Elektrische Antriebe - Regelung von Antriebssystemen

Bearbeitet von Dierk Schröder

4. Auflage 2015. Buch. XXXII, 1879 S. Hardcover ISBN 978 3 642 30095 0 Format (B x L): 16,8 x 24 cm

<u>Weitere Fachgebiete > Technik > Energietechnik, Elektrotechnik > Elektromotoren</u>

Zu Inhaltsverzeichnis

schnell und portofrei erhältlich bei



Die Online-Fachbuchhandlung beck-shop.de ist spezialisiert auf Fachbücher, insbesondere Recht, Steuern und Wirtschaft. Im Sortiment finden Sie alle Medien (Bücher, Zeitschriften, CDs, eBooks, etc.) aller Verlage. Ergänzt wird das Programm durch Services wie Neuerscheinungsdienst oder Zusammenstellungen von Büchern zu Sonderpreisen. Der Shop führt mehr als 8 Millionen Produkte.

# 26 POD zur Optimalsteuerung linearer partieller Differentialgleichungen

Prof. S. Volkwein, Konstanz

# 26.1 Einleitung

Viele Vorgänge in den Naturwissenschaften, in den Ingenieurwissenschaften, aber auch in den Wirtschaftswissenschaften und in medizinischen Anwendungen lassen sich durch Systeme von partiellen Differentialgleichungen beschreiben. Diese Differentialgleichungssysteme sind in der Regel nichtlinear, gekoppelt und beinhalten oft sehr viele Parameter(-funktionen), die geeignet zu wählen sind, um zu garantieren, dass die realen Prozesse hinreichend genau nachgebildet werden. Da meistens nicht alle Parameter bekannt oder messbar sind, werden oft Methoden der Parameterschätzung eingesetzt. Diese Verfahren benötigen in der Regel viele Simulationen des gegebenen Differentialgleichungssystems. In anderen Anwendungen möchte man die Systeme partieller Differentialgleichungen durch eine Steuerung so beeinflussen, dass ein gewünschtes Verhalten erzeugt wird. Dieses führt auf ein sogenanntes Optimalsteuerungsproblem, dass das Lösen von einer nichtlinearen Optimierungsaufgabe erfordert, bei der das Differentialgleichungssystem als Nebenbedingung auftritt. Will man noch Störungen in den Daten beziehungsweise im Modell berücksichtigen, so sind Rückkopplungs- oder Feedbacksteuerungen zu berechnen, was das Lösen mehrerer Optimierungsprobleme erfordert. Alle diese Aufgabenstellungen führen bei der Verwendung klassischer Diskretisierungsverfahren (wie zum Beispiel die Finite-Elemente-, Finite-Differenzen- oder Finite-Volumen-Methode) auf sehr hochdimensionale, nichtlineare diskrete Probleme, deren numerische Behandlung sehr aufwendig beziehungsweise derzeit noch nicht möglich ist (als Beispiel sei hier die Feedbacksteuerung mit der Hamilton-Jacobi-Bellman-Gleichung genannt). Aus diesem Grund sind unterschiedliche Techniken der Modellreduktion entwickelt worden, um hochdimensionale Optimalsteuerprobleme durch niedrigdimensionale zu approximieren, die mit weniger Aufwand (oder überhaupt) numerisch lösbar sind. Heute existiert eine Reihe erfolgreicher Techniken der Modellreduktion wie Balanced Truncation, Moment Matching, Reduced-Basis-Methode oder Proper Orthogonal Decomposition (POD), die im Weiteren kurz beschrieben werden.

D. Schröder, Elektrische Antriebe – Regelung von Antriebssystemen, DOI 10.1007/978-3-642-30096-7\_26

Am flexibelsten ist wahrscheinlich POD, eine in der Praxis sehr häufig angewendete Methode. POD kann man auch bei nichtlinearen Gleichungen bzw. bei linearen Evolutionsgleichungen mit zeitabhängigen Koeffizienten einsetzen, während Balanced Truncation und Moment Matching derzeit mehr oder weniger nur lineare Probleme mit konstanten Koeffizienten zulassen. Andererseits existieren zu den zuerst genannten Methoden zuverlässige A-priori-Fehlerschätzer, während POD unter dem Ruf leidet, wegen der am Anfang zur Gewinnung von Snapshots willkürlich gewählten Steuerfunktion von heuristischer Natur zu sein. Dieser Mangel an zugehöriger Analysis ist ein Nachteil für die Akzeptanz von POD unter Mathematikern, während Ingenieure oft davon Gebrauch machen. Aufgrund neuerer Ergebnisse [1347, 1383, 1387, 1390] ist es aber möglich, mit A-posteriori-Abschätzungen die Qualität berechneter suboptimaler Steuerungen abzuschätzen.

Im Folgenden werden die genannten Techniken kurz beschrieben, wobei nur wenige der zahlreichen zur Thematik erschienenen Arbeiten zitiert werden können.

Eine oft verwendete Methode zur Reduktion hochdimensionaler Systeme ist das *Moment-Matching-Verfahren*, siehe zum Beispiel in [1322, 1324]. Bei dieser Technik wird das dynamische System auf Krylov-Unterräume projiziert, wobei die Krylov-Unterräume durch Arnoldi- oder Lanczos-Verfahren bestimmt werden. Da hierbei nur Matrix-Vektor-Multiplikationen notwendig sind, ist diese Modellreduktion sehr effizient bei hochdimensionalen, dünnbesetzten Systemen, siehe zum Beispiel [1321]. Nachteil der Moment-Matching-Methode ist allerdings, dass Stabilität und Passivität des dynamischen Systems bei der Modellreduktion verloren gehen können. Ferner gibt es derzeit keine globalen Fehlerschätzer[1302, 1332].

Eine weitere, sehr gut analysierte Methode ist *Balanced Truncation* [1326, 1394]. Dieses Verfahren verwendet die Lösungen zweier Lyapunov-Gleichungen, die sogenannten Controllability und Observability Gramians. Bei Balanced Truncation wird das dynamische System in eine balanzierte Form gebracht, so dass die genannten Gramschen Matrizen von Diagonalgestalt und identisch sind. Zustände, die schwer anzusteuern oder zu beobachten sind, werden vernachlässigt. Der Vorteil des Verfahrens ist, dass beim reduzierten dynamischen System die asymptotische Stabilität erhalten bleibt. Ferner gibt es A-priori-Fehlerschätzer. Neulich ist Balanced Truncation auch auf Deskriptorsysteme erweitert worden, siehe zum Beispiel in [1364, 1384] und [1334].

Sowohl Moment Matching als auch Balanced Truncation verwenden keine Snapshots, die mehr oder weniger beliebig gewählt werden müssen. Einen Überblick über Verfahren der Modellreduktion findet man beispielsweise in [1296, 1378]. Allerdings ist derzeit die Anwendung beider Methoden auf lineare, zeitinvariante dynamische Systeme beschränkt. Für zeitvariante oder nichtlineare Probleme sind beide Techniken nicht direkt anwendbar. Es gibt Ansätze, zeitvariante Modelle durch stückweise zeitinvariante zu approximieren, aber diese befinden sich noch in einer Anfangsphase, siehe beispielsweise [1304]. Allerdings sind die Fehlerschätzer für *beliebige* anwendbare Steuerfunktionen einsetzbar. Die beiden Methoden sind also nicht fokussiert auf einen Optimierungsaspekt, sondern weitreichender angelegt.

Die Anwendung von *Reduced-Basis-Methoden* auf zeitvariante nichtlineare Systeme, insbesondere Optimalsteuerprobleme gewinnt derzeit ein zunehmendes Interesse [1316, 1327, 1365]. Diese Verfahren basieren auf einer Projektion des dynamischen Systems auf Unterräume, die von Basisfunktionen aufgespannt werden, die selbst Lösungen des dynamischen Systems zu gewissen Zeitpunkten und/oder zu gewissen Parameterwerten sind. Der wesentliche Unterschied zu Finite-Element-Verfahren ist, dass die Finite-Elemente-Basisfunktionen keine Informationen über die Charakteristiken und Eigenschaften des zu approximierenden dynamischen Systems enthalten. Die Basisfunktionen von Reduced-Basis-Methoden, wie sie etwa in [1328, 1367] und [1309, 1333, 1345] entwickelt worden, enthalten Informationen des dynamischen und/oder parameterabhängigen Systems unter Einfluss von erwarteten Steuerungen.

POD ist die zur Zeit wohl am meisten verwendete Modellreduktionsmethode für zeitvariante oder nichtlineare Systeme, wobei sich die zugehörigen Basisfunktionen aus Lösungen des dynamischen Systems zu gewissen Zeitpunkten und/oder Parameterwerten ergeben. Diese Lösungen werden Snapshots genannt. Aufgrund der möglichen linearen (oder beinahe linearen) Abhängigkeit der Lösungen ist die Verwendung der Snapshots als Basisfunktionen eventuell ungeeignet. Aus diesem Grund wird eine Singulärwertzerlegung einer gewissen Korrelationsmatrix durchgeführt und die zugehörigen verallgemeinerten Eigenfunktionen werden als Basis verwendet. Diese Basis wird POD-Basis genannt. POD wird in vielen Gebieten erfolgreich eingesetzt, z.B. in der Signalanalysis und Mustererkennung [1323], Fluiddynamik und kohärente Strukturen [1294, 1298, 1338, 1343, 1361, 1381] und bei inversen Problemen [1303, 1356]. Ferner wurde POD dazu eingesetzt, Steuerungen reduzierter Ordnung zu berechnen [1300]. Die Beziehung zwischen POD und Balanced Truncation wurde in [1355, 1373, 1393] untersucht. Eine Fehleranalyse für nichtlineare, n-dimensionale dynamische Systeme (ohne Berücksichtigung von für die Optimalsteuerung wichtigen Aspekten der Optimierung) wurde in [1369] durchgeführt. Weiter wurde in [1299] die sogenannte Missing-Point-Technik für Modellreduktionsverfahren entwickelt, um nichtlineare Terme bei n-dimensionalen dynamischen Systemen effizient auszuwerten.

Im Unterschied zu Balanced Truncation besitzt die POD-Methode keine numerisch verwendbaren A-priori-Fehlerschranken. Mit [1341] und [1331, Theorem 4.17] liegen zwar Resultate für eine A-Priori-Analysis vor, aber eine Konvergenzrate kann nur unter dem Verwenden der exakten (unbekannten) optimalen Steuerung hergeleitet werden. Auf Grund einer gewissen Willkür bei der Auswahl der Steuerung zur Berechnung der Snapshots verwundert das auch nicht. Werden die Snapshots nicht aus einem möglichst reichhaltigen Zustandsraum gewählt, dann kann der tatsächliche Fehler zwischen der POD-Lösung und der exakten Lösung der Differentialgleichung sehr groß sein. Auf der anderen Seite hat die POD-Methode den bereits erwähnten Vorteil, dass sie direkt auf zeitvariante nichtlineare Probleme anwendbar ist. Sind die Snapshots gut gewählt, so besitzt der von der POD-Basis aufgespannte Unterraum relevante Informationen über das System. Das und die direkte Anwendbarkeit der POD-Methode machen diese Technik der Modellreduktion so attraktiv für die praktische Anwendung, obwohl dem Verfahren ein gewisser heuristischer Ansatz innewohnt. Aus diesem Grund sind Schritte zur besseren Kontrolle des auftretenden Fehlers von großem Interesse, insbesondere für die Anwender.

In [1347, 1387] wird eine Fehleranalysis für Optimalsteuerungsprobleme mit linearen beziehungsweise nichtlinearen partiellen Differentialgleichungen und quadratischem Zielfunktional hergeleitet. Hauptresultat ist eine A-posteriori-Fehlerabschätzung aus Sicht der Optimalität: Der Abstand einer suboptimalen, auf der Basis des POD-Modells berechneten Steuerung zur optimalen (exakten und unbekannten) Steuerung kann abgeschätzt werden. Numerische Beispiele zeigen, dass diese Strategie überaus effizient sein kann. Es ist auf diese Weise möglich, das Fehlen eines A-priori-Schätzers für POD-Approximationen zu kompensieren. Außerdem kann die Zahl der verwendeten Snapshots auf der Basis des geschätzten Fehlers vergrößert oder verkleinert werden. Zur Berechnung der Fehlerschranke sind jeweils einmal die unreduzierte Zustandsgleichung und die unreduzierte adjungierte Gleichung zu berechnen.

Der entscheidende erste Gewinn dieser Herangehensweise besteht darin, dass Probleme mit linearen zeitabhängigen Koeffizienten behandelt werden können, wozu die anderen verfügbaren Methoden derzeit noch nicht in der Lage sind. Außerdem wurde so die Analysis des POD-Verfahrens ein Stück weiter gebracht.

In diesem Beitrag konzentrieren wir uns zur Vereinfachung der Darstellung auf linear-quadratische Optimalsteuerprobleme und geben an geeigneter Stelle Hinweise auf Erweiterungen, zu denen Arbeiten in der Literatur vorliegen. Wir gliedern diesen Artikel wie folgt: In Abschnitt 26.2 führen wir die POD-Methode anhand der Wärmeleitungsgleichung vor. Diese Vorgangsweise wird in Abschnitt 26.3 auf linear-quadratische Optimalsteuerprobleme angewendet. In Abschnitt 26.4 stellen wir ein paar Literaturhinweise auf numerische Beispiele zusammen.

# 26.2 POD am Beispiel der Wärmeleitungsgleichung

#### 26.2.1 Die schwache Formulierung

Worum geht es? In diesem Abschnitt führen wir für unsere Betrachtungen als Beispiel die lineare Wärmeleitungsgleichung ein. Wir sind an einer numerischen Lösung interessiert. Hierfür wenden wir *Galerkin-Methoden* an. Andere Verfahren wie das Finite-Differenzen-Verfahren oder die Finite-Volumen-Methode könnten natürlich auch verwendet werden. Darauf gehen wir hier aber nicht ein. Für die Galerkin-Verfahren benötigen wir die schwache oder variationelle Formulierung der Wärmeleitungsgleichung (vgl. [1314]). Auf der anderen Seite ermöglichen die Variationsformulierungen einen schwächeren Lösungsbegriff, der für viele Anwendungen ausreichend ist.

Sei  $\Omega \subset \mathbb{R}^d$ ,  $d \in \{1, 2, 3\}$ , eine offene, beschränkte Menge mit einem Lipschitzstetigen Rand  $\Gamma = \partial \Omega$  und Abschluss  $\overline{\Omega} = \Omega \cup \Gamma$ . Mit  $[t_\circ, t_f] \subset \mathbb{R}$  bezeichnen wir ein gegebenes Zeitintervall mit  $0 \leq t_\circ < t_f$ . Wir definieren den Raum-Zeit-Zylinder  $Q = (t_\circ, t_f) \times \Omega$  mit dem Rand  $\Sigma = (t_\circ, t_f) \times \Gamma$ . Weiter benötigen wir folgenden vollständigen Funktionenraum:

$$C^{2,1}(\overline{Q}) = \left\{ \varphi : \overline{Q} \to \mathbb{R} \mid \varphi, \varphi_t, \varphi_{x_i}, \varphi_{x_i x_j} \text{ sind stetig auf } \overline{Q}, \ 1 \le i, j \le d \right\}$$

wobei  $\varphi_t$  und  $\varphi_{x_i}$  die partiellen Ableitung von  $\varphi$  nach t beziehungsweise  $x_i$  bezeichnen. Ferner steht  $\varphi_{x_i x_j}$  für die zweite partielle Ableitungen von  $\varphi$  nach  $x_i$  und  $x_j$ . Der Ausdruck  $\nabla \varphi$  steht für den Gradienten von  $\varphi$ , das heißt, der Gradient  $\nabla \varphi(\boldsymbol{x})$  von  $\varphi$  an der Stelle  $\boldsymbol{x}$  bezeichnet den Zeilenvektor  $(\varphi_{x_1}(\boldsymbol{x}), \ldots, \varphi_{x_d}(\boldsymbol{x})) \in \mathbb{R}^{1 \times d}$ .

Als Zustandsgleichung betrachten wir die lineare Wärmeleitungsgleichung in der sogenannten starken Form: Gesucht ist eine Lösung  $y \in Y_1 = C^{2,1}(\overline{Q})$  mit

$$y_t(t, \boldsymbol{x}) - \operatorname{div} \left( \lambda(\boldsymbol{x}) \nabla y(t, \boldsymbol{x}) \right) = f(t, \boldsymbol{x}), \qquad (t, \boldsymbol{x}) \in Q \qquad (26.1a)$$

$$\frac{\partial y}{\partial n}(t, \boldsymbol{x}) + q(\boldsymbol{x})y(t, \boldsymbol{x}) = u(t, \boldsymbol{x}), \qquad (t, \boldsymbol{x}) \in \Sigma$$
 (26.1b)

$$y(t_{\circ}, \boldsymbol{x}) = y_{\circ}(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega$$
 (26.1c)

In (26.1a) bezeichnet  $\lambda = \lambda(\mathbf{x})$  einen örtlich verteilten, positiven Wärmeleitkoeffizient, das heißt, es gibt eine Konstante  $\lambda_a > 0$  mit  $\lambda(\mathbf{x}) \geq \lambda_a$  für alle  $\mathbf{x} \in \overline{\Omega}$ . Ferner bezeichnet  $f = f(t, \mathbf{x})$  eine verteilte Wärmequelle auf der rechten Seite. In (26.1b) steht  $\partial y / \partial n$  für die Normalableitung von y in Richtung der äußeren Normale n an den Rand  $\Gamma$  von  $\Omega$ . Wir nehmen weiter an, dass  $q = q(\mathbf{x})$  eine gegebene verteilte, nicht negative Funktion ist. Wir verzichten hier zur Vereinfachung der Präsentation darauf, dass  $\lambda$  und q auch von t abhängen. Die Variable  $u = u(t, \mathbf{x})$  steht für die Steuerung. Die Wärmleitungsgleichung wird also durch eine Randkontrolle gesteuert. Wir wollen annehmen, dass die Steuerungen in der zulässigen Menge

$$U_{\mathsf{ad}} = \left\{ u : \Sigma \to \mathbb{R} \left| \int_{t_0}^{t_f} \left| u(t, \boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x} \mathrm{d}t < \infty \text{ and } u_a \leq u \leq u_b \text{ auf } \Sigma \right\}$$

liegen, wobei wir zur Vereinfachung der Darstellung voraussetzen, dass  $u_a$  und  $u_b$  Konstanten sind mit  $u_a < u_b$ . Schließlich nehmen wir in (26.1c) an, dass  $y_{\circ} = y_{\circ}(\mathbf{x})$  eine bekannte Wärmeverteilung zur Zeit  $t = t_{\circ}$  ist.

Angenommen, (26.1) besitzt eine Lösung, die wir als *starke* Lösung bezeichnen. Wir wollen y auf eine Weise beschreiben, die weniger Differenzierbarkeit an y verlangt. Dazu definieren wir den Testraum

 $V_1 = \left\{ \varphi : \overline{\Omega} \to \mathbb{R} \text{ ist stetig und einmal stetig differenzierbar auf } \Omega \right\}$ 

Nun multiplizieren wir (26.1a) mit einer beliebigen Testfunktion  $\varphi \in V_1$ , integrieren über die Menge  $\Omega$  und erhalten nach Anwendung des Gaußschen Integralsatzes

$$\begin{split} &\int_{\Omega} f(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\ &= \int_{\Omega} y_t(t, \boldsymbol{x}) + \lambda(\boldsymbol{x}) \nabla y(t, \boldsymbol{x}) \cdot \nabla \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \int_{\Gamma} \frac{\partial y}{\partial n}(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\ &= \int_{\Omega} y_t(t, \boldsymbol{x}) + \lambda(\boldsymbol{x}) \nabla y(t, \boldsymbol{x}) \cdot \nabla \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma} q(\boldsymbol{x}) y(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\ &- \int_{\Gamma} u(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \end{split}$$

wobei wir die Randbedingung (26.1b) und die Notation

$$\nabla y(t,x) \cdot \nabla \varphi(\boldsymbol{x}) = \sum_{i=1}^{d} \frac{\partial y}{\partial x_i}(t,\boldsymbol{x}) \frac{\partial \varphi}{\partial x_i}(\boldsymbol{x}), \quad (t,\boldsymbol{x}) \in Q$$

verwendet haben. Damit erhalten wir die folgende *schwache* oder *variationelle* Formulierung von (26.1):

$$\int_{\Omega} y_t(t, \boldsymbol{x}) + \lambda(\boldsymbol{x}) \nabla y(t, \boldsymbol{x}) \cdot \nabla \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma} q(\boldsymbol{x}) y(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

$$= \int_{\Omega} f(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma} u(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{für alle } \varphi \in V_1$$
(26.2)

Eine Lösung zu (26.1) bezeichnen wir als *variationelle* oder *schwache Lösung*. Wir können (26.2) auch kompakter schreiben, was uns später hilfreich sein wird. Dazu definieren wir die Abbildung

$$a(\varphi,\phi) = \int_{\Omega} \lambda(\boldsymbol{x}) \nabla \varphi(\boldsymbol{x}) \cdot \nabla \phi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma} q(\boldsymbol{x}) \varphi(\boldsymbol{x}) \phi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{für } \varphi, \phi \in V_1$$

Offenbar ist die Abbildung  $a(\cdot, \cdot)$  linear in beiden Argumenten und wird daher Bilinearform genannt. Ferner seien für jede Zeit  $t \in [t_{\circ}, t_{f}]$  und jede Steuerung  $u \in U_{ad}$  die linearen Funktionale

$$\langle \mathcal{F}(t), \varphi \rangle = \int_{\Omega} f(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \ \langle (\mathcal{B}u)(t), \varphi \rangle = \int_{\Gamma} u(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \text{ für } \varphi \in V_1$$

gegeben. Wir bemerken, dass  $\mathcal{B}u$  linear in der Steuerung u ist, denn es gilt

$$\langle (\mathcal{B}(u+\tilde{u})(t),\varphi) = \int_{\Gamma} \left( u(t,\boldsymbol{x}) + \tilde{u}(t,\boldsymbol{x}) \right) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$
  
$$= \int_{\Gamma} u(t,\boldsymbol{x})\varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma} \tilde{u}(t,\boldsymbol{x})\varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \langle (\mathcal{B}u(t),\varphi) + \langle (\mathcal{B}\tilde{u})(t),\varphi \rangle$$
  
(26.3)

Für  $y \in Y_1$  und für jedes  $\varphi \in V_1$  gilt [1315]

$$\int_{\Omega} y_t(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\Omega} \frac{\mathrm{d}}{\mathrm{d}t} \left( y(t, \boldsymbol{x}) \varphi(\boldsymbol{x}) \right) \, \mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle y(t, \cdot), \varphi \right\rangle_H$$

wobei wir das übliche Skalarprodukt im Raum  $H = L^2(\Omega)$  der quadratisch integrierbaren Funktionen verwendet haben:

$$\langle \varphi, \phi \rangle_H = \int_{\Omega} \varphi(\boldsymbol{x}) \phi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

Weiter steht der Ausdruck  $y(t, \cdot)$  – oder kurz y(t) – für die Funktion y zur Zeit t als Funktion in  $\boldsymbol{x}$ . Nun läßt sich die schwache Formulierung (26.2) wie folgt schreiben:

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle y(t), \varphi \rangle_H + a(y(t), \varphi) = \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \text{für alle } \varphi \in V_1$$
(26.4)

mit  $\langle (\mathcal{F} + \mathcal{B}u)(t), \cdot \rangle = \langle \mathcal{F}(t), \cdot \rangle + \langle (\mathcal{B}u)(t), \cdot \rangle$ . Ist  $y \in Y_1$  eine starke Lösung, so erfüllt y auch die variationelle Formulierung (26.4). Allerdings sind die in (26.4) auftretenden Ableitungen höchstens erster Ordnung. Daher können wir den schwachen Lösbarkeitsbegriff auch wie folgt einführen: Gesucht ist ein  $y \in$  $Y_2 = C^1(\overline{Q}) \supset Y_1$ , welches (26.4) und  $y(t_o) = y_o$  für alle  $\boldsymbol{x} \in \Omega$  erfüllt, wobei wir

$$C^{1}(\overline{Q}) = \left\{ \varphi : \overline{Q} \to \mathbb{R} \mid \varphi, \varphi_{t}, \varphi_{x_{i}} \text{ sind stetig auf } \overline{Q}, \ 1 \leq i, j \leq d \right\}$$

setzen. Es stellt sich aber heraus, dass die Räume  $Y_1$  und  $V_1$  immer noch zu klein sind. Wir führen daher den Sobolev-Raum [1293, 1319]

$$V = \left\{ \varphi \in H \, \middle| \, \int_{\Omega} \sum_{i=1}^{d} \left| \frac{\partial \varphi}{\partial x_{i}}(\boldsymbol{x}) \right| \, \mathrm{d}\boldsymbol{x} < \infty \right\}$$

ein. Dieser Raum ist ein Hilbertraum mit dem Skalarprodukt

$$\langle \varphi, \phi \rangle_V = \int_{\Omega} \varphi(\boldsymbol{x}) \phi(\boldsymbol{x}) + \sum_{i=1}^d \frac{\partial \varphi}{\partial x_i}(\boldsymbol{x}) \frac{\partial \phi}{\partial x_i}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

In der Literatur wird V auch mit  $H^1(\Omega)$  bezeichnet. Unter einer schwachen Lösung der Wärmeleitungsgleichung verstehen wir folgenden Lösungsbegriff:  $y(t) \in V$  heißt schwache oder variationelle Lösung der Wärmeleitungsgleichung, wenn gelten

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle y(t), \varphi \rangle_{H} + a(y(t), \varphi) = \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \forall \varphi \in V, \ t \in (t_{\circ}, t_{f}], 
\langle y(t_{\circ}), \phi \rangle_{H} = \langle y_{\circ}, \phi \rangle_{H} \qquad \forall \phi \in H$$
(26.5)

In [1315] findet man einen Beweis dafür, dass (26.5) eine eindeutige schwache Lösung y besitzt.

#### 26.2.2 Die Finite-Elemente-Methode

Worum geht es? In diesem Abschnitt diskretisieren wir die schwache Formulierung der Wärmeleitungsgleichung. Dazu verwenden wir für die Ortsvariable x die Finite-Elemente-(FE-)Methode und für die Zeitvariable das implizite Eulerverfahren. Das FE-Verfahren ist zuerst in ingenieurwissenschaftlichen Anwendungen entwickelt worden und verbindet die Analysis mit Methoden der Variationsrechnung. Wir verweisen hier beispielsweise auf das das Buch [1395]. Eine mathematische Betrachtung der FE-Methode für elliptische Probleme findet man zum Beispiel in [1306, 1310, 1346, 1382]. Für die hier betrachtete parabolische Wärmeleitungsgleichung verweisen wir auf [1297, 1318, 1385]. Es sei auch auf [1314, Abschnitt 12.4] hingewiesen.

Sei  $\{\Omega_i\}_{i=1}^{n_\Omega}$  eine Triangulierung des Gebiets  $\Omega$  in disjunkte Teilmengen (in der Regel Drei- oder Vierecke)  $\Omega_i \subset \Omega$  (zum Beispiel in Intervalle für d = 1 oder in Dreiecke für d = 2) mit  $\bigcup_{i=1}^{n_\Omega} \overline{\Omega_i} = \overline{\Omega}$ , wobei h > 0 die maximale Seitenlänge der  $\Omega_i$ 's. Damit wird h kleiner, wenn die Triangulierung verfeinert wird. Wir setzen voraus, dass die Winkel der Triangulierung nach unten durch einen von h unabhängingen positiven Winkel beschränkt sind. Die FE-Methode verwendet endlich-dimensionale Ansatzräume  $V^h \subset V$ . Diese bestehen aus stückweisen Polynomen über den Partitionen des Gebiets  $\Omega$  in Elemente wie Dreiecke, Vierecke etc. Wir bezeichnen mit  $\{\varphi_i\}_{i=1}^m$  die linear-unabhängingen Finite-Element-Ansatzfunktionen. Dann ist der m-dimensionale Ansatzraum  $V^h$  gegeben durch

$$V^{h} = \left\{ \varphi^{h} \in V \middle| \varphi^{h} = \sum_{i=1}^{m} c_{i} \varphi_{i} \text{ mit } c_{1}, \dots, c_{m} \in \mathbb{R} \right\} = \operatorname{Span} \left\{ \varphi_{1}, \dots, \varphi_{m} \right\}$$

Offenbar gilt auch  $V^h \subset H$ . Die FE-Galerkin-Approximation von (26.5) lautet dann wie folgt: Gesucht ist  $y^h(t) \in V^h$  für alle  $t \in [t_\circ, t_f]$  mit

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle y^{h}(t), \varphi \rangle_{H} + a(y^{h}(t), \varphi) = \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \forall \varphi \in V^{h}, t \in (t_{\circ}, t_{f}] 
\langle y^{h}(t_{\circ}), \varphi \rangle_{H} = \langle y_{\circ}, \varphi \rangle_{H} \qquad \forall \varphi \in V^{h}$$
(26.6)

Da  $\{\varphi_i\}_{i=1}^m$  eine Basis von  $V^h$  bildet, ist (26.6) äquivalent mit der Formulierung: Gesucht ist  $y^h(t) \in V^h$  für alle  $t \in [t_\circ, t_f]$ , welches für  $1 \le i \le m$ 

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle y^{h}(t), \varphi_{i} \rangle_{H} + a(y^{h}(t), \varphi_{i}) = \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi_{i} \rangle, \quad t \in (t_{\circ}, t_{f}]$$

$$\langle y^{h}(t_{\circ}), \varphi_{i} \rangle_{H} = \langle y_{\circ}, \varphi_{i} \rangle_{H}$$
(26.7)

löst. Aus der Forderung  $y^h(t) \in V^h$  für alle  $t \in [t_\circ, t_f]$  schließen wir, dass es FE-Koeffizienten  $y_1^h(t), \ldots, y_m^h(t)$  gibt für  $t \in [t_\circ, t_f]$  mit

$$y^{h}(t, \boldsymbol{x}) = \sum_{j=1}^{m} y_{j}^{h}(t)\varphi_{j}(\boldsymbol{x}) \quad \text{für } (t, \boldsymbol{x}) \in \overline{Q}$$
(26.8)

Setzen wir (26.8) in (26.7) ein, so erhalten wir das folgende *m*-dimensionale Anfangswertsystem gewöhnlicher Differentialgleichungen:

$$M^{h}(\vec{y}^{h})'(t) + A^{h}\vec{y}^{h}(t) = \vec{f}^{h}(t) + \vec{b}^{h}(u(t)) \quad \text{für } t \in (t_{\circ}, t_{f}]$$

$$M^{h}\vec{y}^{h}(t_{\circ}) = \vec{y}^{h}_{\circ}$$
(26.9)

In (26.9) ist der *m*-dimensionale Vektor  $\vec{y}^h(t) = (y_i^h(t))_{1 \le i \le m}$  der FE-Koeffizienten im FE-Galerkinansatz (26.8) gesucht. Ferner bezeichnen  $M^h$  und  $A^h$  die Masse- beziehungsweise Steifigkeitsmatrix mit den Komponenten

$$M_{ij}^h = \langle \varphi_j, \varphi_i \rangle_H, \quad A_{ij}^h = a(\varphi_j, \varphi_i) \quad \text{für } 1 \le i, j \le m$$

Weiter definieren wir die drei *m*-dimensionalen Vektoren  $\vec{f}^h(t)$ ,  $\vec{b}^h(v)$  mit  $v: \Gamma \to \mathbb{R}$  und  $\vec{y}^h_{\circ}$  mit den Komponenten

$$\mathbf{f}_i^h(t) = \int_{\Omega} f(t)\varphi_i \,\mathrm{d}\boldsymbol{x}, \quad \mathbf{b}_i^h(v) = \int_{\Gamma} v\varphi_i \,\mathrm{d}\boldsymbol{x} \quad \mathbf{y}_{\circ,i}^h = \int_{\Omega} y_\circ \varphi_i \,\mathrm{d}\boldsymbol{x}$$

für die rechte Seite des Differentialgleichungssystems und für die Anfangsbedingung. Aus (26.3) folgt, dass  $\vec{b}(t; u)$  linear in der Steuerung u ist. Die Masseund Steifigkeitsmatrizen sind offenbar symmetrisch. Ferner sind beide Matrizen positiv-definit, das heißt, es gelten

$$\vec{x}^{\top} M^h \vec{x} > 0$$
 und  $\vec{x}^{\top} A^h \vec{x} > 0$  für alle  $\vec{x} \in \mathbb{R}^m \setminus \{0\}$ 

Hier und im Weiteren steht das Symbol " $\top$ " für das Transponieren einer Matrix oder eines Vektors. Symmetrische, positiv-definite Matrizen sind insbesondere invertierbar.

Die Existenz einer eindeutigen Lösung  $y^h(t) \in V^h$  für alle  $t \in [t_o, t_f]$  des Problems (26.7) folgt mit analogen Argumenten wie für die Existenz einer eindeutigen Lösung y für das kontinuierliche Problem (26.5). Der FE-Fehler  $y - y^h$ lässt sich durch Potenzen der maximalen Seitenlänge h abschätzen; vergleiche [1318, 1385].

Zur numerischen Lösung des Anfangswertproblems (26.9) kann man zum Beispiel das *implizite Eulerverfahren* [1314] verwenden. Dazu seien  $t_{\circ} = t_1 < t_2 < \ldots < t_n = t_f$  ein nicht notwendig äquidistantes Zeitgitter in  $[t_{\circ}, t_f]$  und  $\delta t_j = t_j - t_{j-1}, 2 \leq j \leq n$ , die Schrittweiten. Wir setzen  $\bar{f}_j^{h} = \bar{f}^{h}(t_j)$  für  $1 \leq j \leq n$ . In Algorithmus 1 haben wir das implizite Eulerverfahren zusammengestellt.

## Algorithm 1 (FE-Galerkin mit implizitem Eulerverfahren).

**Require:**  $\{\delta t_j\}_{j=1}^n$ ,  $M^h$ ,  $A^h$ ,  $\vec{y}_{\circ}^h$ ,  $\{\vec{f}_j^h\}_{j=1}^n$ ,  $\{\vec{b}^h(u(t_j))\}_{j=1}^n$ ,  $\{\varphi_i\}_{i=1}^m$ . 1: Bestimme  $\vec{y}_1^h$  aus der Gleichung  $M^h \vec{y}_1^h = \vec{y}_{\circ}^h$ . 2: for j = 2 to n do 3: Berechne  $\vec{y}_j^h = ((\vec{y}_j^h)_1, \dots, (\vec{y}_j^h)_m)^\top \in \mathbb{R}^m$  aus der Gleichung

$$(M^{h} + \delta t_{j}A^{h})\vec{y}_{j}^{h} = M^{h}\vec{y}_{j-1}^{h} + \delta t_{j}(\vec{f}_{j}^{h} + \vec{b}_{j}^{h}(u(t_{j}))).$$

4: Setze  $y_j^h = \sum_{i=1}^m (\vec{y}_j^h)_i \varphi_i \in V^h$ . 5: end for 6: return  $Y = [\vec{y}_1^h | \dots | \vec{y}_n^h]$  und  $\{y_j^h\}_{j=1}^n$ .

Es müssen n lineare Gleichungssysteme der Dimension m gelöst werden: eines mit der symmetrischen, positiv-definiten Massematrix  $M^h$  und n-1 mit der symmetrischen, positiv-definiten Koeffizientenmatrix  $M^h + \delta t_j A^h$ . Damit ist die Folge  $\{\vec{y}_j^h\}_{j=1}^n$  eindeutig bestimmt. Wir erhalten somit eine FE-Galerkin-Approximation für die exakte Lösung von (26.5)

$$y(t_j, \boldsymbol{x}) \approx y_j^h(\boldsymbol{x}) = \sum_{i=1}^m Y_{ij} \varphi_i(\boldsymbol{x}) \quad \text{für } 1 \le j \le n \text{ und } \boldsymbol{x} \in \overline{\Omega}$$
 (26.10)

mit der durch Algorithmus 1 berechneten FE-Koeffizientenmatrix  $Y = [\vec{y}_1^h | \dots | \vec{y}_n^h] \in \mathbb{R}^{m \times n}$ , die in der *j*-ten Spalte die FE-Koeffizientenvektoren  $\vec{y}_j^h \in \mathbb{R}^m$  von  $y_j^h$  für  $j = 1, \dots, n$  enthält. Sei y die Lösung der Wärmeleitungsgleichung (26.5). Fehlerabschätzungen für die Differenzen  $\{y(t_j) - y_j^h\}_{j=1}^n$  findet man zum Beispiel in [1318, 1385].

In vielen Anwendungen ist die FE-Dimension m sehr groß oder man ist – zum Beispiel in der Rückkopplungssteuerung – an sehr schnellen Berechnungen der Wärmeverteilung für unterschiedliche Datenvorgaben  $\lambda$ , q, f,  $y_{\circ}$  und u interessiert. In diesen Fällen ist es notwendig, den Aufwand von Algorithmus 1 signifikant zu verringern, ohne dabei die geforderte Genauigkeitsvorgaben zu verletzen. Hierfür verwenden wir Proper Orthogonal Decomposition (POD) als eine Methode der Modellreduktion. Ein weiteres Verfahren ist die Reduced-Basis-Methode [1367]. Wir möchten an dieser Stelle aber auch auf klassische Methoden der Modellreduktion für lineare, dynamische Systeme wie Balanced Truncation oder Moment Matching hinweisen; siehe [1296, 1304, 1394].

## 26.2.3 Die POD-Methode

Worum geht es? In diesem Abschnitt führen wir die POD-Methode für die Wärmeleitungsgleichung (26.1) ein. Für eine andere partielle Differentialgleichung geht man analog vor. Die POD-Methode ist ein Verfahren, POD-Basisfunktionen für (26.1) zu berechnen, die dann an der Stelle der FE-Ansatzfunktionen in einem POD-Galerkinansatz verwendet werden. Aufgrund der Konstruktion enthält die durch die POD-Methode generierte Basis charakteristische Information der Lösung von (26.1). Daher reichen – im Vergleich zur Anzahl m der FE-Ansatzfunktionen – wenige POD-Basisfunktionen aus, um eine gute Approximation der Lösung von (26.1) zu bestimmen. Das führt dann zu den reduzierten Modellen, auf die wir in Abschnitt 26.2.4 eingehen. Es sei hier darauf hingewiesen, dass die POD-Methode auch für nichtlineare Differentialgleichungen verwendet werden kann. Mehr Details findet man zum Beispiel im Buch [1343] und in den Übersichtsartikeln [1331, 1340, 1368, 1374]. In [1356, 1358] wird die POD-Methode auf ein gekoppeltes System von nichtlinearen, partiellen Differentialgleichungen angewendet. Das System besteht aus zwei elliptischen Gleichungen für elektrische Potentiale und einer nichtlinearen parabolischen Gleichung für die Konzentration an Lithium-Ionen-Konzentration in einer Batteriezelle.

In (26.5) wählen wir  $\lambda$ , q, f,  $y_{\circ}$  und eine Steuerung  $u \in U_{ad}$ . Für  $t \in [t_{\circ}, t_f]$  sei  $y(t) \in V$  eine Lösung von (26.5). Die Idee der POD-Methode ist es, die Lösungsmenge

$$\mathscr{V} = \operatorname{Span}\left\{y(t) \in V \mid t \in [t_{\circ}, t_{f}]\right\} \subset V$$

durch wenige orthonormale Ansatzfunktion so gut wie möglich zu approximieren. Da wir aber die Lösungen y(t) nur in Spezialfällen berechnen können, sind wir nur in der Lage, die Menge  $\mathscr{V}$  zu approximieren. Dazu verwenden wir die FE-Funktionen  $\{y_j^h\}_{j=1}^n$  von Algorithmus 1 und führen den approximativen Lösungsraum

$$\mathscr{V}^h = \operatorname{Span}\left\{y_1^h, \dots, y_n^h\right\} \subset V^h$$

ein. Wir bezeichnen  $\mathscr{V}^h$  als *Snapshot-Raum*. Die Elemente  $y_j^h$  nennen wir *Snapshots*. Ziel ist es nun, den Raum  $\mathscr{V}^h$  durch möglichst wenige POD-Ansatzfunktionen zu approximieren. Sei  $d \in \{1, \ldots, n_t\}$  die Dimension von  $\mathscr{V}^h$ . Ferner wählen wir die positiven Trapezgewichte

$$\alpha_1 = \frac{\delta t_1}{2}, \quad \alpha_j = \frac{\delta t_j + \delta t_{j-1}}{2} \text{ für } 2 \le j \le n-1, \quad \alpha_n = \frac{\delta t_n}{2}$$

Da  $\mathscr{V}^h$  ein Teilraum von  $V^h$  ist und  $\mathscr{V}^h$  die Dimension d besitzt, muss d kleiner oder gleich min $\{n, m\}$  sein. Für jedes  $\ell \in \{1, \ldots, d\}$  suchen wir eine *POD-Basis*  $\{\psi_i\}_{i=1}^{\ell}$  vom Rang  $\ell$  als Lösung des Minimierungsproblems

$$\begin{cases} \min \sum_{j=1}^{n} \alpha_{j} \left\| y_{j}^{h} - \sum_{i=1}^{\ell} \left\langle y_{j}^{h}, \psi_{i} \right\rangle_{X} \psi_{i} \right\|_{X}^{2} \\ \text{u.d.N. } \left\{ \psi_{i} \right\}_{i=1}^{\ell} \subset X \text{ and } \left\langle \psi_{i}, \psi_{j} \right\rangle_{X} = \delta_{ij} \text{ für } 1 \leq i, j \leq \ell \end{cases}$$

$$(\mathbf{P}^{\ell})$$

wobei die Abkürzung "u.d.N." für "unter den Nebenbedingungen" steht. In  $(\mathbf{P}^{\ell})$  steht X entweder für den Lebesgue-Raum H oder für den Sobolev-Raum V. Für  $\ell = d$  bildet  $\{\psi_i\}_{i=1}^d$  eine Orthonormal-Basis von  $\mathscr{V}^h$  bezüglich des Skalarprodukts in X. Daher lässt sich jeder Snapshot  $y_j^h \in \mathscr{V}^h$  durch seine Fourierreihe wie folgt approximieren:

$$y_j^h(\boldsymbol{x}) = \sum_{i=1}^d \langle y_j^h, \psi_i \rangle_X \psi_i(\boldsymbol{x}), \quad \boldsymbol{x} \in \overline{\Omega} \text{ und } 1 \le j \le n$$

Damit wird durch ( $\mathbf{P}^{\ell}$ ) für den interessanten Fall  $\ell \ll d$  eine POD-Basis  $\{\psi_i\}_{i=1}^{\ell}$ in X bestimmt, so dass die  $\ell$ -te Partialsumme  $\sum_{i=1}^{\ell} \langle y_j^h, \psi_i \rangle_X \psi_i$  die Snapshots  $y_i^h$  in der X-Norm möglichst gut approximiert.

Die Lösung von  $(\mathbf{P}^{\ell})$  ist durch ein Eigenwertproblem gegeben. Dazu führen wir die lineare Abbildung  $\mathcal{R}: X \to X$  wie folgt ein:

$$\mathcal{R}\varphi := \sum_{j=1}^{n} \alpha_{j} \left\langle y_{j}^{h}, \varphi \right\rangle_{X} y_{j}^{h} \quad \text{für } \varphi \in X$$

Die Abbildung  $\mathcal{R}$  hat folgende Eigenschaften [1331, Lemma 2.2]:

1)  $\mathcal{R}$  ist beschränkt (also auch stetig): Mit der Cauchy-Schwarz-Ungleichung erhalten wir

$$\begin{aligned} \|\mathcal{R}\varphi\|_{X} &= \left\| \sum_{j=1}^{n} \alpha_{j} \left\langle y_{j}^{h}, \varphi \right\rangle_{X} y_{j}^{h} \right\|_{X} \leq \sum_{j=1}^{n} \left( \alpha_{j} \left| \left\langle y_{j}^{h}, \varphi \right\rangle_{X} \right| \left\| y_{j}^{h} \right\|_{X} \right) \\ &\leq \left( \sum_{j=1}^{n} \alpha_{j} \left\| y_{j}^{h} \right\|_{X}^{2} \right) \|\varphi\|_{X} \quad \text{für alle } \varphi \in X \end{aligned}$$

- 2)  $\mathcal{R}(X)$  ist endlich-dimensional: Offenbar folgt aus der Definition von  $\mathcal{R}$  sofort, dass  $\mathcal{R}(X) = \mathscr{V}^h$  gilt. Wir haben also dim  $\mathcal{R}(X) \leq d < \infty$ .
- 3)  $\mathcal{R}$  ist nicht negativ: Für beliebiges  $\varphi \in X$  bekommen wir

$$\left\langle \mathcal{R}\varphi,\varphi\right\rangle_{X} = \left\langle \left.\sum_{j=1}^{n} \alpha_{j}\left\langle y_{j}^{h},\varphi\right\rangle_{X}y_{j}^{h},\varphi\right\rangle_{X} = \sum_{j=1}^{n} \alpha_{j}\left\langle y_{j}^{h},\varphi\right\rangle_{X}^{2} \ge 0\right.$$

4)  $\mathcal{R}$  ist symmetrisch: Das folgt aus der Beziehung

$$\begin{split} \langle \mathcal{R}\varphi, \phi \rangle_X &= \left\langle \sum_{j=1}^n \alpha_j \left\langle y_j^h, \varphi \right\rangle_X y_j^h, \phi \right\rangle_X = \sum_{j=1}^n \alpha_j \left\langle y_j^h, \phi \right\rangle_X \left\langle y_j^h \varphi \right\rangle_X \\ &= \left\langle \sum_{j=1}^n \alpha_j \left\langle y_j^h, \phi \right\rangle_X y_j^h, \varphi \right\rangle_X = \langle \mathcal{R}\phi, \varphi \rangle_X = \langle \varphi, \mathcal{R}\phi \rangle_X \end{split}$$

für beliebige  $\varphi, \phi \in X$ 

Damit können wir die Sätze von Hilbert-Schmidt und Riesz-Schauder für kompakte Operatoren anwenden [1371, Seite 203]. Dabei wird insbesondere genutzt, dass der Raum X in unserem Fall *separabel* ist, also eine abzählbare, dichte Teilmenge besitzt. Ein Beweis des folgenden Satzes findet man in [1331, Theorem 2.7].

**Theorem 26.1** Set X entweder H oder V. Dann gibt es eine Folge nicht negativer Eigenwerte  $\{\lambda_i\}_{i\in\mathbb{N}}$  und eine dazugehörige Orthonormal-Basis in X aus Eigenfunktionen  $\{\psi_i\}_{i\in\mathbb{N}}$  mit

$$\mathcal{R}\psi_i = \lambda_i \psi_i \text{ für } i \in \mathbb{N}, \quad \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_d > \lambda_{d+1} = \ldots = 0$$
 (26.11)

Dann löst  $\{\psi_i\}_{i=1}^{\ell}$  für jedes  $\ell \in \{1, \ldots, d\}$  das Problem ( $\mathbf{P}^{\ell}$ ). Ferner gilt die Fehlerformel

$$\sum_{j=1}^{n} \alpha_{j} \left\| y_{j}^{h} - \sum_{i=1}^{\ell} \left\langle y_{j}^{h}, \psi_{i} \right\rangle_{X} \psi_{i} \right\|_{X}^{2} = \sum_{i=\ell+1}^{d} \lambda_{i}$$
(26.12)

das heißt, der Fehler ist gegeben durch die Summe aller Eigenwerte, deren Eigenfunktionen nicht in der POD-Basis vom Rang  $\ell$  enthalten sind.

### Bemerkung.

1) Die Fehlerformel (26.12) ist essentiell für die Anwendung der POD-Methode. Dabei ist es für den Erfolg der POD-Modellreduktion wichtig, dass  $\sum_{i=\ell+1}^{d} \lambda_i$  klein ist für  $\ell \ll m$ . In der Praxis wird der Rang  $\ell$  oft anhand des (heuristischen) Kriteriums

$$\mathcal{E}(\ell) = \frac{\sum_{i=1}^{\ell} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \in [0, 1]$$

gewählt. Die Eigenwerte  $\lambda_{\ell+1}, \ldots, \lambda_d$  müssen zur Auswertung von  $\mathcal{E}(\ell)$  nicht berechnet werden, da  $\sum_{i=1}^{d} \lambda_i = \sum_{j=1}^{n} \alpha_j \|y_j^h\|_X^2$  gilt [1331, Remark 2.6].

2) Für die Berechnung der POD-Basis haben wir nicht verwendet, dass die Snapshots  $\{y_j^h\}_{j=1}^n$  FE-Lösungen einer linearen Wärmeleitungsgleichung sind. Die Snapshots können auch numerische Lösungen von nichtlinearen, partiellen Differentialgleichungen sein; siehe zum Beispiel [1391, Sections 1.4 und 2.2].

Wir können  $(\mathbf{P}^{\ell})$  in äquivalenter Weise auch anders einführen, und zwar nicht anhand der Snapshots  $\{y_j^h\}_{j=1}^n \subset V^h$  der FE-Lösungen, sondern anhand der Snapshots  $\{\vec{y}_j^h\}_{j=1}^n \subset \mathbb{R}^m$  der FE-Koeffizientenvektoren. Dabei verwenden wir, dass  $V^h$  mit dem euklidischen Raum  $\mathbb{R}^m$  identifiziert werden kann, indem wir in  $\mathbb{R}^m$  ein gewichtetes Skalarprodukt mit der dadurch induzierten Norm einführen.

Für FE-Funktionen  $\varphi^h = \sum_{i=1}^m a_i \varphi_i$  und  $\phi^h = \sum_{j=1}^m b_j \varphi_j$  aus  $V^h$  lassen sich das Skalarprodukt und die Norm in X wie folgt auswerten:

$$\langle \varphi^h, \phi^h \rangle_X = \sum_{i=1}^m \sum_{j=1}^m a_i b_j \left\langle \varphi^h_i, \phi^h_j \right\rangle_X = \vec{a}^\top W \vec{b} =: \langle \vec{a}, \vec{b} \rangle_W$$

$$\| \varphi^h \|_X = \left\langle \varphi^h, \varphi^h \right\rangle_X^{1/2} = \left\langle \vec{a}, \vec{a} \right\rangle_W^{1/2} =: \| \vec{a} \|_W$$

$$(26.13)$$

wobei wir das gewichtete Skalarprodukt  $\langle \cdot, \cdot \rangle_W$  in  $\mathbb{R}^m$  mit der symmetrischen, positiv-definiten Matrix  $W = ((\langle \varphi_i, \varphi_j \rangle_X)) \in \mathbb{R}^{m \times m}$  und der dadurch induzierten Norm  $\| \cdot \|_W$  eingeführt haben. Es seien  $\vec{a} = (a_1, \ldots, a_m)^\top$  und  $\vec{b} = (b_1, \ldots, b_m)^\top$ . Im Fall von der Wahl X = H ist die Matrix W identisch mit der Massematrix  $M^h$ , die wir bereits definiert haben. Im Fall von X = V gilt  $W = M^h + S^h$ , wobei die symmetrische Matrix  $S^h \in \mathbb{R}^{m \times m}$  aus den Elementen

$$S_{ij}^{h} = \left( \left( \int_{\Omega} \nabla \varphi_{i} \cdot \nabla \varphi_{j} \, \mathrm{d}\boldsymbol{x} \right) \right) \quad \text{für } 1 \leq i, j \leq m$$

besteht. In (26.10) haben wir die FE-Koeffizientenmatrix  $Y = [\vec{y}_1^h | \dots | \vec{y}_n^h]$  eingeführt. Aus  $\mathcal{R}(X) = \mathscr{V}^h$  und (26.11) folgt, dass auch die POD-Basisfunktionen  $\{\psi_i\}_{i=1}^{\ell}$  aus  $V^h$  seinen müssen. Es gibt also eine eindeutige FE-Koeffizientenmatrix  $\Psi = [\vec{\psi}_1 | \dots | \vec{\psi}_{\ell}] \in \mathbb{R}^{m \times \ell}$  mit

$$\psi_i(\boldsymbol{x}) = \sum_{i=1}^m \Psi_{ij} \varphi_i(\boldsymbol{x}) \quad \text{für } \boldsymbol{x} \in \overline{\Omega} \text{ und } 1 \leq j \leq \ell$$

Mit (26.13) erhalten wir

$$\left\|y_{j}^{h}-\sum_{i=1}^{\ell}\left\langle y_{j}^{h},\psi_{i}\right\rangle_{X}\psi_{i}\right\|_{X}^{2}=\left\|\vec{y}_{j}^{h}-\sum_{i=1}^{\ell}\left\langle \vec{y}_{j}^{h},\vec{\psi_{i}}\right\rangle_{W}\vec{\psi_{i}}\right\|_{W}^{2},\quad1\leq j\leq m$$

Damit ist jeder POD-Basis  $\{\psi_i\}_{i=1}^{\ell}$  in X eine POD-Basis  $\{\vec{\psi}_i\}_{i=1}^{\ell}$  in  $\mathbb{R}^m$  zugeordnet, wobei die Vektoren  $\vec{\psi}_1, \ldots, \vec{\psi}_{\ell}$  das folgende Minimierungsproblem lösen:

$$\begin{cases} \min \sum_{j=1}^{n} \alpha_{j} \left\| \vec{y}_{j}^{h} - \sum_{i=1}^{\ell} \left\langle \vec{y}_{j}^{h}, \vec{\psi}_{i} \right\rangle_{W} \vec{\psi}_{i} \right\|_{W}^{2} \\ \text{u.d.N. } \{ \vec{\psi}_{i} \}_{i=1}^{\ell} \subset \mathbb{R}^{m} \text{ and } \left\langle \vec{\psi}_{i}, \vec{\psi}_{j} \right\rangle_{W} = \delta_{ij} \text{ für } 1 \leq i, j \leq \ell \end{cases}$$

$$(\mathbf{P}_{W}^{\ell})$$

Wir definieren die Diagonalmatrix  $D = \text{diag}(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^{n \times n}$  mit den Trapezgewichten. Dann erhalten wir

$$(\mathcal{R}\psi_i)(\boldsymbol{x})$$
  
=  $\sum_{j=1}^n \alpha_j \langle y_j^h, \psi_i \rangle_X y_j^h(\boldsymbol{x}) = \sum_{j=1}^n \sum_{l=1}^m \sum_{\nu=1}^m \sum_{\mu=1}^m D_{jj} Y_{lj} W_{l\nu} \Psi_{\nu i} Y_{\mu j} \varphi_\mu(\boldsymbol{x})$ 

1672 26 POD zur Optimalsteuerung Prof. S. Volkwein, Konstanz

$$=\sum_{\mu=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{m}\sum_{\nu=1}^{m}Y_{\mu j}D_{jj}Y_{jl}^{\top}W_{l\nu}\Psi_{\nu i}\varphi_{\mu}(\boldsymbol{x})=\sum_{\mu=1}^{m}\left(YDY^{\top}W\vec{\psi_{i}}\right)_{\mu}\varphi_{\mu}(\boldsymbol{x})$$

für alle  $\boldsymbol{x} \in \overline{\Omega}$  und für  $1 \leq i \leq \ell$ . Ferner gilt

$$\psi_i(\boldsymbol{x}) = \sum_{\mu=1}^m \Psi_{\mu i} \varphi_\mu(\boldsymbol{x}) \quad \text{für alle } \boldsymbol{x} \in \overline{\Omega}, \ 1 \le i \le \ell$$

Also ist die Eigenwertgleichung in (26.11) äquivalent mit dem folgenden Eigenwertproblem für die FE-Koeffizienten der  $\psi_i$ 's:

$$YDY^{\top}W\vec{\psi}_i = \lambda_i\vec{\psi}_i, \quad \lambda_1 \ge \ldots \ge \lambda_d > \lambda_{d+1} = \ldots = 0$$
(26.14)

Da W symmetrisch und positiv-definit ist, besitzt W eine Eigenwertzerlegung der Form  $W = QBQ^{\top}$ , wobei  $B = \text{diag}(\beta_1, \ldots, \beta_m)$  die Eigenwerte  $\beta_1 \ge \ldots \ge \beta_m > 0$  von W enthält und  $Q \in \mathbb{R}^{m \times m}$  eine orthogonale Matrix ist. Wir definieren

$$W^{1/2} = Q \operatorname{diag}\left(\beta_1^{1/2}, \dots, \beta_m^{1/2}\right) Q^{\top}$$

und setzen  $(W^{1/2})^{-1} = W^{-1/2}$ . Für die Diagonalmatrix D erhalten wir  $D^{1/2} =$ diag  $(\alpha_1, \ldots, \alpha_n)$ . Wir setzen  $\vec{u_i} = W^{1/2} \vec{\psi_i}$  in (26.14), multiplizieren (26.14) mit  $W^{1/2}$  von links und führen die Matrix  $\hat{Y} = W^{1/2} Y D^{1/2} \in \mathbb{R}^{m \times n}$  ein. Dann erhalten wir das symmetrische  $m \times m$  Eigenwertproblem

$$\hat{Y}\hat{Y}^{\top}u_i = \lambda_i u_i \quad \text{für } 1 \le i \le \ell$$

Damit ist die POD-Basis  $\{\vec{\psi}_i\}_{i=1}^{\ell}$  vom Rang  $\ell$  gegeben durch Lösen der  $\ell$  linearen Gleichungssysteme  $W^{1/2}\vec{\psi}_i = \vec{u}_i, 1 \leq i \leq \ell$ . Analog zu (26.12) erhalten wir wieder

$$\sum_{j=1}^{n} \alpha_j \left\| \vec{y}_j^h - \sum_{i=1}^{\ell} \left\langle \vec{y}_j^h, \vec{\psi}_i \right\rangle_W \vec{\psi}_i \right\|_W^2 = \sum_{i=\ell+1}^{d} \lambda_i$$

Bemerkung. Aufgrund der Singulärwertzerlegung reeller Matrizen [1314] ergeben sich drei Möglichkeiten zur Berechnung der POD-Basis  $\{\vec{\psi}_i\}_{i=1}^{\ell}$  vom Rang  $\ell$  (siehe [1331, Remarks 2.9 und 2.10] für mehr Details):

1) Löse das  $m \times m$  Eigenwertproblem

$$\hat{Y}\hat{Y}^{\top}u_i = \lambda_i u_i, \ 1 \le i \le \ell, \quad \text{mit} \quad u_i^{\top}u_j = \delta_{ij}, \ 1 \le i, j \le \ell$$

für die größten Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\ell > 0$  und berechne  $\vec{\psi_i} = W^{-1/2}u_i$ . Da  $\hat{Y}\hat{Y}^{\top} = W^{1/2}YDY^{\top}W^{1/2}$  gilt, ist diese Variante numerisch oft sehr teuer, insbesondere wenn wir  $m \gg n$  haben.

2) Löse das  $n \times n$  Eigenwertproblem

$$\hat{Y}^{\top}\hat{Y}v_i = \lambda_i v_i, \ 1 \le i \le \ell, \quad \text{mit} \quad v_i^{\top}v_j = \delta_{ij}, \ 1 \le i, j \le \ell$$
(26.15)

für die größten Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\ell > 0$  und setze  $u_i = YD^{1/2}v_i/\sqrt{\lambda_i}$ . Zur Lösung von (26.15) kann man zum Beispiel in der Programmierumgebung MATLAB die Routine **eigs** verwenden. Wegen  $\hat{Y}^{\top}\hat{Y} = D^{1/2}Y^{\top}WYD^{1/2}$  ist hier die Berechnung von  $W^{1/2}$  nicht erforderlich. Das macht diese Variante insbesondere im Fall von  $n \leq m$  sehr attraktiv bei der numerischen Berechnung der POD-Basis.

3) Berechne die Singulärwertzerlegung [1366] von  $\hat{Y}$ . Dazu bestimmen wir orthonormale Vektoren  $\{u_i\}_{i=1}^{\ell} \subset \mathbb{R}^m$  und  $\{v_i\}_{i=1}^{\ell} \subset \mathbb{R}^n$ , die zu den größten Singulärwerten  $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\ell} > 0$  gehören, das heißt, es gelten

$$\hat{Y}v_i = \sigma_i u_i, \quad \hat{Y}^\top u_i = \sigma_i v_i, \quad 1 \le i \le \ell$$

Es folgt, dass  $\lambda_i = \sigma_i^2$  ind  $\vec{\psi_i} = W^{-1/2}u_i$ . Wir benötigen zwar die Matrix  $W^{1/2}$ , aber diese Variante ist numerisch stabiler, da weder  $\hat{Y}\hat{Y}^{\top}$  noch  $\hat{Y}^{\top}\hat{Y}$  berechnet werden muß. Diese Matrixprodukte haben nämlich eine größere die Konditionszahl als die Matrix  $\hat{Y}$ .

#### 26.2.4 Das reduzierte Modell

Worum geht es? Sei  $\{y_j^h\}_{j=1}^n$  durch Algorithmus 1 zu vorgegebenen Daten  $\lambda, q, f, y_o$  und u berechnet. Wir haben in Abschnitt 26.2.3 die POD-Basis  $\{\psi_i\}_{i=1}^\ell \subset V^h$  vom Rang  $\ell$  eingeführt. Die FE-Koeffizienten  $\{\vec{\psi}_i\}_{i=1}^\ell \subset \mathbb{R}^m$  dieser POD-Basis können mit einer der drei Methoden aus Bemerkung 26.2.3 berechnet werden. Wir wollen in diesem Abschnitt die POD-Basis vom Rang  $\ell$  verwenden, um reduzierte Modelle für die linearen n Gleichungssysteme in Algorithmus 1 herzuleiten. Die Lösung  $\{y_j^\ell\}_{j=1}^n \subset V^h$  dieses reduzierten Modells bezeichnen wir als POD-Lösung der Wärmeleitungsgleichung. Abschließend werden wir einen A-Priori-Fehler für die Differenz von  $y_j^h - y_j^\ell, j = 1, \ldots, n$ , angeben. Mit der berechneten POD-Basis  $\{\psi_i\}_{i=1}^\ell$  vom Rang  $\ell < m$  machen wir den

Mit der berechneten POD-Basis  $\{\psi_i\}_{i=1}^{\ell}$  vom Rang  $\ell < m$  machen wir den folgenden POD-Galerkinansatz für eine POD-Lösung:

$$y_j^{\ell}(\boldsymbol{x}) = \sum_{i=1}^{\ell} Y_{ij}^{\ell} \psi_i(\boldsymbol{x}) = \sum_{l=1}^{m} \left( \sum_{i=1}^{\ell} \Psi_{li} Y_{ij}^{\ell} \right) \varphi_l(\boldsymbol{x}) = \sum_{l=1}^{m} (\Psi Y^{\ell})_{lj} \varphi_l(\boldsymbol{x})$$
(26.16)

für  $x \in \overline{\Omega}$  mit einer zu bestimmenden POD-Koeffizientenmatrix

$$Y^{\ell} = \left[\vec{y}_1^{\ell} \,|\, \dots \,|\, \vec{y}_n^{\ell}\right] \in \mathbb{R}^{\ell \times n}$$

Wir erhalten  $y_j^{\ell} \approx y_j^{h}$  für  $1 \leq j \leq n$ , wenn wir  $\vec{y}_j^{h} \approx \Psi \vec{y}_j^{\ell}$  für  $1 \leq j \leq n$  garantieren. Hier bezeichnet  $\Psi = [\vec{\psi}_1 | \dots | \vec{\psi}_{\ell}]$  die in Abschnitt 26.2.3 eingeführte



http://www.springer.com/978-3-642-30095-0

Elektrische Antriebe – Regelung von Antriebssystemen Schröder, D. 2015, XXXII, 1879 S. 200 Abb. in Farbe., Hardcover ISBN: 978-3-642-30095-0