

Proper Orthogonal Decomposition zur Optimalsteuerung linearer partieller Differentialgleichungen

Stefan Volkwein

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*, das Lösen von nichtlinearen Optimierungsaufgaben erfordert, bei denen 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 mehrere 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 zugekräftiger Techniken der Modellreduktion wie *Balanced Truncation*, *Moment Matching*, *Reduced-Basis-Methode* oder *Proper Orthogonal Decomposition (POD)*, die im Weiteren kurz beschrieben werden.

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 etwas 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 [52, 86, 90, 93] 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 [30, 32]. 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 [29]. 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 [10, 37].

Eine weitere, sehr gut analysierte Methode ist *Balanced Truncation* [33, 97]. Dieses Verfahren verwendet die Lösungen zweier Lyapunov-Gleichungen, die sogenannten Controllability und Observability Gramians. Bei *Balanced Truncation* wird das dynamische System in eine balancierte 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 [69, 87] und [39].

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

[4, 81]. 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 [12]. 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 [24, 34, 70]. 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-Element-Basisfunktionen keine Informationen über die Charakteristiken und Eigenschaften des zu approximierenden dynamischen Systems enthalten. Die Basisfunktionen von Reduced-Basis-Methoden, wie sie etwa in [35, 72] und [17, 50, 38] 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 [31], Fluidodynamik und kohärente Strukturen [2, 6, 43, 48, 66, 84] und bei inversen Problemen [11, 61]. Ferner wurde *POD* dazu eingesetzt, Steuerungen reduzierter Ordnung zu berechnen [8]. Die Beziehung zwischen *POD* und *Balanced Truncation* wurde in [60, 77, 96] untersucht. Eine Fehleranalyse für nichtlineare, n -dimensionale dynamische Systeme (ohne Berücksichtigung von für die Optimalsteuerung wichtigen Aspekten der Optimierung) wurde in [74] durchgeführt. Weiter wurde in [7] 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 [46] und [36, Theorem 4.17] liegen zwar Resultate für eine A-Priori-Analyse 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 [52, 90] wird eine Fehleranalyse 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 unbekannt) 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 Analyse 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 2 führen wir die POD-Methode anhand der Wärmeleitungsgleichung vor. Diese Vorgangsweise wird in Abschnitt 3 auf linear-quadratische Optimalsteuerprobleme angewendet. In Abschnitt 4 stellen wir ein paar Literaturhinweise auf numerische Beispiele zusammen.

2. POD am Beispiel der Wärmeleitungsgleichung

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. 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 Lipschitz-stetigen Rand $\Gamma = \partial\Omega$ und Abschluss $\bar{\Omega} = \Omega \cup \Gamma$. Mit $[t_o, t_f] \subset \mathbb{R}$ bezeichnen wir ein gegebenes Zeitintervall mit $0 \leq t_o < t_f$. Wir definieren den Raum-Zeit-Zylinder $Q = (t_o, t_f) \times \Omega$ mit dem Rand $\Sigma = (t_o, t_f) \times \Gamma$. Weiter benötigen wir folgenden vollständigen Funktionenraum:

$$C^{2,1}(\bar{Q}) = \{ \varphi : \bar{Q} \rightarrow \mathbb{R} \mid \varphi, \varphi_t, \varphi_{x_i}, \varphi_{x_i x_j} \text{ sind stetig auf } \bar{Q}, 1 \leq i, j \leq d \},$$

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

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

$$y_t(t, \mathbf{x}) - \operatorname{div}(\lambda(\mathbf{x})\nabla y(t, \mathbf{x})) = f(t, \mathbf{x}), \quad (t, \mathbf{x}) \in Q, \quad (2.1a)$$

$$\frac{\partial y}{\partial n}(t, \mathbf{x}) + q(\mathbf{x})y(t, \mathbf{x}) = u(t, \mathbf{x}), \quad (t, \mathbf{x}) \in \Sigma, \quad (2.1b)$$

$$y(t_o, \mathbf{x}) = y_o(\mathbf{x}), \quad \mathbf{x} \in \Omega. \quad (2.1c)$$

In (2.1a) 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 $\mathbf{x} \in \bar{\Omega}$. Ferner bezeichnet $f = f(t, \mathbf{x})$ eine verteilte Wärmequelle auf der rechten Seite. In (2.1b) steht $\partial y / \partial n$ für die Normalableitung von y in Richtung der äußeren Normale n an den Rand Γ von Ω . Wir nehmen weiter an, dass $q = q(\mathbf{x})$ eine gegebene verteilte, nichtnegative Funktion ist. Wir verzichten hier zur Vereinfachung der Präsentation darauf, dass λ und q auch von t abhängen. Die Variable $u = u(t, \mathbf{x})$ steht für die *Steuerung*. Die Wärmeleitungsgleichung wird also durch eine Randkontrolle gesteuert. Wir wollen annehmen, dass die Steuerungen in der zulässigen Menge

$$U_{\text{ad}} = \left\{ u : \Sigma \rightarrow \mathbb{R} \mid \int_{t_o}^{t_f} |u(t, \mathbf{x})|^2 \, d\mathbf{x} dt < \infty \text{ and } u_a \leq u \leq u_b \text{ auf } \Sigma \right\}$$

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 (2.1c) an, dass $y_o = y_o(\mathbf{x})$ eine bekannte Wärmeverteilung zur Zeit $t = t_o$ ist.

Angenommen, (2.1) 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

$$V_1 = \{ \varphi : \bar{\Omega} \rightarrow \mathbb{R} \text{ ist stetig und einmal stetig differenzierbar auf } \Omega \}.$$

Nun multiplizieren wir (2.1a) mit einer beliebigen Testfunktion $\varphi \in V_1$, integrieren über die Menge Ω und erhalten nach Anwendung des Gaußschen Integralsatzes

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

wobei wir die Randbedingung (2.1b) und die Notation

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

verwendet haben. Damit erhalten wir die folgende *schwache* oder *variationelle* Formulierung von (2.1):

$$\begin{aligned} & \int_{\Omega} y_t(t, \mathbf{x}) + \lambda(\mathbf{x}) \nabla y(t, \mathbf{x}) \cdot \nabla \varphi(\mathbf{x}) \, d\mathbf{x} + \int_{\Gamma} q(\mathbf{x}) y(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Omega} f(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} + \int_{\Gamma} u(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} \quad \text{für alle } \varphi \in V_1. \end{aligned} \tag{2.2}$$

Eine Lösung zu (2.1) bezeichnen wir als *variationelle* oder *schwache Lösung*. Wir können (2.2) in kompakter schreiben, was uns später hilfreich sein wird. Dazu definieren wir die Abbildung

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

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

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

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

$$\begin{aligned} \langle (\mathcal{B}(u + \tilde{u})(t), \varphi) &= \int_{\Gamma} (u(t, \mathbf{x}) + \tilde{u}(t, \mathbf{x})) \varphi(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Gamma} u(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} + \int_{\Gamma} \tilde{u}(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} = \langle (\mathcal{B}u(t), \varphi) + \langle (\mathcal{B}\tilde{u})(t), \varphi \rangle. \end{aligned} \tag{2.3}$$

Für $y \in Y_1$ und für jedes $\varphi \in V_1$ gilt [23]

$$\int_{\Omega} y_t(t, \mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} = \int_{\Omega} \frac{d}{dt} (y(t, \mathbf{x}) \varphi(\mathbf{x})) \, d\mathbf{x} = \frac{d}{dt} \langle y(t, \cdot), \varphi \rangle_H,$$

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

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

Weiter steht der Ausdruck $y(t, \cdot)$ – oder kurz $y(t)$ – für die Funktion y zur Zeit t als Funktion in \mathbf{x} . Nun läßt sich die schwache Formulierung (2.2) wie folgt schreiben:

$$\frac{d}{dt} \langle y(t), \varphi \rangle_H + a(y(t), \varphi) = \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \text{für alle } \varphi \in V_1 \tag{2.4}$$

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 (2.4). Allerdings sind die in (2.4) 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 (2.4) und $y(t_o) = y_o$ für alle $\mathbf{x} \in \Omega$ erfüllt, wobei wir

$$C^1(\overline{Q}) = \{ \varphi : \overline{Q} \rightarrow \mathbb{R} \mid \varphi, \varphi_t, \varphi_{x_i} \text{ sind stetig auf } \overline{Q}, 1 \leq i, j \leq d \}$$

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* [1, 27]

$$V = \left\{ \varphi \in H \mid \int_{\Omega} \sum_{i=1}^d \left| \frac{\partial \varphi}{\partial x_i}(\mathbf{x}) \right| \, d\mathbf{x} < \infty \right\}$$

ein. Dieser Raum ist ein Hilbertraum mit dem Skalarprodukt

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

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

$$\begin{aligned} \frac{d}{dt} \langle y(t), \varphi \rangle_H + a(y(t), \varphi) &= \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \forall \varphi \in V, t \in (t_o, t_f], \\ \langle y(t_o), \phi \rangle_H &= \langle y_o, \phi \rangle_H \quad \forall \phi \in H. \end{aligned} \tag{2.5}$$

In [23] findet man einen Beweis dafür, dass (2.5) eine eindeutige schwache Lösung y besitzt.

2.2. Die Finite-Elemente-Methode

Worum geht es? In diesem Abschnitt diskreditieren wir die schwache Formulierung der Wärmeleitungsgleichung. Dazu verwenden wir für die Ortsvariable \mathbf{x} die Finite-Elemente-(FE-)Methode und für die Zeitvariable das implizite Eulerverfahren. Der dabei verursachte Diskretisierungsfehler lässt sich quantifizieren. Wir geben hier einen A-Priori-Fehler an.

Das Finite-Elemente-(FE-)Verfahren ist zuerst in ingenieurwissenschaftlichen Anwendungen entwickelt worden und verbindet die Analysis mit Methoden der Variationsrechnung. Wir verweisen hier beispielsweise auf das Buch [98]. Eine mathematische Betrachtung der FE-Methode für elliptische Probleme findet man zum Beispiel in [14, 18, 51, 85]. Für die hier betrachtete parabolische Wärmeleitungsgleichung verweisen wir auf [5, 26, 88]. Es sei auch auf [22, Abschnitt 12.4] hingewiesen.

Sei $\{\Omega_i\}_{i=1}^{n_\Omega}$ eine Triangulierung des Gebiets Ω in disjunkte Teilmen $\Omega_i \subset \Omega$ (zum Beispiel in Intervalle für $d = 1$ oder in Dreiecke für $d = 2$) mit $\bigcup_{i=1}^{n_\Omega} \bar{\Omega}_i = \bar{\Omega}$, wobei $h > 0$ die maximale Seitenlänge der Ω_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 unabhängigen 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 Ω in Elemente wie Dreiecke, Vierecke etc. Wir bezeichnen mit $\{\varphi_i\}_{i=1}^m$ die linear-unabhängigen Finite-Element-Ansatzfunktionen. Dann ist der m -dimensionale Ansatzraum V^h gegeben durch

$$V^h = \left\{ \varphi^h \in V \mid \varphi^h = \sum_{i=1}^m c_i \varphi_i \text{ mit } c_1, \dots, c_m \in \mathbb{R} \right\} = \text{Span} \{ \varphi_1, \dots, \varphi_m \}.$$

Offenbar gilt auch $V^h \subset H$. Die FE-Galerkin-Approximation von (2.5) lautet dann wie folgt: Gesucht ist $y^h(t) \in V^h$ für alle $t \in [t_o, t_f]$ mit

$$\begin{aligned} \frac{d}{dt} \langle y^h(t), \varphi \rangle_H + a(y^h(t), \varphi) &= \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi \rangle \quad \forall \varphi \in V^h, t \in (t_o, t_f], \\ \langle y^h(t_o), \varphi \rangle_H &= \langle y_o, \varphi \rangle_H \quad \forall \varphi \in V^h. \end{aligned} \quad (2.6)$$

Da $\{\varphi_i\}_{i=1}^m$ eine Basis von V^h bildet, ist (2.6) äquivalent mit der Formulierung: Gesucht ist $y^h(t) \in V^h$ für alle $t \in [t_o, t_f]$, welches für $1 \leq i \leq m$

$$\begin{aligned} \frac{d}{dt} \langle y^h(t), \varphi_i \rangle_H + a(y^h(t), \varphi_i) &= \langle (\mathcal{F} + \mathcal{B}u)(t), \varphi_i \rangle, \quad t \in (t_o, t_f], \\ \langle y^h(t_o), \varphi_i \rangle_H &= \langle y_o, \varphi_i \rangle_H \end{aligned} \quad (2.7)$$

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

$$y^h(t, \mathbf{x}) = \sum_{j=1}^m y_j^h(t) \varphi_j(\mathbf{x}) \quad \text{für } (t, \mathbf{x}) \in \overline{Q}. \quad (2.8)$$

Setzen wir (2.8) in (2.7) ein, so erhalten wir das folgende m -dimensionale Anfangswertsystem gewöhnlicher Differentialgleichungen:

$$\begin{aligned} M^h(\bar{y}^h)'(t) + A^h \bar{y}^h(t) &= \bar{f}^h(t) + \bar{b}^h(u(t)) \quad \text{für } t \in (t_o, t_f], \\ M^h \bar{y}^h(t_o) &= \bar{y}_o^h. \end{aligned} \quad (2.9)$$

In (2.9) ist der m -dimensionale Vektor $\bar{y}^h(t) = (y_i^h(t))_{1 \leq i \leq m}$ der FE-Koeffizienten im FE-Galerkinansatz (2.8) gesucht. Ferner bezeichnen M^h und A^h die Masse- beziehungsweise Steifigkeitsmatrix mit den Komponenten

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

Weiter definieren wir die drei m -dimensionalen Vektoren $\bar{f}^h(t)$, $\bar{b}^h(v)$ mit $v : \Gamma \rightarrow \mathbb{R}$ und \bar{y}_o^h mit den Komponenten

$$f_i^h(t) = \int_{\Omega} f(t) \varphi_i \, d\mathbf{x}, \quad b_i^h(v) = \int_{\Gamma} v \varphi_i \, d\mathbf{x} \quad y_{o,i}^h = \int_{\Omega} y_o \varphi_i \, d\mathbf{x}$$

für die rechte Seite des Differentialgleichungssystems und für die Anfangsbedingung. Aus (2.3) folgt, dass $\bar{b}^h(t; u)$ linear in der Steuerung u ist. Die Masse- und Steifigkeitsmatrizen sind offenbar symmetrisch. Ferner sind beide Matrizen positiv-definit, das heißt, es gelten

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

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 $h^h(t) \in V^h$ für alle $t \in [t_o, t_f]$ des problems (2.7) folgt mit analogen Argumenten wie für die Existenz einer eindeutigen Lösung y für das kontinuierliche Problem (2.5). Der FE-Fehler $y - y^h$ lässt sich durch Potenzen der maximalen Seitenlänge h abschätzen; vergleiche [26, 88].

Zur numerischen Lösung des Anfangswertproblems (2.9) kann man zum Beispiel das *implizite Eulerverfahren* [22] verwenden. Dazu seien $t_o = t_1 < t_2 < \dots < t_n = t_f$ ein nicht notwendig äquidistantes Zeitgitter in $[t_o, t_f]$ 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 2.1 haben wir das implizite Eulerverfahren zusammengestellt. 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 $\{\bar{y}_j^h\}_{j=1}^n$ eindeutig bestimmt. Wir erhalten

Algorithm 2.1 (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 $\{y_j^h\}_{j=1}^n$.

somit eine FE-Galerkin-Approximation für die exakte Lösung von (2.5)

$$y(t_j, \mathbf{x}) \approx y_j^h(\mathbf{x}) = \sum_{i=1}^m Y_{ij} \varphi_i(\mathbf{x}) \quad \text{für } 1 \leq j \leq n \text{ und } \mathbf{x} \in \bar{\Omega}, \quad (2.10)$$

mit der durch Algorithmus 2.1 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 (2.5). Fehlerabschätzungen für die Differenzen $\{y(t_j) - y_j^h\}_{j=1}^n$ findet man zum Beispiel in [26, 88].

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 λ , q , f , y_\circ und u interessiert. In diesen Fällen ist es notwendig, den Aufwand von Algorithmus 2.1 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* [72]. Wir möchten an dieser Stelle aber auch auf klassische Methoden der Modellreduktion für lineare, dynamische Systeme wie *Balanced Truncation* oder *Moment Matching* hinweisen; siehe [4, 12, 97].

2.3. Die POD-Methode

Worum geht es? In diesem Abschnitt führen wir die POD-Methode für die Wärmeleitungsgleichung (2.1) ein. Für eine andere partielle Differentialgleichung geht man analog vor. Die POD-Methode ist ein Verfahren, POD-Basisfunktionen für (2.1) zu berechnen, die dann an der Stelle der FE-Ansatzfunktionen in einem POD-Galerkinansatz verwendet werden. Aufgrund der Konstruktion enthalten die durch die POD-Methode generierte Basis charakteristische Information der Lösung von (2.1). Daher reichen – im Vergleich zur Anzahl m der FE-Ansatzfunktionen – wenige POD-Basisfunktionen aus, um eine gute Approximation der Lösung von (2.1) zu bestimmen. Das führt dann zu den reduzierten Modellen, auf die wir in Abschnitt 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 [48] und in den Übersichtsartikeln [36, 45, 73, 78]. In [61, 63] 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 (2.5) wählen wir λ, q, f, y_0 und eine Steuerung $u \in U_{\text{ad}}$. Für $t \in [t_0, t_f]$ sei $y(t) \in V$ eine Lösung von (2.5). Die Idee der POD-Methode ist es, die Lösungsmenge

$$\mathcal{V} = \text{Span} \{y(t) \in V \mid t \in [t_0, t_f]\} \subset V$$

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 \mathcal{V} zu approximieren. Dazu verwenden wir die FE-Funktionen $\{y_j^h\}_{j=1}^n$ von Algorithmus 2.1 und führen den approximativen Lösungsraum

$$\mathcal{V}^h = \text{Span} \{y_1^h, \dots, y_n^h\} \subset V^h$$

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

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

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

$$\begin{cases} \min \sum_{j=1}^n \alpha_j \left\| y_j^h - \sum_{i=1}^{\ell} \langle y_j^h, \psi_i \rangle_X \psi_i \right\|_X^2 \\ \text{u.d.N. } \{\psi_i\}_{i=1}^\ell \subset X \text{ and } \langle \psi_i, \psi_j \rangle_X = \delta_{ij} \text{ für } 1 \leq i, j \leq \ell, \end{cases} \quad (\mathbf{P}^\ell)$$

wobei die Abkürzung „u.d.N.“ für „unter den Nebenbedingungen“ steht. In (\mathbf{P}^ℓ) 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 \mathcal{V}^h bezüglich des Skalarprodukts in X . Daher lässt sich jeder Snapshot $y_j^h \in \mathcal{V}^h$ durch seine Fourierreihe wie folgt approximieren:

$$y_j^h(\mathbf{x}) = \sum_{i=1}^d \langle y_j^h, \psi_i \rangle_X \psi_i(\mathbf{x}), \quad \mathbf{x} \in \bar{\Omega} \text{ und } 1 \leq j \leq n.$$

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

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

$$\mathcal{R}\varphi := \sum_{j=1}^n \alpha_j \langle y_j^h, \varphi \rangle_X y_j^h \quad \text{für } \varphi \in X.$$

Die Abbildung \mathcal{R} hat folgende Eigenschaften [36, Lemma 2.2]:

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

$$\begin{aligned} \|\mathcal{R}\varphi\|_X &= \left\| \sum_{j=1}^n \alpha_j \langle y_j^h, \varphi \rangle_X y_j^h \right\|_X \leq \sum_{j=1}^n \left(\alpha_j |\langle y_j^h, \varphi \rangle_X| \|y_j^h\|_X \right) \\ &\leq \left(\sum_{j=1}^n \alpha_j \|y_j^h\|_X^2 \right) \|\varphi\|_X \quad \text{für alle } \varphi \in X. \end{aligned}$$

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

$$\langle \mathcal{R}\varphi, \varphi \rangle_X = \left\langle \sum_{j=1}^n \alpha_j \langle y_j^h, \varphi \rangle_X y_j^h, \varphi \right\rangle_X = \sum_{j=1}^n \alpha_j \langle y_j^h, \varphi \rangle_X^2 \geq 0.$$

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

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

für beliebige $\varphi, \phi \in X$.

Damit können wir die Sätze von Hilbert-Schmidt und Riesz-Schauder für kompakte Operatoren anwenden [76, 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 [36, Theorem 2.7].

Satz 2.1. *Sei X entweder H oder V . Dann gibt es eine Folge nichtnegativer Eigenwerte $\{\lambda_i\}_{i \in \mathbb{N}}$ und eine dazugehörige Orthonormal-Basis in X aus Eigenfunktionen $\{\psi_i\}_{i \in \mathbb{N}}$ mit*

$$\mathcal{R}\psi_i = \lambda_i \psi_i \quad \text{für } i \in \mathbb{N}, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > \lambda_{d+1} = \dots = 0. \quad (2.11)$$

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

$$\sum_{j=1}^n \alpha_j \left\| y_j^h - \sum_{i=1}^{\ell} \langle y_j^h, \psi_i \rangle_X \psi_i \right\|_X^2 = \sum_{i=\ell+1}^d \lambda_i. \quad (2.12)$$

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

Bemerkung 2.2. 1) Die Fehlerformel (2.12) 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 ℓ 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 $\lambda_{\ell+1}, \dots, \lambda_d$ 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 [36, 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 [94, Sections 1.4 und 2.2]. \diamond

Wir können (\mathbf{P}^ℓ) 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 $\{\bar{y}_j^h\}_{j=1}^n \subset \mathbb{R}^m$ 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:

$$\begin{aligned} \langle \varphi^h, \phi^h \rangle_X &= \sum_{i=1}^m \sum_{j=1}^m a_i b_j \langle \varphi_i^h, \phi_j^h \rangle_X = \vec{a}^\top W \vec{b} =: \langle \vec{a}, \vec{b} \rangle_W, \\ \|\varphi^h\|_X &= \langle \varphi^h, \varphi^h \rangle_X^{1/2} = \langle \vec{a}, \vec{a} \rangle_W^{1/2} =: \|\vec{a}\|_W, \end{aligned} \quad (2.13)$$

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 $\vec{a} = (a_1, \dots, a_m)^\top$ und $\vec{b} = (b_1, \dots, b_m)^\top$. 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

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

besteht. In (2.10) haben wir die FE-Koeffizientenmatrix $Y = [\bar{y}_1^h | \dots | \bar{y}_n^h]$ eingeführt. Aus $\mathcal{R}(X) = \mathcal{V}^h$ und (2.11) 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

$$\psi_i(\mathbf{x}) = \sum_{j=1}^m \Psi_{ij} \varphi_j(\mathbf{x}) \quad \text{für } \mathbf{x} \in \bar{\Omega} \text{ und } 1 \leq j \leq \ell.$$

Mit (2.13) erhalten wir

$$\left\| y_j^h - \sum_{i=1}^{\ell} \langle y_j^h, \psi_i \rangle_X \psi_i \right\|_X^2 = \left\| \vec{y}_j^h - \sum_{i=1}^{\ell} \langle \vec{y}_j^h, \vec{\psi}_i \rangle_W \vec{\psi}_i \right\|_W^2, \quad 1 \leq j \leq m.$$

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 $\vec{\psi}_1, \dots, \vec{\psi}_\ell$ das folgende Minimierungsproblem lösen:

$$\begin{cases} \min \sum_{j=1}^n \alpha_j \left\| \vec{y}_j^h - \sum_{i=1}^{\ell} \langle \vec{y}_j^h, \vec{\psi}_i \rangle_W \vec{\psi}_i \right\|_W^2 \\ \text{u.d.N. } \{\vec{\psi}_i\}_{i=1}^\ell \subset \mathbb{R}^m \text{ and } \langle \vec{\psi}_i, \vec{\psi}_j \rangle_W = \delta_{ij} \text{ für } 1 \leq i, j \leq \ell, \end{cases} \quad (\mathbf{P}_W^\ell)$$

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

$$\begin{aligned} & (\mathcal{R}\psi_i)(\mathbf{x}) \\ &= \sum_{j=1}^n \alpha_j \langle y_j^h, \psi_i \rangle_X y_j^h(\mathbf{x}) = \sum_{j=1}^n \sum_{l=1}^m \sum_{\nu=1}^m \sum_{\mu=1}^m D_{jj} Y_{lj} W_{l\nu} \Psi_{\nu i} Y_{\mu j} \varphi_\mu(\mathbf{x}) \\ &= \sum_{\mu=1}^m \sum_{j=1}^n \sum_{l=1}^m \sum_{\nu=1}^m Y_{\mu j} D_{jj} Y_{jl}^\top W_{l\nu} \Psi_{\nu i} \varphi_\mu(\mathbf{x}) = \sum_{\mu=1}^m (Y D Y^\top W \vec{\psi}_i)_\mu \varphi_\mu(\mathbf{x}) \end{aligned}$$

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

$$\psi_i(\mathbf{x}) = \sum_{\mu=1}^m \Psi_{\mu i} \varphi_\mu(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \bar{\Omega}, \quad 1 \leq i \leq \ell.$$

Also ist die Eigenwertgleichung in (2.11) äquivalent mit dem folgenden Eigenwertproblem für die FE-Koeffizienten der ψ_i 's:

$$Y D Y^\top W \vec{\psi}_i = \lambda_i \vec{\psi}_i, \quad \lambda_1 \geq \dots \geq \lambda_d > \lambda_{d+1} = \dots = 0. \quad (2.14)$$

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

$$W^{1/2} = Q \text{diag}(\beta_1^{1/2}, \dots, \beta_m^{1/2}) Q^\top.$$

und setzen $(W^{1/2})^{-1} = W^{-1/2}$. Für die Diagonalmatrix D erhalten wir $D^{1/2} = \text{diag}(\alpha_1, \dots, \alpha_n)$. Wir setzen $\vec{u}_i = W^{1/2} \vec{\psi}_i$ in (2.14), multiplizieren

(2.14) mit $W^{1/2}$ von links und führen die Matrix $\hat{Y} = W^{1/2}YD^{1/2} \in \mathbb{R}^{m \times n}$ ein. Dann erhalten wir das symmetrische $m \times m$ Eigenwertproblem

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

Damit ist die POD-Basis $\{\vec{\psi}_i\}_{i=1}^\ell$ vom Rang ℓ gegeben durch Lösen der ℓ linearen Gleichungssysteme $W^{1/2}\vec{\psi}_i = \vec{u}_i$, $1 \leq i \leq \ell$. Analog zu (2.12) erhalten wir wieder

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

Bemerkung 2.3. Aufgrund der Singulärwertzerlegung reeller Matrizen [22] ergeben sich drei Möglichkeiten zur Berechnung der POD-Basis $\{\vec{\psi}_i\}_{i=1}^\ell$ vom Rang ℓ (siehe [36, Remarks 2.9 und 2.10] für mehr Details):

- 1) Löse das $m \times m$ Eigenwertproblem

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

für die größten Eigenwerte $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell > 0$ 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

$$\hat{Y}^\top \hat{Y} v_i = \lambda_i v_i, \quad 1 \leq i \leq \ell, \quad \text{mit} \quad v_i^\top v_j = \delta_{ij}, \quad 1 \leq i, j \leq \ell, \quad (2.15)$$

für die größten Eigenwerte $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell > 0$ und setze $u_i = YD^{1/2}v_i/\sqrt{\lambda_i}$. Zur Lösung von (2.15) 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 [71] 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 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_\ell > 0$ gehören, das heißt, es gelten

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

Es folgt, dass $\lambda_i = \sigma_i^2$ und $\vec{\psi}_i = W^{-1/2}u_i$. 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} . \diamond

2.4. Die reduzierte Modell

Worum geht es? Sei $\{y_j^h\}_{j=1}^n$ durch Algorithmus 2.1 zu vorgegebenen Daten λ , q , f , y_0 und u berechnet. Wir haben in Abschnitt 2.3 die POD-Basis $\{\psi_i\}_{i=1}^\ell \subset V^h$ vom Rang ℓ 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 2.3 berechnet werden. Wir wollen in diesem Abschnitt die POD-Basis vom Rang ℓ verwenden, um reduzierte Modelle für die linearen n Gleichungssysteme in

Algorithmus 2.1 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 $y_j^h - y_j^\ell$, $j = 1, \dots, n$, angeben.

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

$$y_j^\ell(\mathbf{x}) = \sum_{i=1}^{\ell} Y_{ij}^\ell \psi_i(\mathbf{x}) = \sum_{l=1}^m \left(\sum_{i=1}^{\ell} \Psi_{li} Y_{ij}^\ell \right) \varphi_l(\mathbf{x}) = \sum_{l=1}^m (\Psi Y^\ell)_{lj} \varphi_l(\mathbf{x}), \quad (2.16)$$

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

$$Y^\ell = [\bar{y}_1^\ell | \dots | \bar{y}_n^\ell] \in \mathbb{R}^{\ell \times n}.$$

Wir erhalten $y_j^\ell \approx y_j^h$ für $1 \leq j \leq n$, wenn wir $\bar{y}_j^h \approx \Psi \bar{y}_j^\ell$ für $1 \leq j \leq n$ garantieren. Hier bezeichnet $\Psi = [\bar{\psi}_1 | \dots | \bar{\psi}_\ell]$ die in Abschnitt 2.3 eingeführte Matrix der FE-Koeffizienten der ψ_i 's. Daher setzen wir in Algorithmus 2.1 für \bar{y}_j^h die Approximation $\Psi \bar{y}_j^\ell$ ein:

$$\begin{aligned} M^h \Psi \bar{y}_1^\ell &= \bar{y}_o^h; \\ \text{Löse für } j &= 2, \dots, n : \end{aligned} \quad (2.17)$$

$$(M^h + \delta t_j A^h) \Psi \bar{y}_j^\ell = M^h \Psi \bar{y}_{j-1}^\ell + \delta t_j (\bar{f}_j^h + b^h(u(t_j))).$$

In (2.17) haben wir n lineare Gleichungssysteme der Dimension m für jeweils ℓ unbekannte Komponenten der n POD-Koeffizientenvektoren $\bar{y}_1^\ell, \dots, \bar{y}_n^\ell \in \mathbb{R}^\ell$. Um dieses Problem aufzulösen, führen wir eine Galerkinprojektion durch, indem wir die Gleichungssysteme von links mit der Matrix Ψ^\top multiplizieren. Die dann auftretenden Matrizen und Vektoren berechnen wir mit Algorithmus 2.2. Aus (2.17) leiten wir Algorithmus 2.3 her, wobei wir die

Algorithm 2.2 (Reduziertes Modell mit POD)

Require: $M^h, A^h, \Psi, \bar{y}_o^h, \{\bar{f}_j^h\}_{j=1}^n, \{\bar{b}^h(u(t_j))\}_{j=1}^n$.

- 1: Setze $M^\ell = \Psi^\top M^h \Psi$ und $A^\ell = \Psi^\top A^h \Psi$.
 - 2: Berechne $\bar{y}_o^\ell = \Psi^\top \bar{y}_o^h$ und $\bar{f}_j^\ell = \Psi^\top \bar{f}_j^h$, $\bar{b}^\ell(u(t_j)) = \Psi^\top \bar{b}^h(u(t_j))$ für $1 \leq j \leq n$.
 - 3: **return** $M^\ell, A^\ell, \bar{y}_o^\ell, \{\bar{f}_j^\ell\}_{j=1}^n, \{\bar{b}^\ell(u(t_j))\}_{j=1}^n$.
-

mit Algorithmus 2.2 berechneten Vektoren und Matrizen verwenden. Die n in Algorithmus 2.3 auftauchenden linearen Gleichungssysteme sind von der Dimension $\ell \ll m$. Daher erwarten wir, dass die Ausführung von Algorithmus 2.3 deutlich schneller geht als die von Algorithmus 2.1. Nun wollen wir der Frage gehen, ob die POD-Lösung, die wir mit Algorithmus 2.3 berechnen, von ähnlicher Güte ist wie die FE-Lösung. Aus der Fehlerformel (2.12) folgt aber nicht automatisch, dass der Fehler

$$\sum_{j=1}^n \alpha_j \|y_j^h - y_j^\ell\|_X^2 = \sum_{j=1}^n \alpha_j \|\bar{y}_j^h - \bar{y}_j^\ell\|_W^2$$

Algorithm 2.3 (POD-Galerkin mit implizitem Eulerverfahren)

Require: $\{\delta t_j\}_{j=1}^n$, M^ℓ , A^ℓ , \vec{y}_0^ℓ , $\{\vec{f}_j^\ell\}_{j=1}^n$, $\{\vec{b}^\ell(u(t_j))\}_{j=1}^n$, $\{\psi_i\}_{i=1}^\ell$.

1: Bestimme \vec{y}_1^ℓ aus der Gleichung $M^\ell \vec{y}_1^\ell = \vec{y}_0^\ell$.

2: **for** $j = 2$ **to** n **do**

3: Berechne $\vec{y}_j^\ell = ((\vec{y}_j^\ell)_1, \dots, (\vec{y}_j^\ell)_\ell)^\top \in \mathbb{R}^\ell$ aus der Gleichung

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

4: Setze $y_j^\ell = \sum_{i=1}^\ell (\vec{y}_j^\ell)_i \psi_i$.

5: **end for**

6: **return** $Y^\ell = [\vec{y}_1^\ell | \dots | \vec{y}_n^\ell]$ und $\{y_j^\ell\}_{j=1}^n$.

ebenfalls mit der Rate $\sum_{i=\ell+1}^d \lambda_i$ abklingt. Daher ist es wichtig, zur Rechtfertigung des Einsatzes der POD-Methode eine A-Priori-Fehleranalyse durchzuführen. Kombinieren wir die Resultate aus [54, Theorem 7], [55, Theorem 4.7] und [36, Theorem 3.11], so erhalten wir den folgenden Satz 2.4. Hier sind auch die neuen Fehlerabschätzungen aus der Arbeit [83] mitberücksichtigt.

Satz 2.4. *Es existiere eine Konstante $c > 0$, die nicht von n abhängt, mit*

$$1 \leq \frac{\max_{1 \leq j \leq n} \delta t_j}{\min_{1 \leq j \leq n} \delta t_j} \leq c.$$

Seien die Vektoren $\{\vec{y}_j^\ell\}_{j=1}^n \subset \mathbb{R}^m$ mit Algorithmus 2.1 berechnet. Für beliebiges $\ell \in \{1, \dots, d\}$ und für die Wahl $X = H$ oder V sei $\{\psi_i\}_{i=1}^\ell \subset X$ eine POD-Basis vom Rang ℓ . Sei $\{y_j^\ell\}_{j=1}^n \subset \mathbb{R}^\ell$ die durch Algorithmus 2.3 berechnete POD-Lösung. Wir setzen

$$y_j^h(\mathbf{x}) = \sum_{j=1}^m (\vec{y}_j^h)_i \varphi_i(\mathbf{x}), \quad y_j^\ell(\mathbf{x}) = \sum_{j=1}^\ell (\vec{y}_j^\ell)_i \psi_i(\mathbf{x}), \quad \mathbf{x} \in \bar{\Omega}, 1 \leq j \leq n.$$

Dann existiert eine Konstante $C > 0$, die unabhängig von ℓ , m , $\{\delta t_j\}_{j=1}^n$ ist, mit

$$\sum_{j=1}^n \alpha_j \|y_j^h - y_j^\ell\|_X^2 \leq C \left(\|\mathcal{P}^\ell y_1^h - y_1^\ell\|_H^2 + \sum_{i=\ell+1}^d \lambda_i \|\psi_i\|_V^2 \right) \quad (2.18)$$

für $X = H$ und

$$\sum_{j=1}^n \alpha_j \|y_j^h - y_j^\ell\|_X^2 \leq C \left(\|\mathcal{P}^\ell y_1^h - y_1^\ell\|_H^2 + \sum_{i=\ell+1}^d \lambda_i \|\psi_i - \mathcal{P}^\ell \psi_i\|_V^2 \right) \quad (2.19)$$

für X_V , wobei $\mathcal{P}^\ell : X \rightarrow \text{Span} \{\psi_1, \dots, \psi_\ell\} \subset X$ die bezüglich X orthogonale, lineare Projektion

$$\mathcal{P}^\ell \varphi = \sum_{i=1}^\ell \langle \psi_i, \varphi \rangle_X \psi_i \quad \text{für } \varphi \in X$$

bezeichnet.

- Bemerkung 2.5.** 1) In [83] wird gezeigt, dass $\{\psi_i\}_{i=1}^d \in V$ auch für die Wahl $X = H$ gilt. Satz 2.4 zeigt, dass der Abfall der Eigenwerte λ_i eine wichtige Rolle spielt für die Konvergenz der POD- gegen die FE-Lösung. Allerdings wird der Abfall mit $\|\psi_i\|_V^2$ im Fall $X = H$ und mit $\|\psi_i - \mathcal{P}^\ell \psi_i\|_V^2$ im Fall von $X = V$ gewichtet.
- 2) Der Fehler $\|\mathcal{P}^\ell y_1^h - y_1^\ell\|_H$ in den Anfangsdaten an $t = t_o$ lässt sich durch den modifizierten POD-Galerkinansatz

$$y_j^\ell(\mathbf{x}) = y_1^h(\mathbf{x}) + \sum_{i=1}^{\ell} (\tilde{y}_j^\ell)_i \psi_i(\mathbf{x}), \quad \mathbf{x} \in \bar{\Omega}, 1 \leq j \leq n. \quad (2.20)$$

vermeiden [36]. In diesem Fall wird die POD-Basis vom Rang ℓ als Lösung von

$$\begin{cases} \min \sum_{j=1}^n \alpha_j \left\| \tilde{y}_j^h - \sum_{i=1}^{\ell} \langle \tilde{y}_j^h, \psi_i \rangle_X \psi_i \right\|_X^2 \\ \text{u.d.N. } \{\psi_i\}_{i=1}^{\ell} \subset X \text{ and } \langle \psi_i, \psi_j \rangle_X = \delta_{ij} \text{ für } 1 \leq i, j \leq \ell, \end{cases}$$

berechnet, wobei wir $\tilde{y}_j^h = y_j^h - y_1^h \in V^h$ gesetzt haben. Wir erhalten $\tilde{y}_1^\ell = 0$ und somit $y_1^\ell = y_1^h$ für alle $\mathbf{x} \in \bar{\Omega}$.

- 3) Wenn man die Differenzenquotienten $(y_j^h - y_{j-1}^h)/\delta t_j$ mit in den Snapshotraum \mathcal{V}^h hineinnimmt und den modifizierten Ansatz (2.20) verwendet, lässt sich der A-Priori-Fehler im Fall $X = V$ von der Ordnung $\sum_{i=\ell+1}^d \lambda_i$.
- 4) Wir weisen auf die Arbeiten [55, 56], wo für die Snapshots und die Fehlerbetrachtung unterschiedliche Zeitgitter verwendet werden. Ferner wird in [56] das Crank-Nicolson-Verfahren anstelle des impliziten Eulerverfahrens für die Zeitintegration verwendet.
- 5) Man kann die Wahl der FE-Snapshots für die POD-Basisberechnung auch so wählen, dass der Fehler (oder ein anderes Kriterium) möglichst klein wird. Diese Methode wird *Optimal Snapshot Location* genannt und ist in [59] eingeführt worden. Verwandte Strategien findet man auch in [15]. Diese Technik kann auch dazu verwendet werden, auf adaptive Weise POD-Bases aufzubauen [62]. \diamond

Die Behandlung nichtlinearer Probleme macht es notwendig, die nichtlinearen Terme effizient auszuwerten. Hier hat sich die Methode der *Empirischen Interpolation* als sehr effizient herausgestellt. Diese Strategie wurde in [9] eingeführt. Wir weisen hier auch auf die Arbeiten [19, 20, 21], wo eine diskrete Variante der Empirischen Interpolation untersucht wird.

3. Die quadratische Optimalsteuerproblem

Die in Satz 2.4 betrachtete Situation ist aus der Sicht der Modellreduktion noch nicht sehr interessant. Wir haben zwar mit der POD-Lösung $\{y_j^\ell\}_{j=1}^n$ eine Approximation von $\{y_j^h\}_{j=1}^n$ berechnet und können den Fehler abschätzen, aber zur Berechnung von der POD-Basis benötigten wir bereits in (\mathbf{P}^ℓ) die FE-Lösung $\{y_j^h\}_{j=1}^n$. Der entscheidende Einsatz der POD-Modellreduktion ist nun, dass die berechnete POD-Basis auch verwendet wird, wenn Daten in der Wärmeleitungsgleichung (2.1) (leicht) verändert werden. In diesem Abschnitt werden wir uns mit der Anwendung der POD-Methode im Bereich der (Open-Loop-)Optimalsteuerung beschäftigen. Wir halten daher die Daten λ , q , y_o , f in (2.1) fest und wollen eine optimale Steuerung, die wir im weiteren mit \bar{u} bezeichnen, bestimmen, so dass ein vorgegebenes Kriterium minimiert wird. Da die optimale Steuerung \bar{u} nicht bekannt ist, muß die POD-Basis $\{\psi_i\}_{i=1}^\ell$ aus einer FE-Lösung zu einer möglichst guten Schätzung u_{exp} von \bar{u} berechnet werden. In der Optimalsteuerung wird, falls die Schätzung nicht so gut ist, eine Anpassung der POD-Basis vorgenommen. Zu diesem Zweck werden A-Posteriori-Fehlerschätzer angegeben.

3.1. Das Optimalsteuerproblem

Worum geht es? In diesem Abschnitt führen wir das Optimalsteuerproblem ein und geben Optimalitätsbedingungen erster Ordnung an. Das führt auf eine duale Gleichung und eine Variationsungleichung.

Wir werden als Beispiel das quadratische *Ziel- oder Kostenfunktion*

$$J(y, u) = \frac{1}{2} \int_{\Omega} |y(t_f, \mathbf{x}) - y_d(\mathbf{x})|^2 \, d\mathbf{x} + \frac{\gamma}{2} \int_{t_o}^{t_f} \int_{\Gamma} |u(t, \mathbf{x})|^2 \, d\mathbf{x} dt$$

betrachten. Wir nehmen an, dass die Zielvorgabe $y_d : \bar{\Omega} \rightarrow \mathbb{R}$ beschränkt ist. Ferner sei $\gamma > 0$ ein gegebener Gewichtungparameter. Die Menge U_{ad} der zulässigen Steuerungen hatten wir bereits in Abschnitt 2.1 eingeführt. Wir betrachten dann das linear-quadratische Optimierungsproblem

$$\min J(y, u) \quad \text{u.d.N.} \quad u \in U_{ad} \text{ und } (y, u) \text{ erfüllt (2.5).} \quad (\mathbf{QP})$$

Man kann zeigen, dass (\mathbf{QP}) genau eine optimale Lösung $\bar{x} = (\bar{y}, \bar{u})$ besitzt; siehe zum Beispiel in [65, 89]. Hier verwendet man unter Anderem, dass $\gamma > 0$ gilt und U_{ad} eine konvexe Menge ist. Da zu jedem $u \in U_{ad}$ eine eindeutige Lösung $y = y(u)$ von (2.5) existiert, ist y keine freie Variable im Optimierungsproblem (\mathbf{QP}) . Man bezeichnet daher (\mathbf{QP}) als *Optimalsteuerproblem*; siehe [89].

Zur Berechnung der optimalen Lösung (\bar{y}, \bar{u}) ist es hilfreich, Optimalitätsbedingungen herzuleiten. Wir verfolgen hier die *First-Optimize-Then-Discretize-Vorgangsweise*. Dabei wird nicht (\mathbf{QP}) zum Beispiel durch die FE-Methode und das implizite Eulerverfahren in ein endlichdimensionales Optimierungsproblem überführt, sondern erst Optimalitätsbedingungen hergeleitet, die dann geeignet diskreditiert werden.

Die Herleitung der Optimalitätsbedingungen basieren auf der Lagrangefunktion für **(QP)**:

$$\begin{aligned} \mathcal{L}(y, u, p, p_o) &= J(y, u) + \int_{t_o}^{t_f} \langle y_t(t), p(t) \rangle_H + a(y(t), p(t)) dt \\ &\quad + \int_{t_o}^{t_f} \langle (\mathcal{F} + \mathcal{B}u)(t), p(t) \rangle dt + \langle y(0) - y_o, p_o \rangle_H. \end{aligned}$$

Zum Beweis des folgenden Satzes verweisen wir auf [42, 65, 89].

Satz 3.1. *Sei $\bar{x} = (\bar{y}, \bar{u})$ die optimale Lösung von **(QP)**. Dann existiert ein eindeutiges Paar (\bar{p}, \bar{p}_o) an Lagrange-Multiplikatoren, die zusammen mit \bar{x} das folgende Optimalitätssystem erfüllen:*

$$\frac{d}{dt} \langle \bar{y}(t), \varphi \rangle_H + a(\bar{y}(t), \varphi) = \langle (\mathcal{F} + \mathcal{B}\bar{u})(t), \varphi \rangle \quad \forall \varphi \in V, \quad t \in (t_o, t_f], \quad (3.1a)$$

$$\langle \bar{y}(t_o), \phi \rangle_H = \langle y_o, \phi \rangle_H \quad \forall \phi \in H, \quad (3.1b)$$

$$-\frac{d}{dt} \langle \bar{p}(t), \varphi \rangle_H + a(\bar{p}(t), \varphi) = 0 \quad \forall \varphi \in V, \quad t \in [t_o, t_f], \quad (3.1c)$$

$$\langle \bar{p}(t_f), \phi \rangle_H = \langle y_d - \bar{y}(t_f), \phi \rangle_H \quad \forall \phi \in H, \quad (3.1d)$$

$$\int_{t_o}^{t_f} \int_{\Gamma} (\gamma \bar{u}(t) - \bar{p}(t)) (u(t) - \bar{u}(t)) \, d\mathbf{x} dt \geq 0 \quad \text{für alle } u \in U_{\text{ad}}, \quad (3.1e)$$

$$\bar{p}_o = \bar{p}(t_o) \text{ in } H, \quad \bar{u} \in U_{\text{ad}}. \quad (3.1f)$$

Die Umkehrung gilt ebenfalls: Lösen $\bar{x} = (\bar{y}, \bar{u})$ und (\bar{p}, \bar{p}_o) das System (3.1), so ist \bar{x} die optimale Lösung von **(QP)**.

Bemerkung 3.2. Wir bezeichnen (3.1a)-(3.1b) als *Zustandsgleichung*, (3.1c)-(3.1d) als *duale* oder *adjungierte Gleichung* und (3.1e) als *Variationsungleichung*. Die duale Gleichung ist die Variationsformulierung des folgenden Anfangs-Randwertproblems für \bar{p}

$$-\bar{p}_t(t, \mathbf{x}) - \operatorname{div}(\lambda(\mathbf{x}) \nabla \bar{p}(t, \mathbf{x})) = 0, \quad (t, \mathbf{x}) \in Q,$$

$$\frac{\partial \bar{p}}{\partial n}(t, \mathbf{x}) + q(\mathbf{x}) \bar{p}(t, \mathbf{x}) = 0, \quad (t, \mathbf{x}) \in \Sigma,$$

$$\bar{p}(t_f, \mathbf{x}) = y_d(\mathbf{x}) - \bar{y}(t_f, \mathbf{x}), \quad \mathbf{x} \in \Omega.$$

Der Lagrange-Multiplikator \bar{p} löst also eine Wärmeleitungsgleichung, die von $t = t_f$ nach $t = t_o$ läuft. Da für \bar{p} aber an $t = t_f$ Anfangsdaten vorliegen, ist das Anfangs-Randwertproblem für \bar{p} wohldefiniert. \diamond

In dem Fall, wenn keine Ungleichungsrestriktionen an die Steuerung vorliegen, das heißt, wir haben $u_a = -\infty$ und $u_b = \infty$, folgt aus der Variationsungleichung (3.1e) die Gleichung

$$\gamma \bar{u}(t, \mathbf{x}) - \bar{p}(t, \mathbf{x}) = 0 \quad \text{für alle } (t, \mathbf{x}) \in \Sigma. \quad (3.1e')$$

Wegen $\gamma > 0$ können wir daher die optimale Steuerung \bar{u} in (3.1) durch den Term \bar{p}/γ auf Σ ersetzen. Das kann in der Numerik genutzt werden, um lineare Gleichungssysteme geringerer Dimension zu erhalten. Wir werden uns

im Folgenden zur Vereinfachung der Darstellung auf das Optimalitätssystem mit $u_a = -\infty$ und $u_b = \infty$ beschränken.

3.2. Die Diskretisierung des Optimalitätssystems

Worum geht es? Wir wollen eine Diskretisierung des Optimalitätssystems durchführen. Dabei verwenden wir, dass wir bereits in Abschnitt 2.2 die Zustandsgleichung (3.1a)-(3.1b) betrachtet haben. Es geht also in diesem Abschnitt um eine Behandlung der dualen Gleichung (3.1c)-(3.1d) und der Gleichung (3.1e'). Für weitere Literatur verweisen wir auf die Arbeiten [41, 44]. Bezüglich der numerischen Verfahren für Optimierungsaufgaben für partielle Differentialgleichungen verweisen wir auf [13, 42].

In Abschnitt 2.2 haben wir die linear-unabhängigen Finite-Element-Funktionen $\{\varphi_i\}_{i=1}^m$ in den m -dimensionalen Ansatzraum V^h eingeführt. Sei $\bar{y}^h(t) \in V^h$ die Lösung von (2.7) für die Steuerung $u = \bar{u}$. Dann können wir analog zu (2.7) eine FE-Galerkin-Diskretisierung für die duale Gleichung betrachten: Gesucht ist $\bar{p}^h(t) \in V^h$ für alle $t \in [t_o, t_f]$ mit

$$\begin{aligned} -\frac{d}{dt} \langle \bar{p}^h(t), \varphi_i \rangle_H + a(\bar{p}^h(t), \varphi_i) &= 0, \quad 1 \leq i \leq m, \quad t \in (t_o, t_f), \\ \langle \bar{p}^h(t_f), \varphi_i \rangle_H &= \langle y_d - \bar{y}^h(t_f), \varphi_i \rangle_H, \quad 1 \leq i \leq m. \end{aligned} \quad (3.2)$$

Wegen $\bar{y}^h(t) \in V^h$ und $\bar{p}^h(t) \in V^h$ für alle $t \in [t_o, t_f]$ schließen wir, dass es FE-Koeffizienten $\bar{y}^h(t) = (\bar{y}_1^h(t), \dots, \bar{y}_m^h(t))^\top$ beziehungsweise $\bar{p}^h(t) = (\bar{p}_1^h(t), \dots, \bar{p}_m^h(t))^\top$ gibt für $t \in [t_o, t_f]$ mit

$$\bar{y}^h(t, \mathbf{x}) = \sum_{j=1}^m \bar{y}_j^h(t) \varphi_j(\mathbf{x}), \quad \bar{p}^h(t, \mathbf{x}) = \sum_{j=1}^m \bar{p}_j^h(t) \varphi_j(\mathbf{x}) \quad \text{für } (t, \mathbf{x}) \in \bar{Q}. \quad (3.3)$$

Setzen wir (3.3) in (3.2) ein, so erhalten wir das folgende m -dimensionale Anfangswertsystem gewöhnlicher Differentialgleichungen:

$$\begin{aligned} -M^h(\bar{p}^h)'(t) + A^h \bar{p}^h(t) &= 0 & \text{für } t \in [t_o, t_f), \\ M^h \bar{p}^h(t_f) &= \bar{y}_d^h - M^h \bar{y}_n^h, \end{aligned} \quad (3.4)$$

wobei wir die FE-Matrizen M^h und A^h bereits in Abschnitt 2.2 definiert haben. Ferner sei $\bar{y}_d^h \in \mathbb{R}^m$ der Vektor mit dem Komponenten

$$y_{d,i}^h = \int_{\Omega} y_d \varphi_i \, d\mathbf{x}, \quad 1 \leq i \leq m.$$

Wenn wir das in Abschnitt 2.2 eingeführte implizite Eulerverfahren verwenden, um das Differentialgleichungssystem (3.4) zu lösen, so erhalten wir das in Algorithmus 3.1 beschriebene Verfahren. Wie in Algorithmus 2.1 müssen wieder 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

Algorithm 3.1 (FE-Galerkin mit impliziten Eulerverfahren für (3.1c)-(3.1d))

Require: $\{\delta t_j\}_{j=1}^n$, M^h , A^h , \bar{y}_d^h , \bar{y}_n^h , $\{\varphi_i\}_{i=1}^m$.

1: Bestimme \bar{p}_n^h aus der Gleichung $M^h \bar{p}_n^h = \bar{y}_d^h - M^h \bar{y}_n^h$.

2: **for** $j = n - 1$ **to** 1 **do**

3: Berechne $\bar{p}_j^h = ((\bar{p}_j^h)_1, \dots, (\bar{p}_j^h)_m)^\top \in \mathbb{R}^m$ aus der Gleichung

$$(M^h + \delta t_j A^h) \bar{p}_j^h = M^h \bar{p}_{j+1}^h.$$

4: Setze $\bar{p}_j^h = \sum_{i=1}^m (\bar{p}_j^h)_i \varphi_i \in V^h$.

5: **end for**

6: **return** $P = [\bar{p}_1^h | \dots | \bar{p}_n^h]$ und $\{\bar{p}_j^h\}_{j=1}^n$.

die Folge $\{\bar{p}_j^h\}_{j=1}^n$ eindeutig bestimmt. Wir erhalten somit eine FE-Galerkin-Approximation für die exakte Lösung von (3.1c)-(3.1d)

$$\bar{p}(t_j, \mathbf{x}) \approx \bar{p}_j^h(\mathbf{x}) = \sum_{i=1}^m P_{ij} \varphi_i(\mathbf{x}) \quad \text{für } 1 \leq j \leq n \text{ und } \mathbf{x} \in \bar{\Omega}. \quad (3.5)$$

Dabei ist $P = [\bar{p}_1^h | \dots | \bar{p}_n^h] \in \mathbb{R}^{m \times n}$ die durch Algorithm 3.1 berechnete FE-Koeffizientenmatrix, die in der j -ten Spalte die FE-Koeffizientenvektoren $\bar{p}_j^h \in \mathbb{R}^m$ von \bar{p}_j^h für $j = 1, \dots, n$ enthält.

Nun kommen wir zur Diskretisierung von (3.1e'). Aufgrund der eingeführten Zeitdiskretisierung bestimmen wir auch nur Approximativ für die optimale Steuerung \bar{u} an den diskreten Zeiten $\{t_j\}_{j=1}^n$, die wir mit $\{\bar{u}_j^h\}_{j=1}^n$ bezeichnen, wobei für jedes $j \in \{1, \dots, n\}$ die diskrete optimale Steuerung \bar{u}_j^h eine reellwertige Funktion auf dem Rand Γ ist. Seien $i_1, \dots, i_{m_\Gamma} \in \{1, \dots, m\}$ die Indizes, für die die FE-Funktionen $\varphi_{i_k} \neq 0$ sind auf Γ . Aus der Gleichung (3.1e') erhalten wir dann

$$\gamma \int_{\Omega} \bar{u}_j^h \varphi_{i_k} \, d\mathbf{x} = \frac{1}{\gamma} \int_{\Omega} \bar{p}_j^h \varphi_{i_k} \, d\mathbf{x} = \sum_{l=1}^m (\bar{p}_j^h)_l \int_{\Gamma} \varphi_l \varphi_{i_k} \, d\mathbf{x} \quad (3.6)$$

für $1 \leq k \leq m_\Gamma$ und $1 \leq j \leq n$. Um (3.6) kompakter zu schreiben, definieren wir zunächst die FE-Matrix $B^h \in \mathbb{R}^{m \times m_\Gamma}$ mit den Komponenten

$$B_{lk}^h = \int_{\Gamma} \varphi_l \varphi_{i_k} \, d\mathbf{x}, \quad 1 \leq l \leq m, \quad 1 \leq k \leq m_\Gamma.$$

Ferner machen wir für die diskrete optimale Steuerung $\{\bar{u}_j^h\}_{j=1}^n$ den Ansatz

$$\bar{u}_j^h(\mathbf{x}) = \sum_{k=1}^{m_\Gamma} (\bar{u}_j^h)_k \varphi_{i_k}(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma \text{ und } j = 1, \dots, n$$

mit dem FE-Koeffizientenvektor $\bar{u}_j^h = ((\bar{u}_j^h)_1, \dots, (\bar{u}_j^h)_{m_\Gamma})^\top$. Dann erhalten wir die Beziehung

$$\sum_{l=1}^m (\bar{p}_j^h)_l \int_{\Gamma} \varphi_l \varphi_{i_k} \, d\mathbf{x} = (B^{h,\top} \bar{p}_j^h)_k, \quad 1 \leq k \leq m_\Gamma,$$

wobei wir $B^{h,\top} = (B^h)^\top \in \mathbb{R}^{m_\Gamma \times m}$ gesetzt haben. Schließlich sei $M_b^h \in \mathbb{R}^{m_\Gamma \times m_\Gamma}$ die reduzierte FE-Massematrix mit den Komponenten

$$(M_\Gamma^h)_{jk} = \int_\Gamma \varphi_{i_j} \varphi_{i_k} \, d\mathbf{x}, \quad 1 \leq j, k \leq m_\Gamma.$$

Dann lässt sich (3.6) in Matrixschreibweise

$$\gamma M_\Gamma^h \vec{u}_j^h - B^{h,\top} \vec{p}_j^h \quad \text{für } 1 \leq j \leq n$$

ausdrücken. Für den in Abschnitt 2.2 definierten FE-Vektor \vec{b}^h gilt die Beziehung

$$\vec{b}^h(\vec{u}_j^h) = B^h \vec{u}_j^h \quad \text{für } 1 \leq j \leq n.$$

Zusammenfassend erhalten wir das folgende gekoppelte, diskrete Optimalitätssystem

$$\begin{aligned} M^h \vec{y}_1^h &= \vec{y}_o^h, \\ -M^h \vec{y}_{j-1}^h + (M^h + \delta t_j A^h) \vec{y}_j^h - \delta t_j B^h \vec{u}_j^h &= \delta t_j \vec{f}_j^h, \quad 2 \leq j \leq n, \\ M^h \vec{y}_n^h + M^h \vec{p}_n^h &= \vec{y}_d^h, \\ (M^h + \delta t_j A^h) \vec{p}_j^h - M^h \vec{p}_{j+1}^h &= 0, \quad 1 \leq j \leq n-1, \\ \gamma M_\Gamma^h \vec{u}_j^h - B^{h,\top} \vec{p}_j^h &= 0, \quad j = 1, \dots, n, \end{aligned} \tag{3.7}$$

bei dem die FE-Koeffizientenvektoren $\{\vec{y}_j^h\}_{j=1}^n \subset \mathbb{R}^m$, $\{\vec{u}_j^h\}_{j=1}^n \subset \mathbb{R}^{m_\Gamma}$ und $\{\vec{p}_j^h\}_{j=1}^n \subset \mathbb{R}^m$ gesucht sind. Mit diesen Vektoren lassen sich dann die FE-Funktionen

$$\begin{aligned} \vec{y}_j^h(\mathbf{x}) &= \sum_{i=1}^m (\vec{y}_j^h)_i \varphi_i(\mathbf{x}), \quad \vec{p}_j^h(\mathbf{x}) = \sum_{i=1}^m (\vec{p}_j^h)_i \varphi_i(\mathbf{x}) \quad \text{für } \mathbf{x} \in \bar{\Omega}, \quad 1 \leq j \leq n, \\ \vec{u}_j^h(\mathbf{x}) &= \sum_{k=1}^{m_\Gamma} (\vec{u}_j^h)_k \varphi_{i_k}(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma, \quad 1 \leq j \leq n, \end{aligned}$$

berechnen. Es sei darauf hingewiesen, dass (3.7) ein lineares Gleichungssystem der Dimension $2m + m_\Gamma$ ist.

3.3. Die POD-Galerkin-Approximation des Optimalitätssystems

Worum geht es? In diesem Abschnitt wollen wir mit den Techniken aus den Abschnitten 2.3 und 2.4 eine POD-Galerkin-Approximation für das Optimalitätssystem (3.7) angeben. Wir erhalten auf diese Weise ein reduziertes Modell, welches gegenüber (3.7) weniger Freiheitsgrade besitzt. Das ermöglicht eine schnellere numerische Lösung des Optimalsteuerproblems.

Ein reduziertes System für (3.7) erhalten wir analog zum Vorgehen in Abschnitt 2.4. Wir gehen davon aus, dass $\{psi_i\}_{i=1}^\ell$ eine POD-Basis vom Rang ℓ . Diese kann zum Beispiel wie in Abschnitt 2.3 beschrieben aus Snapshots $\{y_j^h\}_{j=1}^n \subset V^h$ berechnet werden, wobei die y_j^h 's FE-Lösungen sind, die mit Algorithmus 2.1 mit einer speziellen Wahl $u = u_{\text{ref}}$ generiert werden. Zu der POD-Basis haben wir in Abschnitt 2.2 eine FE-Koeffizientenmatrix $\Psi \in \mathbb{R}^{m \times \ell}$ eingeführt. Mit dieser Matrix ist in Abschnitt 2.4 ein reduziertes

Modell hergeleitet worden. Wir gehen hier analog vor, indem wir eine POD-Galerkin-Approximation für den Zustand- und die duale Variable verwenden, die Steuerung allerdings nicht reduzieren:

$$\bar{y}_j^\ell(\mathbf{x}) = \sum_{i=1}^{\ell} (\bar{y}_j^\ell)_i \psi_i(\mathbf{x}), \quad \bar{p}_j^\ell(\mathbf{x}) = \sum_{i=1}^{\ell} (\bar{p}_j^\ell)_i \psi_i(\mathbf{x}) \quad \text{für } \mathbf{x} \in \bar{\Omega}, \quad 1 \leq j \leq n. \quad (3.8)$$

Mit dem Galerkinansatz (3.8) erhalten wir aus (3.7) nach Projektion auf die POD-Basis das System

$$\begin{aligned} M^\ell \bar{y}_1^\ell &= \bar{y}_o^\ell, \\ -M^\ell \bar{y}_{j-1}^\ell + (M^\ell + \delta t_j A^\ell) \bar{y}_j^\ell - \delta t_j B^\ell \bar{u}_j^\ell &= \delta t_j \bar{f}_j^\ell, \quad 2 \leq j \leq n, \\ M^\ell \bar{y}_n^\ell + M^\ell \bar{p}_n^\ell &= \bar{y}_d^\ell, \\ (M^\ell + \delta t_j A^\ell) \bar{p}_j^\ell - M^h \bar{p}_{j+1}^\ell &= 0, \quad 1 \leq j \leq n-1, \\ \gamma M_\Gamma^h \bar{u}_j^\ell - B^{\ell, \top} \bar{p}_j^\ell &= 0, \quad j = 1, \dots, n. \end{aligned} \quad (3.9)$$

wobei die Matrizen M^ℓ , A^ℓ und die Vektoren \bar{y}_o^ℓ , $\{\bar{f}_j^\ell\}_{j=1}^n$ bereits in Abschnitt 2.4 definiert worden sind. Ferner haben wir in (3.9)

$$B^\ell = \Psi B^h \in \mathbb{R}^{\ell \times m_\Gamma} \quad \text{und} \quad \bar{y}_d^\ell = \Psi \bar{y}_d^h \in \mathbb{R}^\ell$$

gesetzt. Offenbar ist (3.9) ein lineares Gleichungssystem der Dimension $2\ell + m_\Gamma$, bei dem die POD-Koeffizientenvektoren $\{\bar{y}_j^\ell\}_{j=1}^n \subset \mathbb{R}^\ell$, $\{\bar{u}_j^\ell\}_{j=1}^n \subset \mathbb{R}^{m_\Gamma}$ und $\{\bar{p}_j^\ell\}_{j=1}^n \subset \mathbb{R}^\ell$ gesucht sind.

In [46] und [36, Theorem 4.17] werden A-Priori-Fehlerabschätzungen für POD-Galerkin-Approximationen angegeben. Es ist möglich, bei geeigneter Wahl der Snapshots für die POD-Basisberechnung Raten zu erhalten, die sich mit Hilfe der abklingenden POD-Eigenwerte $\{\lambda_i\}_{i=\ell+1}^d$ ausdrücken lassen. Allerdings ist dafür die Kenntnis der Lösung von (3.7) erforderlich, was aus praktischer Sicht eine unrealistische Annahme ist. Es sind daher Aufdatierungsstrategien für die POD-Basis notwendig, da ein zu Anfang gewähltes u_{ref} häufig zu einer POD-Basis führt, die keine hinreichend genaue POD-Lösung zulässt. Wir verweisen hier auf die Arbeiten [36, 86, 93], insbesondere auch für numerische Beispiele. Aufdatierungsstrategien für POD-Basen findet man in zum Beispiel [2, 6, 58, 75, 82].

Zum Schluss wollen wir noch auf eine A-Posteriori-Fehlerabschätzung eingehen, mit der es möglich ist, den Fehler zu einer berechneten suboptimalen Steuerung $\{\bar{u}_j^h\}_{j=1}^n$ abzuschätzen. Wir bezeichnen mit $\{\bar{u}_h^j\}_{j=1}^n$ die Koeffizientenvektoren der suboptimalen Steuerung \bar{u}_j^h . Das Vorgehen ist wie folgt:

- 1) Bestimme $\{\bar{y}_j^h\}_{j=1}^n \subset \mathbb{R}^m$ mit Algorithmus 2.1 für die Wahl $u(t_j) = u_j^h$, $j = 1, \dots, n$.
- 2) Berechne $\{\bar{p}_j^h\}_{j=1}^n \subset \mathbb{R}^m$ mit Algorithmus 3.1 unter Verwendung des Vektors \bar{y}_n^h aus 1).
- 3) Setze $\bar{\zeta}_j^h = \gamma M_\Gamma^h \bar{u}_j^h - B^{h, \top} \bar{p}_j^h \in \mathbb{R}^{m_\Gamma}$ für $j = 1, \dots, n$.

Dann erhalten wir für den Fehler [86, 90]

$$\sum_{j=1}^n \alpha_j \|\vec{u}_j^h - \vec{u}_j^\ell\|_U^2 \leq \frac{1}{\gamma^2} \sum_{j=1}^n \alpha_j \|\vec{\zeta}_j^h\|_U^2$$

mit den Normen

$$\begin{aligned} \|\vec{u}_j^h - \vec{u}_j^\ell\|_U^2 &= \sum_{k=1}^{m_\Gamma} \sum_{l=1}^{m_\Gamma} (\vec{u}_j^h - \vec{u}_j^\ell)_k (\vec{u}_j^h - \vec{u}_j^\ell)_l \int_\Gamma \varphi_{i_k} \varphi_{i_l} \, d\mathbf{x}, \\ \|\vec{\zeta}_j^h\|_U^2 &= \sum_{k=1}^{m_\Gamma} \sum_{l=1}^{m_\Gamma} (\vec{\zeta}_j^h)_k (\vec{\zeta}_j^h)_l \int_\Gamma \varphi_{i_k} \varphi_{i_l} \, d\mathbf{x}. \end{aligned}$$

4. Literaturhinweise auf numerische Beispiele

Wir wollen in diesem Abschnitt ein paar POD-Anwendungen auflisten mit dem Literaturhinweis auf numerische Testbeispiele dazu. Die angeführten Referenzen erheben keinen Anspruch auf Vollständigkeit.

- nichtlineare Wärmeleitungsgleichung [25, 64],
- Phasen-Feldmodell [91],
- Laseroberflächenhärtung [47],
- Helmholtzgleichung für die Innenraumakustik [63, 92, 95],
- Batteriemodelle einer Lithium-Ion-Zelle [16, 61, 62],
- Maxwellgleichung [68],
- Navier-Stokes-Gleichung [2, 6, 28, 43, 45, 66, 67, 75]
- Optionspreisberechnung [79, 80, 82]
- Gleichungen mit zweiten Zeitableitungen [40]
- Feedbackstrategien [3, 53, 57, 64]

Literatur

- [1] R.A. Adams. *Sobolev Spaces*. Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London, 1975.
- [2] K. Afanasiev and M. Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. *Lecture Notes in Pure and Applied Mathematics*, 216:317-332, 2001.
- [3] A. Alla und S. Volkwein. Asymptotic stability of POD based model predictive control for a semilinear parabolic PDE. Eingereicht, 2013.
- [4] A.C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Advances in Design and Control, SIAM, Philadelphia, 2005.
- [5] W. Arendt und K. Urban. *Partielle Differenzialgleichungen. Eine Einführung in analytische und numerische Methoden*. Spektrum Akademischer Verlag, Heidelberg, 2010.
- [6] E. Arian, M. Fahl, and E.W. Sachs. Trust-region proper orthogonal decomposition for flow control. Technical Report 2000-25, ICASE, 2000.

- [7] P. Astrid, S. Weiland, K. Willcox, and T. Backx. Missing point estimation in models described by proper orthogonal decomposition. Proceedings 43rd IEEE Conference on Decision and Control, December, 2004.
- [8] J.A. Atwell, J.T. Borggaard und B.B. King. Reduced-order controllers for Burgers' equation with a nonlinear observer. *Int. J. Appl. Math. Comp. Sci.*, 11:1311-1330, 2001.
- [9] M. Barrault, Y. Maday, N.C. Nguyen and A.T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667-672, 2004.
- [10] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Appl. Numer. Math.*, 43:9-44, 2002.
- [11] H.T. Banks, M.L. Joyner, B. Winchesky und W.P. Winfree. Nondestructive evaluation using a reduced-order computational methodology . *Inverse Problems*, 16:1-17, 2000.
- [12] P. Benner, V. Mehrmann, and D.C. Sorensen (eds.). *Dimension Reduction of Large-Scale Systems*. Lecture Notes in Computational Science and Engineering, Springer-Verlag Berlin, 2005.
- [13] A. Borzi und V. Schulz. *Computational Optimization of Systems Governed by Partial Differential Equations*. Computational Sciences and Engineering, SIAM, Philadelphia, 2012.
- [14] S.C. Brenner und L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics, vol. 15, Springer, 2008.
- [15] T. Bui-Thanh. *Model-Constrained Optimization Methods for Reduction of Parameterized Systems*. Ph.D. thesis, MIT, USA, 2007.
- [16] L. Cai und R. E. White. Reduction of model order based on proper orthogonal decomposition for lithium-ion battery simulations. *Journal of The Electrochemical Society*, 156(3):A154-A161, 2009.
- [17] C. Canuto, T. Tonn und K. Urban. A-posteriori error analysis of the reduced basis method for non-affine parameterized nonlinear pde's *SIAM Journal on Numerical Analysis*, 47:2001-2022, 2009.
- [18] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.
- [19] S. Chaturantabut und D.C. Sorensen. Application of POD and DEIM on a dimension reduction of nonlinear miscible viscous fingering in porous media. *Technical Report*, TR09-25, RICE University, 2009.
- [20] S. Chaturantabut und D.C. Sorensen. A state space estimate for POD-DEIM nonlinear model reduction. *SIAM Journal on Numerical Analysis*, 50:46-63, 2012.
- [21] S. Chaturantabut und D.C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737-2764, 2010.
- [22] W. Dahmen und A. Reusken. *Numerik fr Ingenieure und Naturwissenschaftler*. Zweite, korrigierte Ausgabe, Springer, Heidelberg, 2008.
- [23] R. Dautray und J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology. Volume 5: Evolution Problems I*. Springer, Berlin, 2000.

- [24] L. Dede. Reduced basis method and a posteriori error estimation for parametrized linear-quadratic optimal control problems. *SIAM Journal on Scientific Computing*, 32:997-1019, 2010.
- [25] F. Diwoky and S. Volkwein. Nonlinear boundary control for the heat equation utilizing proper orthogonal decomposition. In K.-H. Hoffmann, R. H. W. Hoppe, V. Schulz, editors, *Fast Solution of Discretized Optimization Problems*, International Series of Numerical Mathematics, 138:73-87, 2001.
- [26] G. Dzuik. *Theorie und Numerik partieller Differentialgleichungen*. Walter de Gruyter, Berlin, 2010.
- [27] L.C. Evans. *Partial Differential Equations*. American Math. Society, Providence, Rhode Island, 2008.
- [28] M. Fahl. Computation of POD basis functions for fluid flows with Lanczos methods. *Math. Comput. Modelling*, 34:91-107, 2001.
- [29] R.W. Freund. Krylov-subspace methods for reduced-order modelling in circuit simulation. *J. Comput. Appl. Math.* 123:395-421, 2000.
- [30] P. Feldmann und R.W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design*, 14:639-649, 1995.
- [31] K. Fukunaga. *Introduction to Statistical Recognition*. NewYork, Academic Press, 1990.
- [32] K. Gallivan, E. Grimme und P. Van Dooren. Asymtotic waveform evaluation via the Lanczos method. *Appl. Math. Lett.*, 7:75-80, 1994.
- [33] M. Green und D.J.N. Limebeer. *Linear Robust Control*. Prentice-Hall, Englewood Cliffs, NJ, 1995.
- [34] M. Grepl and M. Kärcher. A posteriori error estimation for reduced order solutions of parametrized parabolic optimal control problems. Submitted, 2013.
- [35] M.A. Grepl, Y. Maday, N.C. Nguyen and A.T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Math. Model. Numer. Anal.*, 41:575-605, 2007.
- [36] M. Gubisch und S. Volkwein. *Proper Orthogonal Decomposition for Linear-Quadratic Optimal Control*. Eingereicht, 2013.
<http://nbn-resolving.de/urn:nbn:de:bsz:352-250378>
- [37] S. Gugercin. *Projection methods for model reduction of large-scale dynamical systems*. Ph.D. thesis, Rice University, Houston, 2003.
- [38] B. Haasdonk und M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42:277-302, 2008.
- [39] M. Heinkenschloss, D.C. Sorensen und K. Sun. Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations. *SIAM J. Sci. Comput.*, 30:1038-1063, 2008.
- [40] S. Herkt, M. Hinze, and R. Pinnau. *Convergence analysis of Galerkin POD for linear second order evolution equations*. *Electronic Transactions on Numerical Analysis*, 40:321-337, 2013.
- [41] M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *Computational Optimization and Applications*, 30:45-61, 2005.

- [42] M. Hinze, R. Pinnau, M. Ulbrich und S. Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.
- [43] M. Hinze und K. Kunisch. Three control methods for time-dependent fluid flow. *Flow, Turbulence and Combustion*, 65:273-298, 2000.
- [44] M. Hinze und F. Tröltzsch. Discrete concepts versus error analysis in pde constrained optimization. *GAMM-Mitteilungen*, 33:148-162, 2010.
- [45] M. Hinze und S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control. *Lecture Notes in Computational Science and Engineering*, 45:261-306, 2005.
- [46] M. Hinze und S. Volkwein. Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. *Computational Optimization and Applications*, 39:319-345, 2008.
- [47] D. Hömberg und S. Volkwein. Control of laser surface hardening by a reduced-order approach using proper orthogonal decomposition. *Mathematical and Computer Modelling*, 38:1003-1028, 2003.
- [48] P. Holmes, J.L. Lumley, G. Berkooz und C.W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics, Cambridge University Press, second edition, 2012.
- [49] K. Ito und K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control, SIAM, Philadelphia, 2008.
- [50] K. Ito and S.S. Ravindran. A reduced basis method for control problems governed by PDEs. In W. Desch, F. Kappel, and K. Kunisch, eds., *Control and Estimation of Distributed Parameter Systems*. Proceedings of the International Conference in Vorau, 1996, Birkhäuser-Verlag, Basel, 126:153-168, 1998.
- [51] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, Cambridge, 1987.
- [52] E. Kammann, F. Tröltzsch und S. Volkwein. A method of a-posteriori error estimation with application to proper orthogonal decomposition. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47:555-581, 2013.
- [53] K. Kunisch und S. Volkwein. Control of Burgers' equation by a reduced order approach using proper orthogonal decomposition. *Journal on Optimization Theory and Applications*, 102:345-371, 1999.
- [54] K. Kunisch und S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik*, 90:117-148, 2001.
- [55] K. Kunisch und S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40:492-515, 2002.
- [56] K. Kunisch und S. Volkwein. Crank-Nicolson Galerkin proper orthogonal decomposition approximations for a general equation in fluid dynamics. Proceedings of the 18th GAMM Seminar on *Multigrid and related methods for optimization problems*, Leipzig, 97-114, 2002.
- [57] K. Kunisch und S. Volkwein und L. Xie. HJB-POD based feedback design for the optimal control of evolution problems. *SIAM Journal on Applied Dynamical Systems*, 3:701-722, 2004.
- [58] K. Kunisch und S. Volkwein. Proper orthogonal decomposition for optimality systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42:1-23, 2008.

- [59] K. Kunisch und S. Volkwein. Optimal snapshot location for computing POD basis functions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44:509-529, 2010.
- [60] S. Lall, J.E. Marsden und S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. Robust Nonlinear Control*, 12:519-535, 2002.
- [61] O. Lass und S. Volkwein. POD Galerkin schemes for nonlinear elliptic-parabolic systems. *SIAM Journal on Scientific Computing*, 35(3):A1271-A1298, 2013.
- [62] O. Lass and S. Volkwein. Adaptive POD basis computation for parametrized nonlinear systems using optimal snapshot location. Eingereicht, 2013.
<http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-191730>
- [63] O. Lass and S. Volkwein. Parameter identification for nonlinear elliptic-parabolic systems with application in lithium-ion battery modeling. Eingereicht, 2013.
- [64] F. Leibfritz and S. Volkwein. Reduced order output feedback control design for PDE systems using proper orthogonal decomposition and nonlinear semidefinite programming. *Linear Algebra and Its Applications*, 415:542-575, 2006.
- [65] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer, Berlin, 1971.
- [66] D.M. Luchtenburg, B.R. Noack und M. Schlegel. An introduction to the POD Galerkin method for fluid flows with analytical examples and MATLAB source codes. Institut für Strömungsmechanik und Technische Akustik, TU Berlin, 2009.
- [67] H.V. Ly und H.T. Tran. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor. *Quarterly Appl. Math.* 60:631-656, 2002.
- [68] R. Mancini und S. Volkwein. An inverse scattering problem for the time-dependent Maxwell equations: nonlinear optimization and model-order reduction. *Numerical Linear Algebra with Applications*, 20:689-711, 2013.
- [69] V. Mehrmann und T. Stykel. Balanced truncation model reduction for large-scale systems in descriptor form. In *Dimension Reduction of Large-Scale Systems*, Lect. Notes Comput. Sci. Eng. 45, P. Benner, V. Mehrmann, and D.C. Sorensen, eds., Springer, Berlin, 2005.
- [70] F. Negri, G. Rozza, A. Manzoni, and A. Quateroni. Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing*, to appear.
- [71] B. Noble. *Applied Linear Algebra*. Englewood Cliffs, NJ : Prentice-Hall, 1969.
- [72] A.T. Patera und G. Rozza. Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations. MIT Papalardo Graduate Monographs in Mechanical Engineering, 2006.
- [73] R. Pinnau. Model reduction via proper orthogonal decomposition. In *Model Order Reduction: Theory, Research Aspects and Applications*, Mathematics in Industry, vol. 13, pp. 95-109, Springer, 2008.
- [74] M. Rathinam und L. Petzold. Dynamic iteration using reduced order models: a method for simulation of large scale modular systems. *SIAM Journal on Numerical Analysis*, 40:1446-1474, 2002.

- [75] S.S. Ravindran. Reduced-order adaptive controllers for fluid flows using POD. *SIAM Journal on Scientific Computing*, 15:457-478, 2000.
- [76] M. Reed und B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*. Academic Press, New York, 1980.
- [77] C.W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *Int. J. on Bifurcation and Chaos*, 15:997-1013, 2005.
- [78] E.W. Sachs und S. Volkwein. POD Galerkin approximations in PDE-constrained optimization. *Gamm-Mitteilungen*, 33:194-208, 2010.
- [79] E.W. Sachs und M. Schu. Reduced order models (POD) for calibration problems in finance. In K. Kunisch, G. Of und O. Steinbach (Editoren), *Numerical Mathematics and Advanced Applications*, ENUMATH, Springer, Seiten 735-742, 2007.
- [80] E.W. Sachs und M. Schu. Reduced order models in PIDE constrained optimization. *Control and Cybernetics*, 39: 661-675, 2001.
- [81] W.H.A. Schilders, H.A. van der Vorst, and J. Rommes. *Model Order Reduction: Theory, Research Aspects and Applications*. Mathematics in Industry, vol. 13, Springer, 2008.
- [82] M. Schu. *Adaptive Trust-Region POD Methods and their Application in Finance*. Ph.D thesis, University of Trier, 2013.
- [83] J.R. Singler. New POD expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs. Eingereicht, 2013.
- [84] L. Sirovich. Turbulence and the dynamics of coherent structures. Parts I-II. *Quarterly of Applied Mathematics*, XVI:561-590, 1987.
- [85] G. Strang und G.J. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, Eaglewood Cliffs, New Jersey, 1973.
- [86] A. Studinger und S. Volkwein. Numerical analysis of POD a-posteriori error estimation for optimal control. *International Series of Numerical Mathematics*, 164:137-158, 2013
- [87] T. Stykel. Balanced truncation model reduction for semidiscretized Stokes equation. *Linear Algebra Appl.*, 415:262-289, 2006.
- [88] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Springer series in Computational Mathematics, Springer, Berlin, 1997.
- [89] F. Tröltzsch. *Optimal Control of Partial Differential Equations. Theory, Methods and Applications*. American Mathematical Society, Providence, volume 112, 2010.
- [90] F. Tröltzsch and S. Volkwein. POD a-posteriori error estimates for linear-quadratic optimal control problems. *Computational Optimization and Applications*, 44:83-115, 2009.
- [91] S. Volkwein. Optimal control of a phase-field model using proper orthogonal decomposition. *Zeitschrift für Angewandte Mathematik und Mechanik*, 81:83-97, 2001.
- [92] S. Volkwein. Admittance identification from point-wise sound pressure measurements using reduced-order modelling, *Journal of Optimization Theory and Applications*, 14:166-193, 2010.
- [93] S. Volkwein. Optimality system POD and a-posteriori error analysis for linear-quadratic problems. *Control and Cybernetics*, 40:1109-1125, 2011.

- [94] S. Volkwein. *Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling*. Lecture notes, University of Konstanz, Department of Mathematics and Statistics, 2013.
www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/scripts.php
- [95] S. Volkwein and A. Hepberger. Impedance identification by POD model reduction techniques. *at-Automatisierungstechnik*, 8:437-446, 2008.
- [96] K. Willcox und J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *American Institute of Aeronautics and Astronautics (AIAA)*, 2323-2330, 2002.
- [97] K. Zhou, J.C. Doyle und K. Glover. *Robust and Optimal Control*. Prentice-Hall, Upper Saddle River, NJ, 1996.
- [98] O.C. Zienkiewicz. *The Finite Element Method in Engineering Scienc.* McGraw-Hill, London, New York, 1977.

Stefan Volkwein
University of Konstanz
Department of Mathematics and Statistics
Universitätsstraße 10
D-78457 Konstanz
Germany
e-mail: Stefan.Volkwein@uni-konstanz.de