Dynamic Model Reduction: 
An Overview of Available Techniques 
with Application to Power Systems 

Savo D. Đukić¹, Andrija T. Sarić²

Abstract: This paper summarises the model reduction techniques used for the 
reduction of large-scale linear and nonlinear dynamic models, described by the 
differential and algebraic equations that are commonly used in control theory. 
The groups of methods discussed in this paper for reduction of the linear 
dynamic model are based on singular perturbation analysis, modal analysis, 
singular value decomposition, moment matching and methods based on a 
combination of singular value decomposition and moment matching. Among the 
nonlinear dynamic model reduction methods, proper orthogonal decomposition, 
the trajectory piecewise linear method, balancing-based methods, reduction by 
optimising system matrices and projection from a linearised model, are 
described. Part of the paper is devoted to the techniques commonly used for 
reduction (equivalencing) of large-scale power systems, which are based on 
coherency, synchrony, singular perturbation analysis, modal analysis and 
identification. Two (most interesting) of the described techniques are applied to 
the reduction of the commonly used New England 10-generator, 39-bus test 
power system.

Keywords: Dynamic system, Model reduction, Equivalencing, Control theory, 
Power system.

1 Introduction

In many engineering situations and application, the model of the dynamic 
system under study can be rather complex. High order and complicated 
mathematical models accurately represent the problem at hand, but are 
unsuitable for the desired application; for instance, for analysis, optimization or 
control design. Due to limited computational, accuracy and storage capabilities, 
simplified models that capture the main features of the original dynamic 
systems have evolved [1, Preface].

In large-scale settings, the system dimension makes the computation 
ifeasible due to memory and time limitations as well as the ill-conditioning.

¹Scholar of the Ministry of Education and Science, Government of the Republic of Serbia, Futoška 111, 21000 
Novi Sad, Republic of Serbia (home address); E-mail: savodj@uns.ac.rs
²Technical Faculty Cacak, Svetog Save 65, 32 000 Čačak, Serbia; E-mail: asaric@tfc.kg.ac.rs
One approach to overcoming this problem is through model reduction [2]. It must be stressed that achieving faster simulation and optimization times is not the only goal for applying the model reduction. It is just as important to gain insight into the true cause of the observable dynamics of a system [3]. Sometimes, it is most important to get the model with the lowest number of variables [4].

For example, world-wide power systems increase in size and complexity year-by-year due to the rapid growth of widespread interconnections and much higher penetration of distributed resources. Today, the interconnected power systems cover large geographical areas and comprise thousands of devices, so the dimension of the models may easily reach the order of several thousands of state variables or more. For such large-scale power and other systems and their full detailed models, it is neither practical nor necessary to perform dynamic studies, such as the electromagnetic transient analysis, on-line dynamic security assessment, off-line stability studies, trajectory sensitivity analysis, design of different controls etc. [5, 6].

When approximating the system, one must be aware of the characteristics and properties that should be preserved in the reduced model. In certain applications one is interested in the behaviour of an autonomous dynamic system, that is one with no external inputs driving the system. However, in the case of controlled systems, one is concerned with modelling the relationship between the system behaviour and the system inputs and outputs, or actuators and sensors [7].

Model reduction involves a trade-off between model order and the degree to which the characteristics of the system are reflected by the model. Because the relative importance of various system characteristics is highly dependent upon the application, there can be no universal model reduction algorithm. The best one can hope for is a good set of tools and some reliable guidelines for using them [8]. When performing the model reduction for any system, it is important to keep in mind the limitations of the reduced model. Strictly speaking, the reduced model is valid only over the range of conditions used to generate it [4].

The problem of model reduction is to replace a given mathematical model of a system or process by a model that is much smaller than the original ones, yet still describes (at least approximately) certain aspects of the system or process (in control theory that is input-output behaviour of the system). If the approximation error is within a given tolerance, only the smaller system’s model needs to be simulated, which will in general take much less time and computer memory than the original large-scale system would do [9, Part I, Paper 6].

Model reduction involves a number of interesting issues [10, Part I, Paper 8]:
• First and foremost is the issue of selecting the appropriate approximation schemes that allow the definition of suitable reduced-order models. It is important that an appropriate dimension of the reduced model is found and that the states that should be retained are identified.

• In addition, it is often important that the reduced-order model preserves certain crucial properties of the original system.

• Other issues include the characterization of the quality of the models, the extraction of the data from the original model that is needed to actually generate the reduced-order models, and the efficient and numerically stable computation of the models.

The model reduction problem can also be defined as follows [1, Chapter 1, 2]: given the dynamic system, find a reduced order model such that the following properties are satisfied: (1) the approximation error is small, and there exists a global error bound; (2) system properties (like stability or passivity) are preserved; (3) the procedure is computationally stable and efficient.

Model reduction resolves redundancies and replaces less relevant quantities by the most significant ones. Solving lower dimensional problems one can get the statements on the system’s performance more quickly [9, Preface]. The reduced model might be used to replace the original system as a component in a large simulation, or it might be used to develop a low dimensional controller suitable for real-time applications [2]. In model reduction process we must to be careful, because if we are reducing them too much one cannot trust the results anymore.

There are some interesting issues concerning the model reduction. It would be of interest to be able to predict the error on reducing a system, i.e., how “small” can the approximate model be and yet accurately represent the process? One reliable method is a straightforward comparison of the time responses of the reduced system with the original system [11]. An important question also is how to assess the accuracy of methods, or how can we guarantee that the reduced model is a sufficiently accurate approximation of the original model [9, Part I, Paper 1]? Next question in model reduction is the choice of the order of the approximation, since it affects the quality of the approximation. Sometimes it is possible to choose the order during the construction of the reduced order model, and sometimes quality measures (like the Hankel singular values (HSVs)) can be evaluated in advance [10, Part I, Paper 5].

Much work has been done on obtaining the low-order models for large-scale systems, and many methods have been proposed, particularly in the areas of electrical and mechanical engineering, control design and computational fluid dynamics. Industry applied systems usually are nonlinear and therefore methods addressing nonlinear system approximation should be primarily considered. However, all physical systems are locally linear; in applications typically one
linearizes around an operating point of interest. Also, linear theory is reach and extensive and offers a coherent picture. There are attempts in developing a nonlinear approximation theory, but they remain mostly ad hoc [1, Chapter 1].

Some methods retain the important eigenvalues or states of the system, but some of them determine the reduced model in such a manner that it is, in some sense, an optimum approximation of the original system, without the constraints of matching eigenvalues or states [12 and references therein]. Some methods reduce the model utilising a subset of the original state variables, but some of them use a modified set of state variables, where the new system states lose their original interpretation. In computer simulation studies of physical systems, the need to retain the physical interpretation of the states is usually pre-eminent [13]. The loss of the physical significance of the states also leads to some difficulties when the original model is a part of a larger system. Such a situation occurs quite frequently in modelling the large-scale power systems [12]. Some methods are suitable for reduction of stable, others for reduction of unstable systems. Methods for effective reduction of very high-order systems are also developed. Much progress in this area has been made in the field of fluid dynamics [4 and references therein].

Sometimes, the different disciplines have a preferred model reduction technique, as modal analysis (MA) and Guyan reduction in structural dynamics, proper orthogonal decomposition (POD) in computational fluid dynamics, Padé and Padé-like approximation techniques based on Krylov methods (KMs) in circuit simulation and microsystem technology, etc. [10, Part I, Paper 1]. Model reduction is a flourishing field of research, both within systems and control and in numerical mathematics. Future developments need mathematical methods from a wide variety of fields.

This paper focuses on methods used in control theory, that can be applied to reduce the large-scale power system dynamic models. Two of these methods are applied to the dynamic models of New England 10-generator, 39-bus power system, to illustrate the way these methods could be applied to power system models.

The paper is organized as follows: problem that is elaborated in paper is formulated in Section 2. Sections 3 and 4 give an overview of model reduction techniques commonly used in control theory, for linear and nonlinear models, respectively. In Section 5, techniques used to reduce large-scale power system models are presented. Well-known the New England 10-generator, 39-bus power system is used in Section 6 to illustrate two of the described techniques. The last Section 7 is devoted to concluding remarks and future work.
2 Problem Formulation

Model of power system is a simplified representation of the power system suitable for being expressed in terms of mathematical equations and translated into computer programing code [14, Chapter 1]. The individual models of the generators, automatic voltage regulators (AVRs), turbine–governor systems, dynamic loads, etc., are given by the differential and algebraic equations, where the transmission and/or distribution networks are modelled by the algebraic equations. These equations together form a complete mathematical model of the system [15, Chapter 13].

A power system can be generally described by the following set of differential and algebraic equations, respectively [14, Chapter 1]:

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

where:

- \(x\) – vector of state variables;
- \(y\) – vector of algebraic variables;
- \(\eta\) – vector of discrete variables;
- \(u\) – vector of controllable (input) variables;
- \(t\) – time;
- \(f, g\) – set of differential and algebraic equations, respectively.

The most convenient power system model for transient stability analysis is a set of differential algebraic equations (DAEs) [14, Chapter 8]:

\[
\begin{align*}
\dot{x} &= f(x, y, t) ; \\
0 &= g(x, y, 0) ; \\
(x(0)) &= x_0, \\
(y(0)) &= y_0.
\end{align*}
\]

Small signal stability analysis studies use the properties of equilibria, or stationary points \((x_0, y_0)\) that satisfies [14, Chapter 7]:

\[
\begin{align*}
0 &= f(x_0, y_0) ; \\
0 &= g(x_0, y_0),
\end{align*}
\]

through an eigenvalue analysis of the state matrix \((A_s)\) of the system. This matrix is obtained by manipulating the complete Jacobian matrix \((A_c)\), that is defined by the linearization of the DAEs at the equilibrium point [14, Chapter 7]:

\[
\begin{bmatrix}
\Delta \dot{x} \\
0
\end{bmatrix} =
\begin{bmatrix}
\Delta x \\
\Delta y
\end{bmatrix} =
\begin{bmatrix}
f_x & f_y \\
g_x & g_y
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta y
\end{bmatrix} = A_c
\begin{bmatrix}
\Delta x \\
\Delta y
\end{bmatrix}.
\]

The state matrix \((A_s)\) is obtained by eliminating the algebraic variables (it is implicitly assumed that \(g_y\) is not singular) and, thus, one obtain [14, Chapter 7]:

\[
A_s = f_x - f_y g_y^{-1} g_x ;
\]

\[
\Delta \dot{x} = A_s \Delta x.
\]
Model reduction consists of replacing the original system (given by (1), (2), (4) or (6)) with one of a much smaller dimension according to the following guidelines [5]:

1. The reduced system must be an accurate representation of the original one for the analysis performed.
2. The cost of generating the reduced model must be much smaller than the cost of performing the analysis using the original model.

3 Dynamic Model Reduction of Linear Systems

In this section, techniques for the reduction of linear models, commonly used in control theory, are considered. Most model reduction methods focus on linear time-invariant (LTI) continuous-time and discrete-time systems. LTI continuous-time systems (time-invariant means that matrices $A$, $B$, $C$ and $D$ in (7) are time independent) are described by the following equations [10, Part I, Paper 1]:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \quad x(t_0) = x_0,$$

(7)

where:

- $x(t)$ – vector of state variables;
- $u(t)$ – vector of inputs;
- $y(t)$ – vector of outputs;
- $x_0$ – initial values of vector of state variables;
- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$;
- $n$, $m$, $p$ – order of the system, number of input and output variables, respectively.

The associated transfer function matrix (the Laplace transform of the impulse response) obtained from Laplace transforms in (7) and assuming $x_0 = 0$ is [10, Part I, Paper 1]:

$$G(s) = C (sI - A)^{-1} B + D.$$

(8)

The task of model reduction is to find a reduced-order LTI system (described by equations (7)) of order $r$ ($r \leq n$) and associated transfer function matrix $\hat{G}(s)$ which approximates $G(s)$.

Set of DAEs given by (7) is called a standard state-space system. A descriptor system or generalized state space system has the form [10, Part I, Paper 3]:

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \quad x(t_0) = x_0.$$

(9)
Most methods of model reduction focus on linear systems, which, in many cases, provide accurate descriptions of the physical systems.

Depending on the properties of the original system that are retained in the reduced model, there are different model reduction methodologies. Hence, there are techniques based on:

- **Singular Perturbations Analysis (SPA)**.
- **Modal Analysis (MA)**, such as Modal Truncation (MT).
- **Singular Value Decomposition (SVD)**, such as balanced truncation (BT), Hankel norm approximation, etc.
- **Moment Matching (MM)**, known as Krylov Methods (KMs), such as Lanczos method, Arnoldi method, etc.
- Combination of SVD and MM, known as SVD-Krylov Methods, such as ADI-type and Smith-type methods.

Many of the model reduction methods can be classified as projection based, such as MT, BT, MM methods, etc. These methods construct a reduced order model via the Petrov–Galerkin (Galerkin) projection using two (one) projection matrices whose columns form bases for relevant subspaces of the state-space. Projection methods differ in the way the projection matrices are chosen. These also determine which properties are preserved after reduction [9, Part II, Paper 12].

### 3.1 Singular Perturbations Analysis (SPA)

The interaction of fast and slow phenomena in dynamic systems results in ‘stiff’ numerical problems which require expensive integration routines. One of the most famous methods which alleviates both dimensionality and stiffness difficulties and is applicable for both linear and nonlinear systems is singular perturbation [16, 17]. This method is based on the assumption that the system equations can be separated into two parts, so-called fast and slow modes. This method decreases the order of the model, first by ignoring the fast modes of the system, then it improves the quality of the approximation by reintroducing their effect as ‘boundary layer’ corrections calculated in separate time scales. In this method, the concept of dominant subspace is bypassed by assuming that in modelling of some dynamic systems, there are some fast and slow modes and instead of just trimming the nondominant part, its steady state effect is taken into account [18]. Some issues, like initial value problems, boundary value problems and stability of reduced model are discussed in [16].

### 3.2 Modal Analysis (MA)

Model reduction methods based on MA identify and preserve certain modes of interest. MT is probably one of the oldest model reduction techniques [11, 19]. The basic idea of MT is to project the dynamics of the LTI system onto an $A$-invariant subspace corresponding to the dominant modes of the system.
(poles of $G(s)$, eigenvalues of $A$). An obvious, though certainly not always optimal, choice of dominant modes is to select those eigenvalues of $A$ having the nonnegative, or small negative real parts. Basically, these eigenvalues dominate the long-term dynamics of the solution (solution components corresponding to large negative real parts decay rapidly and mostly play a less important role). Such a view may not be the best as it neither takes into account the transient behaviour of the dynamic system nor the oscillations caused by large imaginary parts or the sensitivity of the eigenvalues with respect to small perturbations. An advantage of MT is that the poles of the reduced-order system are also poles of the original system [10, Part I, Paper 1].

In [11] is proposed a method for reducing linear systems by constructing a system of lower order which has the same dominant eigenvalues and eigenvectors as the original system. The principle of the method is to neglect eigenvalues of the original system which are farthest from the origin and retain only dominant eigenvalues and hence dominant time constants of the original system in the reduced model. The relevant variables with respect to that principle are kept and the remaining variables are expressed in terms of those variables. In [20] author is interested in determining which modes to retain in a reduced-order model for essential states. In [21] modes are used to determine the order of an accurate reduced model and the states which should be retained. In [13] a method of essential-state selection is proposed. First the dominant modes are identified, and then these are used to identify the essential states in which these modes make a significant contribution.

3.3 Singular Value Decomposition (SVD) based methods

Moore in [8] introduced balancing (balanced realization) with the aim of using it as a tool for model reduction. The main idea is that the singular values of the controllability Gramian correspond to the amount of energy that has to be put into the system in order to move the corresponding states. For the observability Gramian, its singular values refer to the energy that is generated by the corresponding states. If a linear system is in the balanced form, the HSVs provide a measure for the importance of the states, because the state with the largest singular value is the one which is the most affected by control moves and the output is the most affected by a change of this state. Therefore, the states corresponding to the largest singular values make the most influence to the input–output behaviour [3]. Truncation of the states corresponding to the smaller singular values will result in a model whose input-output behaviour closely approximates the behaviour of the original model. As the model is reduced further, the larger deviations are expected between the behaviour of the original model and the reduced one for a given input [22].

Balanced Truncation (BT) is probably the most popular projection and SVD-based method. This is mainly due to its simplicity: the construction is
based on simple linear algebra decompositions and there is no need to first choose a set of essential parameters [10, Part I, Paper 5]. To apply balanced reduction, first the system is transformed to a basis where the states which are difficult to reach are simultaneously difficult to observe. This is achieved by simultaneously diagonalizing the reachability (controllability) and the observability Gramians, which are solutions to the reachability (controllability) and the observability Lyapunov equations. Positive decreasing diagonal entries of diagonal Gramians are called the HSVs of the system. The reduced model is obtained by truncating the states corresponding to the \((n - r)\) smallest HSVs. The result is a system which contains fewer states than the original system. The number of states that can be truncated depends on the system itself and on the accuracy that is required for the system behaviour [3].

One property of BT, is that the approximation of the dynamic system is explicitly based on its input-output properties [9, Part II, Paper 14]. BT preserves several system properties, like stability and passivity and provides a bound on the approximation error [10, Part I, Paper 1]. The existence of a priori error bounds allows an adaptive choice of the state space dimension of the reduced model depending on how accurate the approximation is needed [10, Part I, Paper 3]. The existence of computable error bounds essentially distinguishes the BT technique from other model reduction approaches [9, Part I, Paper 3]. BT results in a good approximation to the original system over the whole frequency range. The disadvantage of this approach is that it does not preserve the steady state behaviour of the original system. If it is important to maintain the steady state behaviour of the original system, residualization can provide better results than truncation. Residualization is based on the idea that the derivatives of the states corresponding to small HSVs can be approximated to zero, while the rest of the system is retained [3].

Model reduction by BT requires balancing the whole system, followed by truncation. This approach may turn out to be numerically inefficient and ill-conditioned, especially for large-scale problems. The reason is that often Gramians have numerically low rank (compared to \(n\)). In many cases this is due to the rapid decay of the HSVs. Therefore, it is important to avoid formulae involving the matrix inverses. In [1, Chapter 7] several algorithms for balancing and BT are given. Although in theory these methods are similar, in practice the algorithms have quite different numerical properties. Also, in a series of papers [23 – 26] the underlying numerical algorithms for BT have been presented. The extension of BT model reduction to descriptor systems has been considered in [27 – 29].

Besides the basic (Lyapunov) balancing method, other types of balancing exist such as: stochastic balancing, bounded real balancing, positive real balancing, linear quadratic Gaussian (LQG) balancing, frequency weighted
balancing, etc. The stochastic balancing method was first proposed in [30] for balancing stochastic systems and later generalized in [31]. The relative error bound for stochastic balancing is discussed in [31]. Unlike the Lyapunov balancing method, the stochastic balancing algorithm requires solving one Lyapunov and one Riccati equation. A closely related to the balancing method is positive real balancing [30], which is applied for model reduction of positive real (passive) systems as an important subclass of dynamic systems. The positive real balancing method can be viewed as the stochastic balancing method applied to the spectral factor of the given passive system and requires solving two positive real Riccati equations. Another method which also requires solving two Riccati equations is bounded real balancing which is applied to the bounded real systems. This method, together with the absolute error bound, was first introduced in [32]. LQG balancing, also referred as the closed-loop balancing first introduced in [33], is mainly used for reduced-order controller design. Several ways of frequency weighted balancing have been introduced in the literature. Most of these methods need input and output weights [34 – 38], which are usually not explicitly specified, and try to find a reduced-order model which minimizes the weighted error. However, often the original problem is to approximate the transfer function over a frequency interval $[\omega_1, \omega_2]$ and no input or output weights are provided. In [39] type of weighted balanced reduction where for a given frequency band $[\omega_1, \omega_2]$, the construction of the weights are avoided, simply by using the frequency domain representation of the Gramians. Although the method works quite efficiently in practice, stability is not guaranteed and no error bound exists. Similarly to the band-limited frequency weighted balancing method, a time-limited balancing method where the Gramians are computed over a finite time interval $[t_1, t_2]$, is also introduced in [39].

In [1, Chapter 7] balancing of unstable systems is discussed. BT method is able to handle the reduction of unstable systems either via modal decomposition or coprime factorization techniques [10, Part I, Paper 9 and references therein]. In [10, Part I, Paper 3] a survey on BT model order reduction for linear time-invariant continuous-time descriptor systems is presented. In [40] the BT approach to model reduction is considered for linear discrete-time periodic systems with time-varying dimensions.

A model reduction technique closely related to BT and HSVs is Hankel norm approximation, which is based on a remarkable result of three Russian mathematicians (Arov, Adamjan and Krein) [1, Foreword]. When applied to the stable dynamic systems, this approach guaranteed to preserve stability and provides bounds on the approximation error [41].
3.4 Krylov Methods (KMs) (or Moment Maching (MM) methods)

The starting point for model reduction in the electronics industry is usually attributed to a method termed Asymptotic Waveform Evaluation (AWE) [42]. The underlying idea of this method is simple, approximating the moments of the transfer function of the system. The idea being that moments will decay, so that calculating a sufficient, but finite, number of moments will eventually lead to an accurate approximation of the transfer function. Soon after its publication, it was realized that the method suffers from numerical problems. In order to overcome these numerical difficulties associated with AWE, use of KMs is proposed in [43]. The method was termed Padé-via-Lanczos (PVL), and shown to match the moments of the transfer function. It demonstrated that KMs, developed mainly in the area of numerical linear algebra, can also be used to perform model order reduction [9, Part I, Paper 1].

The property of MM methods is the leading coefficients of a power series expansion of the transfer function of the reduced system around a user-defined point that have to match those of the original system transfer function. While SVD-based methods are shown to yield excellent global approximation in the frequency domain, MM techniques have good performance in a limited range of frequencies. Nevertheless, MM methods present less computational effort and less storage requirements, requiring little empirical parameter adjusting [5].

In [1, Chapter 11] a set of KMs is introduced. KMs can be iteratively implemented. The iterative process is usually stopped when the difference between subsequent iterations is below a certain threshold [9, Part I, Paper 1]. These schemes were originally developed for computing eigenvalues and eigenvectors, but can be applied to model reduction via moment matching (MM). Typically, these methods require only of the order of $n^2$ operations and MM is achieved without computation of moments. KMs are attractive for large-scale sparse systems, since only matrix-vector multiplications are required, and they can easily be generalized for descriptor systems [10, Part I, Paper 3]. Their disadvantage, however, is that they lack global error bounds, depend significantly on the choice of certain parameters, the resulting reduced models have only locally good approximation properties and they have difficulties when special system properties, such as stability or passivity, are to be preserved by the reduced model (usually post-processing is needed to realize these properties [9, Part I, Paper 3]). Broader list of properties of KMs is given in [1, Chapter 11].

In [44] KMs are discussed and three algorithms are proposed: 1) dual rational Arnoldi, 2) rational Lanczos and 3) rational power KM. Techniques for choosing the matching frequencies, estimating the modeling error, insuring the model’s stability, treating multiple input and multiple output systems, implementing parallelism etc. are examined to various degrees.
3.5 Singular Value Decomposition-Krylov methods (SVD-KMs)

Both SVD-based methods and KMs have their own set of advantages and disadvantages, as mentioned earlier. To be applied to large-scale problems SVD approximation methods need to be combined with iterative methods which can be implemented by means of vector-matrix multiplications exclusively. This leads to KMs. Methods that combine the best attributes of both SVD-based methods and KMs are called SVD-KMs [1, Chapter 12].

For small-to-medium scale problems, the BT can be implemented efficiently using the Bartels-Stewart [45] method, as modified by Hammarling [46], to solve the two Lyapunov equations. However, the method results in \(O(n^3)\) arithmetic operations and \(O(n^2)\) storage. For large-scale sparse problems, iterative methods are preferred, since they retain the sparsity of the problem and are much more suitable for parallelization. The Smith method [47], the alternating direction implicit (ADI) iteration method [48], and the Smith(l) method [49] are the most popular iterative schemes developed for large sparse Lyapunov equations. Unfortunately, even though the number of arithmetic operations is reduced, all of these methods compute the solution in dense form and hence require \(O(n^2)\) storage [10, Part I, Paper 2].

It is well-known that the Gramians often have low numerical rank (i.e. the eigenvalues of Gramians decay rapidly). One must take advantage of this low-rank structure to obtain approximate solutions in low-rank factored form. Most low-rank methods, such as [50 – 53], are KMs based. As stated in [49], even though these methods reduce the memory requirement, they usually fail to yield approximate solutions of high accuracy. To reach accurate approximate solutions, one usually needs a large number of iterations, and therefore obtain approximations with relatively high numerical ranks; see [49]. For large-scale sparse Lyapunov equations, a more efficient low-rank scheme based on the ADI iteration was introduced, independently in [49] and [54]. The method was called the low-rank ADI iteration (LR-ADI) in [49] and the Cholesky factor ADI iteration (CF-ADI) in [54]. Even though LR-ADI and CF-ADI are theoretically the same, CF-ADI is less expensive and more efficient to implement. Indeed, LR-ADI can be considered as an intermediate step in deriving the CF-ADI algorithm. Another low-rank scheme based on the ADI iteration was also introduced in [49]. The method is called the cyclic low-rank Smith (LR-Smith(l)) method and is a special case of LR-ADI, where \(l\) number of shifts are re-used in a cyclic manner. While solving the Lyapunov equation, where \(B\) has \(m\) columns, the LR-ADI and the LR-Smith(l) methods add \(m\) and \((m \times l)\) columns respectively to the current solution at each step. Therefore, for slowly converging iterations and for the case where \(m\) is too big, e.g. \(m = 10\), the number of columns of the approximate low-rank Cholesky factor can exceed manageable memory capacity. To overcome this, a Modified LR-Smith(l)
method that prevents the number of columns from increasing arbitrarily at each step is introduced in [55]. In fact, the method only requires the number of columns \( r \) which are needed to meet the pre-specified BT tolerance. Due to the rapid decay of the HSVs, this \( r \) is usually quite small relative to \( n \). Consequently, the memory requirements are drastically reduced. Because these low-rank methods produce the Cholesky factor of the solution to the Lyapunov equation, they are especially well-suited to be used in conjunction with approximate BT to reduce large-scale linear systems [10, Part I, Paper 2].

Ref. [10, Part I, Paper 2] surveys Smith-type and ADI-type methods used for solving large-scale sparse Lyapunov equations and consequently for BT of the underlying large sparse dynamic system. These allow the computation of the factors at a computational cost and a memory requirement proportional to the number of nonzeros in matrix \( A \). Thus, implementations of BT based on these ideas are in the same complexity class as Padé-approximation and POD [10, Part I, Paper 1]. Connections between different Smith-type and ADI-type methods, convergence results, and upper bounds for the approximation errors are discussed in [10, Part I, Paper 2].

In [41] a comparative study of seven algorithms for model reduction is presented: BT, approximate BT, SPA-based method, Hankel norm approximation, Arnoldi procedure, Lanczos procedure and Rational KM.

### 3.6 Other methods

In [12] a criterion is proposed for selecting the most important states of a large-scale linear system to be retained in a reduced model. The effective participation of each state is estimated by evaluating its contribution to the total impulse response energy at the output of the system and a procedure for obtaining the reduced model, based on this criterion, is described. The states of the reduced model retain their physical meaning. The method may be regarded as the combination of SPA and aggregation [56].

In [4] a method that combines POD (see Subsection 4.1) and concepts from balanced realization theory, for performing a reduction of a large-scale linear system is presented. The method is particularly effective when a small number of outputs is of interest.

In [57] both particle swarm and genetic algorithm optimization are employed for finding reduced models of single-input-single-output (SISO) large-scale linear systems. For the given criteria, parameters of the transfer function approximation are optimized. Both the techniques guarantee stability of reduced order model if the original model is stable. It was shown that the steady state responses of both the proposed reduced order models were exactly matching with that of the original model, and that the transient responses of evolutionary reduced models were very close to that of original model.
4 Dynamic Model Reduction of Nonlinear Systems

A nonlinear dynamic system with its control can be generally described by the following set of DAEs [15, Chapter 1]:

\[
\dot{x} = f(x(t),u(t)) \quad \text{and} \quad y = g(x(t),u(t)) \quad \text{with} \quad x(t_0) = x_0,
\]

where:

- \( x(t) \) – vector of system state variables;
- \( u(t) \) – vector of control signals which affect the system to achieve a desired behaviour (input vector);
- \( y(t) \) – vector of output signals which serve to assess whether the control achieved the desired goal (output vector);
- \( x_0 \) – initial value;
- \( f, g \) – sets of differential and algebraic equations, respectively, which describe the relationship between input, state and output variables, respectively.

Goal is to replace the original model (10) with one less complex, such that the input-output behaviour of the system is sufficiently well approximated.

Methods and supporting theories for the reduction of linear models are well established in the past. For nonlinear systems no complete theory for model reduction currently exists [3]. Since the days of 19th century mathematician Henri Poincaré, we have known that it is impossible to find general analytical solutions to nonlinear systems. Furthermore, the development of nonlinear dynamics has proven that such systems, even when they have very few state variables, can produce highly complex and intricate behaviour that would be impossible to anticipate, let alone analyse, directly from their structure. Thus, in the absence of a “grand unified theory” of dynamic systems, we shall always have to rely on simulation to discover the dynamics implied by the structure [58].

When transferring approaches from model reduction of linear systems, especially projection based methods, fundamental differences emerge (for details see [9, Part II, Paper 17]). With respect to computation time no reduction will be obtained unless additional measures are taken or other strategies are pursued [9, Part II, Paper 17]. One might be sceptical that a method derived from linear systems theory may have any use for the nonlinear models found in system dynamics [58].

Despite considerable progress in model reduction techniques over the last few decades, robust general procedures for nonlinear model reduction have not become available yet. It cannot be expected that truly very large scale nonlinear systems will easily be reduced unless the system have a specific simplifying structure [9, Part II, Paper 18]. Larger nonlinear systems can only be tackled by
using specific information about the underlying nonlinear system. It is to be expected that successfully generated reduced order models must also make use of huge numbers (millions) of measured or simulated results [9, Part I, Paper 1].

There has been a surge of interest in reduced-order modelling of nonlinear systems, but still there exist only a few group of methods that have turned out to be useful in practical applications, like:

- **Proper Orthogonal Decomposition (POD) (Karhunen–Loève decomposition).**
- **Trajectory Piecewise Linear (TPWL) method.**
- **Balancing-based methods.**
- **Reduction by optimizing system matrices.**
- **Projection from a linearized model.**

One of the most popular methods for reducing nonlinear systems is POD. It generates matrix of snapshots (in time), then calculates the correlation matrix and its singular value decomposition. The vectors corresponding to the largest singular values are used to form a basis for solutions. The TPWL is also very popular. However, care has to be taken when POD and TPWL are being applied, as the approaches seem to be very sensitive to various heuristics [9, Part II, Paper 17]. In the area of dynamical systems and control, methods developed by Fujimoto and Scherpen [59] and Verriest [60] are promising, although so far the methods are limited to small nonlinear systems [9, Part I, Paper 1].

### 4.1 Proper Orthogonal Decomposition (POD)

The POD is method widely used to determine efficient bases to construct the projection matrix. In this method for a fixed input the state variables trajectories at certain instances of time are measured (sampled) and saved in the matrix of snapshots. If the singular values of this matrix decrease rapidly, this matrix could be approximated by a low-order matrix. Dominant subspace in the sense of POD is the part of the state space which absorbs the most energy from specific inputs [18]. It does not take the state-to-output behaviour into account [61].

The range over which the sampling is performed is determined by assessing the important frequency range in the problems at hand. To determine the specific snapshot locations within this range, one uses a combination of experience and intuition. Often, the required density of snapshots will be determined a posteriori from the performance of the reduced model. If the desired dynamics cannot be accurately captured, more snapshots must be included in the POD process and the basis vectors recalculated [4].

The POD essentially provides an orthonormal basis for representing the given data in a certain least squares optimal sense. Truncation of the optimal
basis provides a way to find optimal lower dimensional approximations of the given data. In addition to being optimal in a least squares sense, the POD has the property that it uses a modal decomposition that is completely data dependent and does not assume any prior knowledge of the process that generates the data [62].

The POD does not provide reduced model, but only the basis for the solution space. Combined with the Galerkin projection procedure, it provides a powerful method for generating lower dimensional models of dynamic systems [62], [1, Chapter 9], [10, Part I, Paper 10]. Error analysis for nonlinear dynamic systems and the relationship with balancing have been discussed in [10, Part I, Paper 10 and references therein], [1, Chapter 9]. The sensitivity of the results of POD to perturbations in the data used to form the reduced model, as well as computational efficiency gained in using POD in model reduction, are studied in [62].

4.2 Trajectory Piecewise Linear (TPWL) method

The TPWL method applied to nonlinear DAEs is a promising technique. The idea of TPWL method is to reproduce the typical behaviour of the full nonlinear system by a varying combination of a set of order reduced linear models. For this purpose a training input is chosen that drives the system into typical states, i.e., situations. A transient simulation with the chosen input yields a collection of points approximating the trajectory at timepoints. On the trajectory, points are chosen around which the nonlinear functions are linearized. The linear models are reduced individually. This delivers local reduced subspaces. A common subspace is constructed that describes the primary information of all local subspaces. All linear models are projected on this space. Finally a weighting is used to select linear models which are valid in a certain situation [9, Part II, Paper 17].

The first crucial point in TPWL method is to decide when to add a linear substitute for the nonlinear problem automatically during the training simulation (selection of linearization points). Too many linear models could make the final model slow, too few could make it inaccurate. To have minimum complexity one aims at having to deal with a combination of just a small number of linear models. Hence, the weighting function is chosen such that only a few (in the ideal case just one) dominant linear models are chosen. When reducing the linear models any linear model reduction technique can be applied. However, the reduction of linear models has the effect on the overall accuracy of the approximation [9, Part II, Paper 17 and references therein].

The TPWL method has several advantages compared to other methods. First it can speed up a simulation, because only small linear systems to approximate original system are solved. Then the well-developed linear model reduction techniques can be used. The properties of the TPWL model can be
increased if a very good weighting procedure is constructed. The TPWL method has also the very nice property that it is scalable. This means that by using a different linear model reduction techniques and weighting methods, one can change the method from a very fast, but not so accurate, to a slower, but also much more accurate method. This means that the user himself can decide what he desires: speed or accuracy [63, Chapter 7].

4.3 Balancing-based methods

Another well-known method is nonlinear balancing which is an extension of balancing for linear systems (see Subsection 3.3) in the sense that it is based on extended definition of balancing and Hankel singular functions [64]. This method similarly to the concept of balancing for linear systems finds a coordinate transformation that balances the system due to extended definition of balancing for nonlinear systems. Some related results are given in [65, 66]. As it is apparent, on the one hand, the procedure for nonlinear balancing presents computational difficulties, which restricts its application to very low order nonlinear systems. On the other hand it finds a very meaningful coordinate transformation as a point of view of dynamic systems which specifies the dominant subspace of state space by assuming the effect of input and state variables on output in the sense of energy [18].

The main drawback of the procedure proposed in [64] is its extensive numerical requirement even for small systems [67]. The only numerical implementation of method presented in [64], which uses a Monte Carlo approach, is given in [67]. However, after the coordinate transformation is applied and even without reduction of the model, the transformed system does not exhibit the same input-output behaviour as the original system, because of the approximations that were applied during the computational procedure [22].

An approximation to balancing of nonlinear systems via the concept of empirical Gramians has been introduced in [68, 3, 7]. The empirical Gramians capture some of the nonlinear behaviour of the system, while being simple to compute [3, 7]. They are calculated using data from either simulation or experiment (for different values of inputs and initial conditions) to identify the dynamics relevant to the input-output map of the system and then balanced by the same procedure as is used for linear systems. The balancing transformation is used within a Galerkin projection in order to transform the nonlinear system into balanced form. The resulting nonlinear equations can be reduced using different truncation or residualization methods. The number of remaining states can be adjusted by a trial and error procedure to achieve optimum performance.

In [22, 69] controllability and observability covariance matrices have been proposed as an extension of the empirical Gramians. Covariance matrices are determinated from data collected in the operating region of the process and have some advantages over empirical Gramians. These covariance matrices can be
used to compute a balancing-like transformation followed by a model reduction step via truncation or residualization [61]. The procedure will result in a model of reduced size that describes the input-output behaviour locally, but no conclusions can be drawn about the global behaviour of the reduced system [22]. The reason the reduced model does not describe the global behaviour of the original one is that the covariance matrices are only locally defined. However, the information captured in the covariance matrices gives a better reflection of the system dynamics in this operating region than would result from linearizing the model at the operating point.

In [9, Part II, Paper 17 and references therein] BT in nonlinear model reduction is discussed.

In [63] several options for reducing a nonlinear differential-algebraic model are investigated: POD, TPWL, Empirical balanced truncation and Volterra series. Complexity analysis of these techniques, choice of the optimal linearization points and weighting procedure are also discussed.

### 4.4 Reduction by optimizing system matrices

Model reduction by optimizing system matrices is another method for reduction of nonlinear systems with specific structure [70]. In this method determination of dominant state state variables (dominant subspace) plays a significant role in the quality of model reduction. The first idea for pointing out the dominant state variables is engineering impression, which could be helpful in many practical problems but for complex technical systems, choosing dominant state variables usually is not a straightforward task and some more advanced methods are required [18]. Once the dominant variables are determined, the system matrices of the reduced model are determined by solving optimization problems, defined to minimize the difference in the behaviour of the original and reduced model.

### 4.5 Projection from a linearized model

In methods belonging to this group, the reduced model is found from analysing a linear model, calculated by linearizing the original nonlinear system at some operating points. Because the linearized model is a good approximation of the original system for the points close to the operating point, the linear reduced system found in this way approximates the behaviour of the original model only for a range around the operating point. This range can be extended by adding some nonlinearities to the reduced model. This can be done by first representing the linear reduction procedure by a projection and calculate the projection matrices. Then, the projection is directly applied to the original nonlinear system. Although the approach uses the linearization of the nonlinear model, the final reduced system is nonlinear. However, that finding a linearized model can be difficult [71 and references therein].


4.6 Other methods

In model reduction of linear systems, KMs construct order reduced models of the system, such that the moments of the transfer functions of full and reduced problem match up to a certain order. As there is no direct counterpart of the transfer function for the nonlinear problem, KMs are applied to two types of systems derived from the general problem: bilinear systems and linear periodically time-varying systems [9, Part II, Paper 17].

In [72] a procedure, where the linear transformation is found from a covariance matrix that is computed from data collected along system trajectories, is investigated. These trajectories represent the system behaviour under a constant input, but starting from different initial conditions. In [73] this model reduction approach to models described by differential-algebraic equation (DAE) systems is extended. The method consists of generating state trajectories, computing a covariance matrix from data, extracting the principal components of the covariance matrix and performing a projection on the original system. The case of trajectories generated by different initial conditions under constant inputs as well as the case where the trajectories start at the steady-state operating point and are generated by step changes in the inputs to the system are investigated. While similarities exist between these methods and balanced model reduction, it should be pointed out that the emphasis of these two methods is not to retain or approximate the input-output behaviour of the system [22].

The model reduction procedure presented in [61] combines elements from balanced model reduction with system identification techniques in order to reduce the differential and the algebraic equations, while retaining the control-relevant properties of the model. The procedure consists of three major steps. In a first step of the procedure, transformations are applied to both differential and algebraic equations. The transformation for the differential equations is computed via balancing of the covariance matrices or via POD, while the transformation matrix for the algebraic equations is obtained by singular value decomposition (SVD) of the state covariance matrix computed for different excitations of the system. In a second step, both differential and algebraic equations of the transformed system can then be reduced via a truncation procedure. The model is further reduced (third step) by replacing the algebraic equations with an identified model. A comparison of the results obtained from POD and balancing indicates that balancing performs better than POD since the input-to-state and the state-to-output behaviours are simultaneously taken into account while POD only uses the input-to-state behaviour.

In [18] a procedure for the order reduction of nonlinear systems by combining linear balancing with an optimization procedure is proposed. The method consists of the following steps: production of the snapshots,
linearization of the nonlinear system, finding the dominant subspace and application of system matrices optimization method.

In [71] a approach to the model order reduction of nonlinear systems, which does not need a simulation of the original system, is presented. By separating the linear and nonlinear parts of the original nonlinear model, the idea is to consider the nonlinearities of the resulting system as additional inputs. The results confirm that simulation-free reduction can deliver acceptable results, while being fast and simple to implement.

5 Dynamic Model Reduction in Power Systems

The problem of modelling a large-scale power system arises for a number of reasons including [15, Chapter 14]:

- **Practical limitations on the size of computer memory.**
- **The excessive computing time required by large-scale power systems, particularly when running dynamic simulation and stability programs.**
- **Parts of the system far away from a disturbance have little effect on the system dynamics and it is therefore unnecessary to model them with great accuracy.**
- **Often parts of large interconnected systems belong to different utilities, each having its own control centre which treats the other parts of the system as external subsystems.**
- **In some countries private utilities compete with each other and are reluctant to disclose detailed information about their business. This means that vital data regarding the whole system may not be available.**
- **Even assuming that full system data are available, maintaining the relevant databases would be very difficult and expensive.**

Power system dynamic model is difficult to analyze because of its size, complexity and nonlinearity (in general). The dynamic equivalent is an important tool in power systems to deal with size and complexity, since it provides a reduced representation of the system, while preserving, with reasonable accuracy, the dynamic characteristics of interest. The problem of obtaining an equivalent representation of a portion of a system modeled by the nonlinear model is difficult, because of the lack of general analytical methods for dealing with the nonlinear equations required to describe them [74].

The desired characteristics of the reduced model will depend on its application and use. Although it seems that a large number of variables can be involved in the problem of interest, typically only a subset of these variables is essential. System contains information that is not vital for the problem of interest. Also, a reduced power system model provides better insight into the system characteristics of interest.
The key component in model reduction is knowing how far the model reduction can be taken without destroying the accuracy of the reduced order model vis-a-vis the unreduced or "full" model [75]. Given the complexity of the context in which dynamic equivalencing is carried out, the real validation of any approach to it has to lie in the quality of the results one obtains in practice, and the ease with which these results are obtained [76].

While analyzing a large system, the engineers are usually interested in the behaviour of a certain part of the system. Such a part of the large system is called internal or study (sub)system (area) and the rest of the system is referred to as external (sub)system (area). Model reduction or equivalencing is the process of reducing the complexity of external system model while retaining its effect on the study system [6]. The study system is kept untouched. A reduced model of the external system is created assuming that the disturbance occurs only inside the internal system. The border nodes between the internal and external system are sometimes referred to as the boundary nodes [15, Chapter 14].

The methods by which the equivalent of a system can be produced can be broadly divided into two groups, depending on whether or not they require any knowledge of the configuration and parameters of the system itself. Methods that do not require any knowledge of the system are used for online security assessment and typically use the measurement of certain electrical quantities to form the equivalent. Methods that do require knowledge of the system are used for off-line system analysis [15, Chapter 14].

The bulk of model reduction techniques in power systems are tailored for the tasks of control design and transient/small signal stability analysis [5]. Concepts like:

- **Coherency** treated in [74, 77, 78, 79, 80, 81],
- **Synchrony** introduced in [82, 83],
- **Singular Perturbations Analysis** (SPA) given in [84 – 87],
- **Modal Analysis** (MA) presented in [88, 89, 90, 85, 91, 92, 93], and
- **Identification** used in [94 – 103, 6]

form the basis for a wide variety of model reduction tools developed.

Among existing software tools for equivalencing are DlgSILENT, DYNRED and PSS/E [6].

### 5.1 Coherency

A well-known concept for dynamic equivalencing of power systems is coherency. Generator coherency describes the similarity of generator responses following disturbances [104]. A coherent group of generating units, for a given perturbation, is a group of generators oscillating with the same angular speed, and terminal voltages in a constant complex ratio [105]. It is usually possible to
find groups of coherent generators because some groups of generators in the system have a natural tendency to swing together [15, Chapter 14]. One hopes that for a sufficiently wide range of disturbances, say all faults that are electrically far away, the composition of these groups can be considered constant [74].

Each coherent group of machines can be replaced by a large equivalent machine. The dynamic equivalent is a single generating unit that exhibits the same speed, voltage and total mechanical and electrical power as the group during any perturbation where those units remain coherent [105]. The equivalent generator is usually represented by the classical model with constant equivalent transient electromotive force and by the swing equation. If more detailed models are used, then parameters of the equivalent unit can be found by matching the frequency response characteristics of the equivalent unit to the characteristics of the aggregated units. Details can be found in [105, 106].

A set of sufficient conditions for theoretical coherency consists of [107]: (1) identical machine and control device models, parameters, and power output levels; (2) identical admittances connecting the generator buses to each boundary bus. If the theoretical coherency conditions are satisfied, a disturbance will not provoke any relative motions between the machines in the area. To an observer, the motions of these machines are seen as if they were originated from one single machine [107]. In MA, this corresponds to a situation where modal variables representing the swinging of generator rotors inside the coherent group are not excited by disturbances. Disturbances excite only that modal variable that represents the swinging of the whole coherent group with respect to the rest of the system [15, Chapter 14]. In real-world power systems (apart from the trivial case of identical generating units operating in parallel on the same busbar) the exact coherency practically does not appear, but the definition is useful for theoretical considerations. Inaccuracy of coherency means that all the dynamic properties of the original (unreduced) model will be maintained only to some degree by the equivalent (reduced) model. Hence it may be expected that also eigenvalues and eigenvectors of the equivalent model will be only approximately equal to eigenvalues and eigenvectors of the original model. It is important here that the equivalent model maintains as precisely as possible those modal variables that are strongly excited by disturbances in the internal system and which therefore have the strongest influence on power swings in the internal system (dominant modal variables) [15, Chapter 14].

Analytical conditions for coherency are discussed in [108–112]. Approximated coherency has been studied in [81, 80]. Slow coherency, or coherency in the slowest modes of the system is discussed in [110, 113]. Slow coherency concept clusters generators based on the a priori assumption that the equivalent should retain only the slowest system modes.
Dynamic Model Reduction: An Overview of Available Techniques with Application...

The process to reduce the dynamic model of a power system using the concept of coherency can be divided into three stages: 1) coherency identification, 2) the aggregation of the coherent generators and their control devices, and 3) network reduction. In the literature, different methods have been proposed for coherency identification [114, 115, 116, 80, 117, 118, 119, 120, 121]. To construct equivalent systems, various methods have been developed [79, 113, 122]. In 1993, EPRI released the dynamic reduction software package (DYNRED), which includes a range of coherency determination and dynamic reduction options [122, 123, 124]. Ref. [125] describes the authors’ experience with the application of DYNRED to perform dynamic reductions on three large power systems with significantly different characteristics for the purpose of studying transient and small-signal stability. In [126] a software tool, that simplifies and speed-up the computation involved in three mains steps of the dynamic equivalent determination procedure based on coherency, is presented.

Coherency equivalents do not need any special interfacing with the study system model, because the equivalent is in the form of a model of an actual physical component. The coherency technique has been found to be cost-efficient and reasonably accurate. Nonlinearities can be introduced in the equivalent machine model, thus extending the validity of the coherency equivalents to large disturbances. However the coherency technique suffers from the disadvantage of being a purely empirical. The quality of the equivalent obtained is dependent on the perturbation chosen to determine coherency, but it is difficult to lay down guidelines for choosing the most appropriate perturbation [127]. Generator coherency is also dependent on system conditions, and coherency information obtained under one particular condition might not be applicable to another condition [104]. For a new operating condition, generator coherency behaviour needs to be re-evaluated. This process is time-consuming, especially for large interconnected power systems. In [104] a systematic approach on the basis of eigen-sensitivity, to predicting the changing patterns of generator slow coherency for different operating conditions and forming an appropriate study area boundary by including the critical generators in the external area that become strongly coherent with the study area, is presented. Since the coherency impose the regions into which the network can be divided, it is not always possible to reduce a given part of the network [6]. Approaches that are able to retain a given part of the network are [128, 129, 95].

Another important observation in the simulation of power system dynamic response concerns the influence of the coherency error of a group of generators aggregated in the external system on the simulation accuracy of the internal system. The more the group of aggregated generators is remote from the internal system, the smaller is the influence of the coherency error on the simulation accuracy of dynamic response in internal system [15, Chapter 14].
5.2 Synchrony

Another well-known concept for dynamic equivalencing of power systems, which generalizes the concept of slow-coherency, synchrony, is introduced in [82]. Synchrony is defined with respect to a selected subset of modes of a linearized model. Such a subset is called a chord. Two generators are termed exactly or approximately synchronic in a chord (ν), or are ν-synchronic, if their angular variations are exactly or approximately in constant proportion for any transient in which only the modes in ν are excited. In contrast, ν-coherency requires exact or approximate equality of the angular variations, and slow-coherency further restricts ν to contain only the slowest electromechanical modes of the system. A (maximal) set of ν-synchronic generators constitutes a synchronic group, or a synchronic area [130]. An algorithm to identify such areas and assign a reference machine to each area, is proposed in [82].

Ref. [131] elaborates synchrony and outlines how it can form the basis for efficient construction of dynamic equivalents by aggregation. An approach for selecting the inter-area modes that are to be represented by the aggregate model is described. An algorithm for recognizing the approximate synchrony is presented, and improvements over the standard slow-coherency recognition algorithm are noted.

In [83] a method for dynamic equivalencing, termed synchronic modal equivalencing (SME), based on synchrony and modal equivalencing, is proposed, and algorithms for constructing effective dynamic equivalents for classical swing-equation models are presented. [130] generalizes the framework to obtain dynamic equivalents for detailed models described in structure-preserving, DAE form. The objectives of [76] are to summarize certain extensions and simplifications of SME and layout of the recommended SME procedure.

The SME procedure for constructing a dynamic equivalent, as proposed in [76], occurs in three stages. The first stage involves preparation of the various dynamic models that are necessary or useful for the process. The second stage is devoted to analysis of the modal structure of a simple, low-order, linearized, electromechanical model of the system, such as a linearized, undamped, swing-equation model. This analysis leads to the selection of a suitable chord, and to the decomposition of the model into synchronic groups, each associated with a distinct basis (reference) generator. One of these synchronic groups (or possibly the union of several synchronic groups) is selected as a study group. In the third stage procedure returns to the original, large, unreduced, nonlinear model of the system, to impose on it the structural decomposition suggested by the preceding analysis of the simplified model. The models of all the generators of the study group and of all the basis generators external to the study group are retained in full nonlinear detail. Each of the remaining ,,less-relevant“ generator models is
Dynamic Model Reduction: An Overview of Available Techniques with Application...

replaced by a simple nondynamic linear circuit that contains a dependent current source driven by the motions of the basis generators. Impedance-load buses that interconnect only the replaced generators can be eliminated, with their effects being accounted for by including the bus voltages of the replaced generators as additional driving variables for the dependent current sources. The remainder of the network is left unmodified.

The SME is aimed at structure-preserving dynamic equivalencing of large power system models. It is motivated by the slow-coherency method for system partitioning and aggregation, but its formulation and associated computational algorithms are more general in some important aspects [76].

5.3 Singular Perturbations Analysis (SPA)

Power system dynamic analysis encompasses a wide time span of responses, ranging from lightning phenomena in microseconds to automatic generation control over periods of minutes. Within the time span of stability analysis, there are also time scales arising from the various speeds of responses of different devices such as synchronous machines and excitation systems, as well as from the interconnections within large power systems [87]. The natural time scale separations resulted in time scale modelling and SPA.

Singular perturbations are most conveniently performed on two-time-scale systems in the explicit form elaborated in [16, 17]. Not all models of power systems with time scales are in the explicit form. Thus the first task of time scale modelling is to identify singular perturbation parameter (the ratio of the time-scales of the slow and fast phenomena [85]), which could be due to ratios of small and large time constants, stray and linkage inductances, weak and strong connections, etc., and to formulate physical transformations to obtain the slow and fast variables. Once the two-time-scale system is in the explicit form, a slow subsystem can be obtained by setting singular perturbation parameter equal to zero, while a fast subsystem can be obtained to model the fast transients of vector of fast variables from its 'quasi-steady’ state. The subsystems can then be used for analysis in separate time scales [87].

In [85] it is shown how other important forms of singularly perturbed systems can be reduced to the explicit form. In [87] the two-time-scale modelling technique is applied to several modelling problems occurring in power systems. The results described in [87] illustrate the role of singular perturbations as an analytical tool in power system modelling at both the device and system level. They also illustrate how additional insights into power system dynamics can be obtained in the process.

5.4 Modal Analysis (MA)

In MA, the complex nonlinear model of the system (2) is reduced to a linear form of state equation representation (4). The MA technique requires the
computation of eigenvalues of matrix $A_S$ defined by (5), which is time-consuming, and provides a reduced system of differential equations which cannot be interpreted, in general, as representing models of physical units (equivalents do not have structural identity) [105]. However, MA methods have a rigorous mathematical basis and provide a good insight into the various modes of oscillations present in the system. In contrast to the coherency equivalent, the quality of the modal equivalent does not depend on the appropriateness of the perturbation chosen to construct the equivalent [127]. The MA has been proven efficient, yet the determination of the full set of truly dominant modes of the system is not a completely resolved question, although significant progress has been made [91, 92].

The modal-coherency method [132, 133, 134, 135, 127] is developed for deriving the dynamic equivalents combining the advantages of modal and coherency analysis. A modal-coherent equivalent can preserve not only the coherent groups of the original system model, but also some original modes of oscillations. A modal-coherent equivalent represents a valuable tool for transient stability analysis since it is constructed only once for a given utility and can then be used in the transient stability study of any disturbance that might occur in the utility.

5.5 Identification

Identification based approaches are based on parametric identification techniques. The objective is to estimate a set of parameters belonging to a model that is assumed to represent some part of a power system, based on measurements of important signals. This kind of technique has the attraction of not needing detailed information of the subsystem to be identified. So, the essence of system identification consists of matching signals from a real system that is undergoing a random perturbation, with the values of these signals calculated on a model of the system, and adjusting the model to reduce the difference [102].

5.6 Other methods

Well-established solutions for the reduction of dynamic models exist in control theory based on the so-called balanced realization (see Subsection 3.3). However, a direct application of this approach to the case of power systems faces two major difficulties [136]. First, the reduced model should preserve the structure of the physical system, i.e., differential equations for generators and their regulations and algebraic equations for the grid. Thus, nonphysical state variables are not appropriate. Next, the size of the problem is important when large-scale power systems are addressed.

In [137] balanced realization theory is combined with modal-coherency analysis to provide accurate estimates for dynamic equivalents for use in
transient stability studies based on coherency aggregation techniques. Relationship between modal-coherency and balanced realization theory is investigated. A method of estimating the feasible order reduction using the theory of balanced realizations is proposed. This estimate of order reduction is used to determine when to stop the aggregations dictated by the modal-coherent analysis. This approach is appropriate for moderate sized power systems [138]. In this reference also an alternative approach appropriate for larger systems, based on the intergenerator coherency ranking table, is discussed and compared to the balanced realization approach. Ref. [75] presents the theoretical results that establish an analytical bridge between coherency and balanced realizations. Analytical connections that relate coherent behaviour (perfect coherency) of generators to redundancy in the Gramians (redundancy in the states) of the linear model of the power system are determined.

In [136], the construction of the newly defined border synchrony chord was combined with the aggregation techniques [11, 56] to provide a structure-preserving reduced model by balancing techniques. The relation between synchrony and the lack of controllability and observability is investigated in the general context of power systems. The vision obtained fills the gap between the two approaches and makes a bridge between several existing characterizations of the redundancy in power systems, like the synchrony and the slow-coherency and the recent reduction principles of the control system theory. The main advantage against the classical balanced-reduction is that the result is a physical power system and not only a mathematical object (state-space realization with state-variables with no physical meaning). In [139] a method to construct power system dynamic equivalents for day-ahead stability studies is proposed. The notion of synchrony with respect to a given border introduced in [136] was used. It allows one to determine classes of machines of the external area which have the same influence on the study system in a robust manner and without using information from the study system. Starting from these classes, aggregation methods [11, 56] are used to provide a physical dynamic equivalent (a dynamic equivalent which consists in real machines, regulations, lines, etc. and not only a mathematical equivalent), i.e. structure-preserving reduced model.

In [140] the model reduction technique, based on the computation of a subspace of the product of the Gramians using the formulation proposed in [25] with a modification to allow for the retention of lightly damped and unstable modes in the reduced system, is proposed. The reduced system exhibits modal characteristics similar to the unreduced system in the frequency range associated with swing modes and can be used for the design of damping controllers.

In [141] a technique, called sparse LRCF-ADI, is used to construct reduced-order state-space models of large scale power systems described in descriptor form (9). The proposed technique is an improvement of the LR(CF)-
ADI (see Subsection 3.5) and exploits the sparsity of the power system model described in descriptor form.

The external system is important only as far as it influences the analysis in the study system and is often represented by a linear model for studies such as small signal stability analysis. Furthermore, it is often the case that the external system input/output behaviour is of interest only in very low frequencies (less than 2 Hz) depending on the nature of its interconnection to the study system and the level of generator modeling. This characteristic makes KMs (see Subsection 3.4) suitable for model reduction application on the external area. In [5] the use of KMs in the model reduction of power systems is described. Additionally, a connection between the KMs and coherency in power systems is proposed, aiming at retaining some physical relationship between the reduced and the original system. The reduction process considered the external area of the power system as an input-output system.

6 Application

In this section, two of the most frequently used techniques for the reduction of linear models (BT and MT) are applied to the New England 10-generator, 39-bus test power system, shown by single-line diagram in Fig. 1. System parameters and data are given in [142].
After the load flow calculation and data initialization, the test system described by DAEs (2) is linearized at the operating point (see eqs. (3) and (4)). Algebraic variables (i.e. vector of perturbations of algebraic variables) are eliminated, as in (5) and (6). The reference voltage of the AVR connected to the synchronous machine G1, and voltage magnitudes at buses 1, 5, 9, 14 and 17 (denoted in red in Fig. 1), are chosen to be input and output variables, respectively. Thus, the model used for the application of the two chosen techniques, is described by (7).

Two techniques applied to the example (linear system described by (7)) are briefly described in formulaes in Subsections 6.1 and 6.2.

### 6.1 Balanced Truncation (BT)

BT consists of three main steps [141].

The first step is to compute the Gramians $P$ and $Q$ by solving the Lyapunov equations:

$$AP + PA^T = -BB^T; \quad A^TQ + QA = -C^TC.$$  \hspace{1cm} (11)

The second step is to factor $P$ and $Q$ as $P = UU^T$ and $Q = LL^T$, where $U$ and $L$ are the Cholesky factors. These factors are then multiplied and their product decomposed as follows:

$$U^TL = W \Sigma Y^T = [W_1 \quad W_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} Y_1^T \\ Y_2^T \end{bmatrix},$$  \hspace{1cm} (12)

where $W_1$ and $Y_1$ are composed of the leading $r$ columns of $W$ and $Y$, respectively. Matrix $\Sigma = \text{diag} \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ contains the singular values of $U^TL$, that are also known as the HSVs of the system.

The third step is to construct a reduced-order model (in the form of (7)) of order $r \ll n$ via matrices:

$$A_1 = T_L^TA_R, \quad B_1 = T_L^TB, \quad C_1 = CT_R, \quad D_1 = D,$$  \hspace{1cm} (13)

where $T_L$ and $T_R$ are given by:

$$T_L = LY_1\Sigma_1^{-1/2}, \quad T_R = UW_1\Sigma_1^{-1/2}.$$  \hspace{1cm} (14)

$Y_1$, $W_1$ and $\Sigma_1$ are chosen such that $r$ largest HSVs and $r$ corresponding state variables are preserved in the reduced model.

### 6.2 Modal Truncation (MT)

In MT transfer function (see (8)) of the system is decomposed:

$$G(s) = G_1(s) + G_2(s),$$  \hspace{1cm} (15)

so that $G_1(s)$ contains $r$ dominant modes, while $G_2(s)$ contains the other $n-r$ modes of the system. Dominant modes are those that predominantly affect the
behavior of the system. In this example, for dominant modes are chosen those with the smallest absolute value, i.e. modes with small real and imaginary (low magnitude and frequency) values. The MT truncate modes contained in $G_2(s)$, and the model corresponding to the transfer function $G_1(s)$ approximates the original system.

**Fig. 2** – Voltage magnitudes at buses 1, 5, 9, 14 and 17, when increasing the reference voltage of the AVR connected to the synchronous machine G1 by 5 %, for:
1) original model, 2) model reduced by BT, and 3) model reduced by MT.
Fig. 2 gives a comparative view of voltage magnitudes at buses 1, 5, 9, 14 and 17, when increasing the reference voltage of the AVR connected to the synchronous machine G1 by 5%, for original model, model reduced by BT and model reduced by MT. State variables (modes) were truncated as far as the form of voltage magnitudes at buses 1, 5, 9, 14 and 17 corresponded to those obtained with the original model. Of the 109 state variables, BT method truncated 106, while MT method truncated 77 state variables (modes). MT preserved 32 dominant modes of original model, while neglecting less important ones. After balancing the system (where the state variables lost their physical meaning), BT method truncated the states which are less important for input-output behaviour of the system. BT method did not preserve the steady state behaviour of the original model (see voltage magnitudes at buses 5, 14 and 17). As already mentioned, if it is important to maintain the steady state behaviour of the original model, residualization can provide better results than truncation.

7 Conclusion

The need for reduced models has been driven by different reasons: limited computational, accuracy and storage capabilities, need to gain insight into the true cause of the observable dynamics of a system, etc. The problem of model reduction can be viewed as a trade-off between complexity and misfit. A model of minimal complexity that approximates the original one (optimally) with a maximum allowed misfit is required. This paper gives a brief review of the model reduction techniques, commonly used in control theory, for large-scale dynamic systems. Advantages and disadvantages of described techniques were discussed. Techniques commonly used for reduction of the power system models are also described. Two of the described techniques are applied to the New England 10-generator, 39-bus test power system with discussion of obtained results.

Control theory provides lots of methods which can be used for reduction of different large-scale physical systems. These methods mostly have a rigorous mathematical basis and aim to approximate the input-output behaviour of the considered system. On the other hand, for power system model reduction, empirical methods are preferred, since they retain the physical structure of the model, while being simple to use. Connections between those two groups of methods are investigated in several papers. Our future future research direction is to apply the methods used in control theory to reduce the dynamic model of the large-scale power system, so that the state variables in the reduced model preserve their physical meaning.
8 References


Dynamic Model Reduction: An Overview of Available Techniques with Application...


Dynamic Model Reduction: An Overview of Available Techniques with Application...

S.D. Dukić, A.T. Sari


Dynamic Model Reduction: An Overview of Available Techniques with Application...


S.D. Đukić, A.T. Sari


Dynamic Model Reduction: An Overview of Available Techniques with Application...


