^j \underline{\mathbf{M}}_j(s_0) (s - s_0)^j = \\ &\quad \sum_j (-1)^j \mathbf{L} [\underline{\mathbf{M}}_{1,j}(s_0), \dots, \underline{\mathbf{M}}_{K,j}(s_0)] (s - s_0)^j \end{aligned} \quad (\text{D.7})$$

By equating the moments in the corresponding terms of the Taylor series in the both sides of (D.7), the original moments in (5.8) are now obtained based on the corresponding moments of the submodels, as

$$\mathbf{M}_j(s_0) = \mathbf{L} [\underline{\mathbf{M}}_{1,j}(s_0), \dots, \underline{\mathbf{M}}_{K,j}(s_0)] \quad (\text{D.8})$$

Substituting (C.20) in (D.8), we get

$$\mathbf{M}_j(s_0) = \mathbf{L} \left[\hat{\underline{\mathbf{M}}}_{1,j}(s_0), \dots, \hat{\underline{\mathbf{M}}}_{K,j}(s_0) \right], \quad 0 \leq j < \mathcal{M}_i. \quad (\text{D.9})$$

Next, the (input-to-output) transfer function for the reduced submodels in (5.13) is obtained as

$$\hat{\mathbf{H}}_i(s) = \hat{\mathbf{L}}_i \left(\hat{\mathbf{G}}_i + s \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{B}}_i \quad (\text{D.10})$$

and considering the definition of $\hat{\mathbf{L}}_i$ in (5.14), we have

$$\hat{\mathbf{H}}_i(s) = \mathbf{L} \mathbf{Q}_i \left(\hat{\mathbf{G}}_i + s \hat{\mathbf{C}}_i \right)^{-1} \hat{\mathbf{B}}_i. \quad (\text{D.11})$$

By comparing (5.26) and (D.11), it is

$$\hat{\mathbf{H}}_i(s) = \mathbf{L} \hat{\underline{\mathbf{H}}}_i(s). \quad (\text{D.12})$$

Then, starting from

$$\hat{\mathbf{I}}(s) = \sum_{i=1}^K \hat{\mathbf{I}}_i(s). \quad (\text{D.13})$$

and following the similar steps in (D.2)-(D.6), we get

$$\hat{\mathbf{H}}(s) = \mathbf{L} \left[\hat{\underline{\mathbf{H}}}_1(s), \dots, \hat{\underline{\mathbf{H}}}_K(s) \right] \quad (\text{D.14})$$

The Taylor expansion in proximity of s_0 for $\hat{\mathbf{H}}(s)$ in (5.26) and (D.14) leads to

$$\begin{aligned} \hat{\mathbf{H}}(s) &= \sum_j (-1)^j \mathbf{L} \hat{\underline{\mathbf{M}}}_j(s_0) (s - s_0)^j = \\ &\quad \sum_j (-1)^j \mathbf{L} \left[\hat{\underline{\mathbf{M}}}_{1,j}(s_0), \dots, \hat{\underline{\mathbf{M}}}_{K,j}(s_0) \right] (s - s_0)^j \end{aligned} \quad (\text{D.15})$$

As explained before, by equating the corresponding moment matrices in the both sides of (D.15), we have

$$\hat{\mathbf{M}}_j(s_0) = \mathbf{L} \left[\hat{\underline{\mathbf{M}}}_{1,j}(s_0), \dots, \hat{\underline{\mathbf{M}}}_{K,j}(s_0) \right] \quad (\text{D.16})$$

Comparing (D.9) and (D.16) indicates that, the corresponding entries in the moments matrices $\mathbf{M}_j(s_0)$ in (D.9) and $\hat{\mathbf{M}}_j(s_0)$ in (D.16) are equal for $0 \leq j < \mathcal{M}$, which explicitly proves that these two moment matrices are identical up to a certain order $\mathcal{M}$.

$$\mathbf{M}_j(s_0) = \hat{\underline{\mathbf{M}}}_j(s_0), \quad \text{for } 0 \leq j < \mathcal{M} \quad (\text{D.17})$$

The order of matching $\mathcal{M}$ is decided by the lowest count of the moments $\mathcal{M}_i$, matched between the subsystems and their associated reduced models as $\mathcal{M} = \min_{i=1,\dots,K} (\mathcal{M}_i)$. ■