Verfahren zur Modellordnungsreduktion: Two - Step - Lanczos

Damian Belz, Mandy Domke, Marie Krause

***Zusammenfassung*—Das Übertragungsverhalten eines Bauelements wird in der Regel durch sehr große Systeme beschrieben. Diese sollen bei geringem Fehler stark verkleinert werden, um eine effizientere Berechnung zu ermöglichen. Im Folgenden wird dafür der Two-Step-Lanczos zur Modellordnungsreduktion vorgestellt. Es zeigt sich, dass, mit den aus der Finiten Integrations Technik gewonnenen Systemen, die Ergebnisse das ursprüngliche Verhalten gut wiedergeben. Krause**

***Index Terms*—Finite Integration Technik (FIT), Two-Step-Lanczos (TSL), Krylov-Unterräume, partielle Realisierung, Padé-Via-Lanczos (PVL)**

I. Einführung

UM große Modelle berechnen zu können, gibt es mehrere Möglichkeiten. Eine Variante ist es Hochleistungsrechner zu verwenden, was allerdings kostspielig ist. Eine Andere die Details im Modell zu vernachlässigen, was zu ungewollten Fehlern führen kann. Die in diesem Paper gezeigte Variante approximiert das originale, verlustfreie System ohne auf wichtige Details verzichten zu müssen. Dies ist mit der Modellordnungsreduktion, genauer mit dem Two-Step-Lanczos (TSL), möglich, der im Folgenden vorgestellt wird.

Dazu wird zunächst auf die Erstellung der Übertragungsfunktion (ÜF) mit der Finiten Integration Technik (FIT) eingegangen. Anschließend wird der TSL bzgl. des Problems vorgestellt. Im nächsten Abschnitt wird dieses Verfahren auf elektromagnetische Bauelemente angewendet. Der Vergleich der reduzierten ÜF mit der originalen ÜF erfolgt danach. Zum Schluss wird ein Fazit gegeben. Domke

II. Aufstellen der Impedanzfunktion

Die vorzustellende Methode zur Reduzierung der Modellordnung setzt die Beschreibung des Übertragungsverhaltens im Frequenzbereich voraus. Durch die Diskretisierung der Struktur mit Hilfe der FIT (vgl. [1, S. 5ff.]) erhält man ein in der Frequenz lineares System. Die Ordnung beträgt $n = n_{\widehat{\mathrm{e}}} + n_{\widehat{\mathrm{h}}}$ mit allen Einträgen aus den Vektoren der elektrischen und magnetischen Kantenspannungen $\widehat{\mathrm{e}}$ bzw. $\widehat{\mathrm{h}}$ als Unbekannte. Das dual-orthogonale Gitter wird mit Hilfe von CST STUDIO SUITE ®, [5], erstellt. Die Materialmatrizen M_ε, M_μ und M_κ können über die CSTResultreader.dll mittels einer Funktion aus der Matlab-Bibliothek m2m ausgelesen werden, der Rotationsoperator C wird in Matlab 2016b, [4], erstellt. Durch das Kombinieren der Materialgleichungen mit den Gitter-Maxwellgleichungen erhält man

$$M_\varepsilon \frac{d}{dt}\widehat{\mathrm{e}} = C^T \widehat{\mathrm{h}} - M_\kappa \widehat{\mathrm{e}} - \widehat{\widehat{\mathrm{j}}}_s \; und \tag{1}$$

$$M_\mu \frac{d}{dt}\widehat{\mathrm{h}} = -C\,\widehat{\mathrm{e}}. \tag{2}$$

Der Quellenstrom $\widehat{\widehat{\mathrm{j}}}_s$ soll aus dem Vektor der einzelnen Portströme i erzeugt werden. Dazu wird eine Einkoppelmatrix R erstellt. Die Einträge in R werden so gewählt, dass die Anregung durch $\widehat{\widehat{\mathrm{j}}}_s$ zu dem zweidimensionalen H-Feld der entsprechenden Portmode H_{2D} korrespondiert. Der Vektor der äquivalenten Portspannungen u kann über eine Auskoppelmatrix L aus den elektrischen Kantenspannungen $\widehat{\mathrm{e}}$ extrahiert werden.
Diese ergibt sich aus der Beziehung $E_{2D} \times H_{2D} = 1W$. Bei geeigneter Normierung gilt somit $L = R^T$ und es ergibt sich

$$-\widehat{\widehat{\mathrm{j}}}_s = R\,i, \quad u = R^T\,\widehat{\mathrm{e}}\,.$$

Durch Elimination von $\widehat{\mathrm{h}}$ in (1) kann ein alternatives System aufgestellt werden. Das sogenannte Curl-Curl-System ist nicht mehr linear in der Frequenz, enthält jedoch nur noch $n_{\widehat{\mathrm{e}}}$ Unbekannte. Da in dieser Arbeit nur der verlustfreie Fall mit $M_\kappa = 0$ betrachtet wird, ergibt sich nach der Transformation in den Frequenzbereich

$$(j\omega)^2 M_\varepsilon \widehat{\mathrm{e}} = -C^T M_\mu^{-1} C\,\widehat{\mathrm{e}} + j\omega\, R\, i\,,$$
$$u = R^T \widehat{\mathrm{e}}\,.$$

Durch Einführen der Variablen $x = M_\varepsilon^{1/2}\widehat{\mathrm{e}}$ wird das System symmetrisiert. Für das weitere Vorgehen werden die Produkte der konstanten Matrizen in

$$(j\omega)^2 x = -\underbrace{M_\varepsilon^{-1/2} C^T M_\mu^{-1} C M_\varepsilon^{-1/2}}_{A}\, x + j\omega \underbrace{M_\varepsilon^{-1/2} R}_{B} i,$$
$$u = B^T x$$

zu A und B zusammengefasst. Mit $s = j\omega$ ergibt sich die Impedanzmatrix aus dem Curl-Curl-System über die Beziehung $u = Z(s)\,i$ zu

$$Z(s) = sB^T(s^2 I + A)^{-1} B\,.$$

Belz

III. Two-Step-Lanczos

Beim TSL handelt es sich um ein zweistufiges Verfahren zur Modellordnungsreduktion. In der ersten Stufe wird die partielle Realisierung durchgeführt. Anschließend kommt der Padé-via-Lanczos (PVL)-Algorithmus zum Einsatz. Dabei

Danke an Prof. Dr.-Ing. Schuhmann und unseren Betreuer Philipp Jorkowski (wiss. Mitarbeiter) von der TU Berlin, Fakultät IV: Elektrotechnik und Informatik, Fachgebiet Theoretische Elektrotechnik, für die gute Zusammenarbeit.

sorgt die partielle Realisierung für eine Verringerung des Systems auf mittlere Größe. Die anschließende Anwendung des PVL-Algorithmus reduziert das System noch weiter. Eine grobe Abschätzung, um wie viel das System jeweils reduziert wird, ist Abb. 1 zu entnehmen.
In den folgenden Abschnitten wird der TSL aus [1, S. 88] schrittweise erläutert. Domke

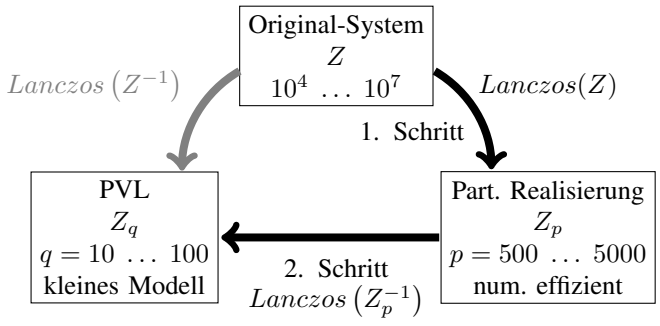

Abbildung 1. Ablauf des TSL [2] (angepasst)

A. Partielle Realisierung

Im ersten Schritt wird die ÜF umformuliert zu

$$Z(s) = sB^T(s^2I + A)^{-1}B = \sum_{k=0}^{\infty} B^T(-A)^k B\frac{1}{s^k},$$

um dann mit einem geeigneten Verfahren die Summe zu approximieren. Die partielle Realisierung (PR) soll dabei eine Ordnungsreduktion des Problems bewirken und grundlegende Eigenschaften des Original-Systems erhalten. Für Matrizen bedeutet dies den Erhalt dominanter Eigenwerte. Es wird also ein Verfahren gesucht, welches eine Matrix T_p findet, die ähnliche Eigenschaften wie die System-Matrix A hat und von geringerer Ordnung ist.

Diese Bedingungen erfüllen Krylov-Unterraum (KU) Verfahren. Für die vorliegenden Probleme werden in der Regel m-Port-Systeme, $m \geq 2$, betrachtet. Es werden demnach die Block-Krylov-Unterräume benötigt, welche definiert sind durch

$$\mathcal{K}_{pm}(A, B) \coloneqq span\left\{B, AB, ..., A^{p-1}B\right\}, \tag{3}$$

wobei $p \ll n$ und m die Anzahl der Spalten von B ist. In der Praxis relevante Verfahren sind der Arnoldi-Algorithmus, der eine symmetrisch, positiv definite Matrix A voraussetzt, der Band-Lanczos-Algorithmus, der ein symmetrisches A benötigt und der Bi-Lanczos-Algorithmus, welcher keine Bedingungen an A stellt (vgl. [1, S. 52ff.]).

Unter Beachtung der Gegebenheiten, die FIT mit sich bringt, ist bekannt, dass die System-Matrix A immer symmetrisch, aber nicht notwendiger Weise positiv definit, ist. Für die PR ist hier also der Band-Lanczos am sinnvollsten. Ein Pseudocode für dieses Verfahren, mit einer kleinen Korrektur (Z.6: $A_i v_{i-m} \rightarrow A v_{i-m}$), findet sich in [1, S. 54].

Bei der PR wird das KU Verfahren der Wahl p-mal für die System-Matrix A und die Port-Matrix B aufgerufen. In jedem Schritt wird der Rang der A-approximierenden Matrix

Input: A, B
  $p \leftarrow m$
  while $error \geq tol$ **do**
    $[B_p, T_p, V_p, \hat{V}] \leftarrow BandLanczos(A, B, p)$
    $error \leftarrow \sum_{k=1}^{p} \|\lambda_{k,p+\Delta p} - \lambda_{k,p}\| \,/\, \|\lambda_{k,p}\|$
    $p \leftarrow p + \Delta p$
  end while
Output: T_p, B_p, V_p

Abbildung 2. partielle Realisierung - Algorithmus

T_p erhöht und B an die Basis des KU angepasst, man erhält B_p. Nach Δp Schritten wird überprüft, ob die Approximation gut genug ist oder ob der Rang von T_p weiter erhöht werden muss.

Auf die Wahl eines geeigneten Abbruch-Kriteriums wird in III-C eingegangen.

Nun kann das Verfahren der PR zusammengefügt werden, welches in Abb. 2 dargestellt ist. Die Original-ÜF wurde nun mittels einer KU-Basis zu

$$Z_p = sB_p^T(s^2I_p + T_p)^{-1}B_p \tag{4}$$

transformiert. Krause

B. Padé - via - Lanczos

Wie im vorherigen Abschnitt beschrieben, erhält man das reduzierte System der Ordnung p aus (4). Dieses lässt sich mit einer Padé-Approximation im Entwicklungspunkt

$$s_0 = j\left(\frac{\omega_{max} - \omega_{min}}{2} + \omega_{min}\right)$$

auch in der Form

$$Z_p(s) \approx s\sum_{k=0}^{2q} \underbrace{B_p^T\left(\hat{T}_p\right)^k \hat{B}_p}_{M_k} \underbrace{\left(s^2 - s_0^2\right)^k}_{\sigma^k} = s\sum_{k=0}^{2q} M_k\sigma^k \tag{5}$$

schreiben, wobei $\hat{T}_p = \left(T_p + s_0^2 I_p\right)^{-1}$ und $\hat{B}_p = \hat{T}_p B_p$ gilt. Dabei stimmen die Koeffizienten $M_k, k = 0, \ldots, 2q$ mit den entsprechenden ersten $2q + 1$ Koeffizienten der Taylor-Approximation von Z_p überein. Um das direkte Bilden einer Inversen zu vermeiden, wird $\hat{T}_p$ mit der LU-Zerlegung in Dreiecksmatrizen zerlegt. Dies ist bei großen Systemen nur durch vorherige Anwendung der PR möglich. Ferner ist zu beachten, dass $\hat{T}_p$ lediglich von s_0, also der Mittenfrequenz, abhängt. Das heißt, die Berechnung der Inversen findet an nur einem Frequenzpunkt statt, was die Rechenzeit ebenfalls deutlich verkürzt.
Betrachtet man folgende aus (5) resultierende Darstellung

$$Z_q(s) = sB_p^T\left(I_p + \hat{T}_p\sigma + \hat{T}_p^2\sigma^2 + \cdots + \hat{T}_p^{2q}\sigma^{2q}\right)\hat{B}_p,$$

so ist auffällig, dass sie eine ähnliche Struktur wie die KU in Gleichung (3) aufweist. Daher ist für die weitere Berechnung von Z_q eine ähnliche Vorgehensweise zur PR angebracht um damit das System weiter zu reduzieren. Während allerdings in der vorherigen Variante das Abbruch-Kriterium alle Δp-Schritte überprüft wurde, findet dies nun in jeder Iteration statt,

wodurch das System stärker reduziert wird. Damit funktioniert der PVL-Algorithmus wie in Abb. 3 dargestellt. Dabei werden, wie in Abschnitt III-A, die Matrizen T_q, B_q und V_q ausgegeben. Das mit diesem Algorithmus reduzierte System wird nun durch

$$Z_q(s) = s\ B_p^T V_q \left(s^2 I + V_q^T T_p V_q\right)^{-1} V_q^T B_p \qquad (6)$$

beschrieben und hat jetzt eine Ordnung von $q \ll p$.

Domke

Input: s_0, T_p, B_p
// Padé - Approximation
$p \leftarrow size\,(T_p,\, 1)$
$\hat{T}_p \leftarrow -inv\left(T_p + s_0^2 I_p\right)$
$\hat{B}_p \leftarrow \hat{T}_p B_p$
// modifizierte partielle Realisierung
$q \leftarrow 1$
berechne Eigenwerte λ_p von T_p
while $error \geq tol$ **do**
$[B_q, T_q, V_q, \hat{V}_q] \leftarrow BandLanczos(\hat{T}_p, \hat{B}_p, q)$
$q \leftarrow q + 1$
berechne Eigenwerte λ_q von T_q
$\lambda_q \leftarrow -1/\lambda_q$
$error \leftarrow \sum_{k=1}^{p} \|\lambda_{k,p} - \lambda_{k,q}\| \,/\, \|\lambda_{k,p}\|$
end while
Output: T_q, B_q, V_q

Abbildung 3. PVL - Algorithmus

C. Abbruch-Kriterien

Die Wahl eines geeigneten Abbruch-Kriteriums ist entscheidend für den TSL. Eine Möglichkeit besteht darin den relativen Fehler der reduzierten ÜF Z_q zur Original-ÜF Z, an ausgewählten Stellen s_i, $i = 1, ..., N$, zu vergleichen. Da die Original-ÜF in der Regel nicht bekannt ist und es passieren kann, dass die s_i ungünstig gewählt sind, ist dieses Kriterium ungeeignet. Eine bessere Möglichkeit bietet der Vergleich der Eigenwerte. Für den relativen Fehler wird dann gefordert

$$\delta_{eig,i} = \frac{\|\lambda_{i,p+\Delta p} - \lambda_{i,p}\|}{\|\lambda_{i,p}\|} < \varepsilon_{eig}.$$

Bei dem Versuch das Kriterium anzuwenden stellt man fest, dass es nicht ohne Weiteres möglich ist, die approximierten Eigenwerten einander zuzuordnen. Denn T_p besitzt p Eigenwerte und $T_{p+\Delta p}$ bereits $(p + \Delta p)$. Ein Lösungsansatz ist, die Eigenwerte aus T_p denen aus $T_{p+\Delta p}$ zuzuordnen, die am nächsten liegen. Dieses Vorgehen ist anscheinend ungenau, da aber der Großteil der Eigenwerte dicht beieinander liegt, ist von einer richtigen Zuordnung auszugehen. Dieses Vorgehen hat leider zu einem zu frühen Abbruch geführt, sodass hier noch Zeit in die Fehlersuche investiert werden sollte. In [1, S. 62] wird als eine weitere Möglichkeit die Berechnung der Ritz-Vektoren $x_{i,p}$ zu den Eigenwerten von T_p angegeben. Es wird überprüft, ob

$$|\hat{V} x_{i,p}| < \varepsilon_{eig,p},$$

wobei das $\hat{V}$ aus dem Lanczos-Verfahren resultiert. Ein Verfahren zur Bestimmung der Ritz-Paare lässt sich [3, S. 28] entnehmen. Dieses Kriterium ist jedoch beim Testen durchgefallen, da sich, unabhängig von p, nur eine Genauigkeit von 10^{-2} erreichen ließ. An dieser Stelle sollte noch Zeit investiert werden, um mögliche Fehler zu finden. Krause

D. Two - Step - Lanczos

Als Hintereinander-Ausführung der PR und des PVL vereint der TSL die Vorteile beider Verfahren und vermeidet deren Nachteile, sodass eine starke Reduzierung des Systems bei geringer Laufzeit erreicht werden kann. Das Verfahren der PR ist numerisch schnell, reduziert das System jedoch nur auf mittlere Größe. Anders ist es bei der PVL-Algorithmus. Hier wird das System stark reduziert, allerdings ist diese Methode numerisch langsam.
Insgesamt läuft der TSL-Algorithmus wie in Abb. 4 ab.

Domke

Input: -
// System
Lade System-Matrizen A, B
$s_0 \leftarrow (\omega_{max} - \omega_{min})\,/2 + \omega_{min}$
// TSL
$[T_p, B_p, V_p] \leftarrow partReal\,(A, B)$
$[T_q, B_q, V_q] \leftarrow PVL\,(s_0, T_p, B_p)$
// reduziertes System
$Z_q \leftarrow s\ B_p^T V_q \left(s^2 I + V_q^T T_p V_q\right)^{-1} V_q^T\ B_p$
Output: Z_q

Abbildung 4. TSL - Algorithmus

IV. Anwendung und Vergleich

Im Folgenden wird der TSL auf zwei Filter angewendet und sowohl für die Original-ÜF als auch für die Ergebnisse des TSL die S-Parameter für 200 Frequenzpunkte berechnet. Bis auf die Diskretisierung der Anordnungen wurden sämtliche Berechnungen mit Matlab 2016b, [**?**], [4] durchgeführt. Verglichen werden die Laufzeiten sowie die erhaltenen Übertragungsfunktionen. Die Angaben zur Laufzeit beziehen sich auf einen Rechner mit einem Intel(R) Core i5 6700-Prozessor mit $4 \times 3.50\,GHz$ Taktfrequenz und $32\,GB$ RAM.

Belz

A. Schmalband- und Langer-Filter

Das Übertragungsverhalten für den in Abb. 5 gezeigten symmetrischen Schmalband-Filter wurde für die Frequenzen von 0.5 bis $2\,GHz$ bestimmt. Die Struktur wird über zwei koaxiale Ports angeregt. Mit einer Abtastung von mindestens 13 Gitterlinien pro Wellenlänge ergibt sich ein Curl-Curl-System mit $63\,756$ Unbekannten.
Beim Langer-Filter, s. Abb. 6, wird das Resonanzverhalten durch zwei dielektrische Ringe mit $\varepsilon_r = 38$ gesteuert. Die Anregung erfolgt erneut über koaxiale Ports und auch die Abtastung ändert sich nicht. Für Frequenzen von 3.5 bis $6.5\,GHz$ entsteht ein Curl-Curl System mit $222\,264$ Unbekannten.

Belz

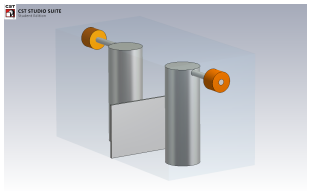

Abbildung 5. Schmalbandfilter

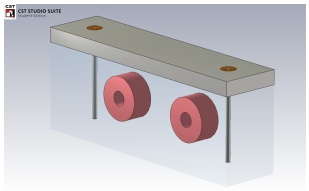

Abbildung 6. Langer-Filter

B. Auswertung

Betrachtet man die Werte in Tabelle I, so lässt sich feststellen, dass in beiden Fällen durch die PR Systeme der Größenordnung 10^3 erzeugt werden. Der vollständige TSL-Algorithmus reduziert die Größenordnung auf 10^1. Auch diee Laufzeit für die Berechnung des Systems Z wird um mehr als Faktor 10 verringert.

Tabelle I
SYSTEM-ORDNUNGEN UND LAUFZEITEN FÜR DEN SCHMALBAND- UND DEN LANGER-FILTER

	Schmalbandfilter		Langer-Filter	
	Ordnung	Laufzeit [s]	Ordnung	Laufzeit [s]
Orginal-System	63 756	240.82	222 264	6502.11
TSL-Algorithmus	20	21.15	22	230.77
Part. Realisierung	962	19.55	1362	228.97
PVL	20	0.33	22	0.72

Wie in Abb. 7 zu sehen ist, steigt die Laufzeit linear mit der Dimension der Curl-Curl-Systemmatrix A. Zudem ist eine Abhängigkeit der Laufzeit des TSL von der Anzahl der Iterationen p der PR erkennbar. Die Anwendung des PVL-Algorithmus auf das System mittlerer Ordnung hat kaum Einfluss auf die Gesamtlaufzeit, da durch die LU-Zerlegung die Inverse leicht gebildet werden kann und hier nur wenige Iterationen durchlaufen werden.

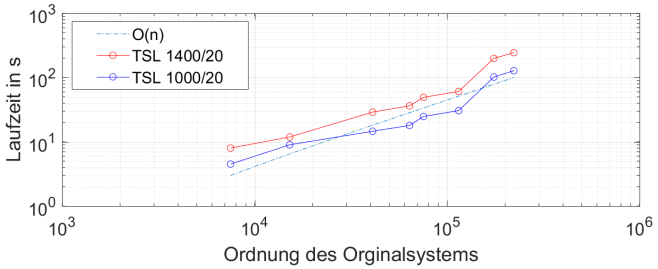

Abbildung 7. Laufzeit-Verhalten des TSL für $p = 1000; q = 20$ sowie $p = 1400; q = 20$ (doppelt-logarithmisch)

Das Übertragungsverhalten, s. Abb. 8 und 9, wird in beiden Fällen durch die PR fast perfekt wiedergegeben. Erst durch die drastische Reduzierung mit dem PVL entstehen grobe Approximationsfehler. Denn bei der drastischen Reduzierung bleiben nur wenige dominante Eigenwerte erhalten, sodass der Einfluss der weggefallenen Eigenwerte eventuell überwiegen kann. Eine höhere Gewichtung der Eigenwerte innerhalb des betrachteten Frequenzbandes kann möglicherweise

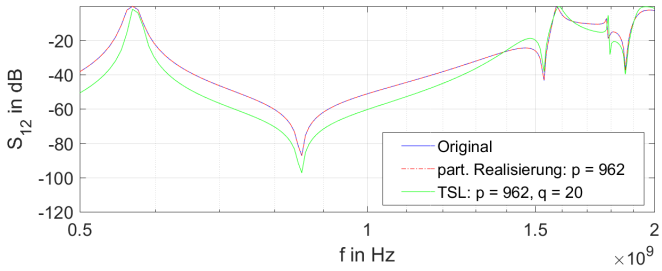

Abbildung 8. S-Parameter der originalen und mit dem TSL reduzierten ÜF für den Schmalbandfilter

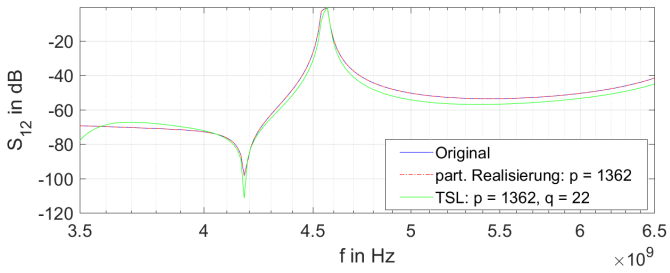

Abbildung 9. S-Parameter der originalen und mit dem TSL reduzierten ÜF für den Langer-Filter

zu einer genaueren Abbildung des Übertragungsverhaltens führen. Ohne diese Ergänzung wird eine Verbesserung der Abbruch-Kriterien kaum bessere Resultate erzeugen. Obgleich Abweichungen zum Teil deutlich sichtbar sind, ist das grobe Übertragungsverhalten klar erkennbar.

Belz, Domke, Krause

V. FAZIT

In diesem Paper wird der TSL als ein Verfahren vorgestellt, das große Systeme stark reduzieren kann ohne das Übertragungsverhalten wesentlich zu verändern. Als „System“ dient hierbei die ÜF verlustfreier elektromagnetischer Bauelemente, die mit FIT-Matrizen ermittelt wurde. Um die Reduzierung zu erzielen, arbeitet der TSL-Algorithmus in zwei Schritten. Im ersten Schritt wird mit der PR das System auf mittlere Größe gebracht und im zweiten Schritt mit der PVL-Methode noch einmal stark reduziert. Die betrachteten Abbruch-Kriterien waren in der Anwendung wenig zielführend, sodass eine weitere Betrachtung dessen nötig ist. Auch ist es sinnvoll weitere Beispiel-Anordnungen zu testen.
Insgesamt ist nachgewiesen, dass es mit dem TSL-Algorithmus möglich ist, sehr große Systeme effizient und immer noch recht präzise zu simulieren.

Domke

LITERATUR

[1] T. Wittig, *Zur Reduzierung der Modellordnung in elektromagnetischen Feldsimulationen* Göttingen, Deutschland: Cuvillier Verlag
[2] T. Wittig, *Efficient Model Order Reduction Based on a Two-Step-Lanzcos Approach* CST GmbH Darmstadt, Deutschland
[3] W. Vogt, *Zur Numerik großdimensionaler Eigenwertprobleme* TU Ilmenau, Deutschland
[4] MATLAB 2016b, The MathWorks® Inc. Natick, Massachusetts, USA
[5] CST STUDIO SUITE 2016®, CST AG, Darmstadt, Deutschland