Schnelle Algorithmen für die Zustands- und Parameterschätzung auf bewegten Horizonten

Fast Algorithms for State and Parameter Estimation on Moving Horizons

Moritz Diehl, Peter Kühl, Hans Georg Bock und Johannes P. Schlöder

Wir behandeln das Problem einer gleichzeitigen Zustands- und Parameterschätzung auf bewegtem Horizont (Moving Horizon Estimation, MHE). Wir schlagen eine MHE Formulierung für differential-algebraische Systemmodelle vor, die deterministisches Systemverhalten auf dem Schätzhorizont annimmt und eine Regularisierung des Anfangswertes verwendet. Zu ihrer numerischen Lösung stellen wir effiziente Methoden zur schnellen On-line-Lösung der MHE Probleme vor, die auf dem Mehrzielverfahren für die Parameterschätzung beruhen. In der vorgeschlagenen MHE Echtzeititeration wird zu jedem MHE Optimierungsproblem nur eine einzige Optimierungsiteration durchgeführt, und die aufeinander folgenden Probleme werden durch ein Verschieben der bereits geschätzten Trajektorie effizient initialisiert. Die Methode wird auf das Problem zweier thermisch gekoppelter Destillationskolonnen angewandt und mit einem erweiterten Kalman-Filter verglichen. Neben dem Systemzustand aus über 400 Variablen werden drei veränderliche Betriebsparameter mitgeschätzt. Die Rechenzeiten für dieses schwierige Problem liegen bei rund 5 Sekunden pro MHE Optimierung.

We treat the problem of moving horizon state and parameter estimation (MHE) and propose a formulation for differential-algebraic systems that assumes a deterministic system behaviour on the estimation horizon, and which uses a regularization of the initial value. We present an efficient method for fast online solution of the subsequent MHE problems. The method is based the multiple shooting method for parameter estimation. The proposed MHE real-time iteration scheme performs only a single optimization iteration for each new MHE optimization problem. Subsequent problems are initialized via a shift of the best available trajectory estimate. The method is applied to the problem of two thermally coupled distillation columns and compared to an extended Kalman Filter. In addition to the system state of over 400 variables we also estimate three varying feedflow parameters. The algorithm uses a CPU time of around 5 seconds per MHE optimization for this challenging application.

Schlagwörter: Mehrzielverfahren, Parameterschätzung, Zustandsschätzung, nichtlineare Modellbasierte Prädiktive Regelung, differential-algebraische Gleichungen

Keywords: Multiple shooting, parameter estimation, state estimation, nonlinear model predictive control, differential algebraic equations

1 Einführung

Bei jeder Anwendung Modellbasierter Prädiktiver Regelungsmethoden (Model Predictive Control, MPC) sind in Echtzeit zwei Problemstellungen zu addressieren: zum ersten das Schätzproblem, das aus den vorliegenden Mess-

daten den aktuellen Systemzustand und eventuell gestörte Systemparameter zu bestimmen sucht, und zum zweiten das Optimalsteuerungsproblem, das aus den aktuellen Informationen das zukünftige Prozessverhalten simuliert und optimiert. Beide Problemstellungen können in Form von On-line-Optimierungsproblemen formuliert werden, deren

Lösung für gegebene Problemdaten in Echtzeit berechnet werden muss. Aufgrund der dafür notwendigen Rechenzeiten wurde die nichtlineare MPC (NMPC) - die, im Gegensatz zur bereits weitverbreiteten linearen MPC, auf ersten Prinzipien beruhende, nichtlineare Prozessmodelle für die Schätzung und Prädiktion verwendet - noch vor wenigen Jahren als ein schwer zu realisierendes und nur aus theoretischer Sicht interessantes Konzept angesehen, und es gab lange Zeit nur wenige echte NMPC Anwendungen [34]. In den letzten Jahren sind jedoch gewaltige Fortschritte auf algorithmischer Seite erzielt worden, die mit den hohen Rechenleistungen heutiger CPUs eine Anwendung der NMPC auch auf komplexe, durch nichtlineare Differentialgleichungen beschriebene Prozessmodelle möglich macht. Der Vorteil im Vergleich zu linearer MPC liegt in der höheren prädiktiven Genauigkeit von auf ersten Prinzipien beruhenden Modellen, die sich besonders in transienten Phasen und bei Anfahr-Manövern auszahlt. Zudem bietet NMPC die Möglichkeit, im Rahmen eines "optimizing control" direkt ökonomische Kriterien in die Zielfunktion einfließen zu lassen, und so von einem sturen Einregeln gegebener Sollwerte weg zu kommen, hin zu einer größeren Flexibilität, die z.B. auch erlaubt, günstige Störungen auszunutzen, statt sie zu unterdrücken. Die heute für NMPC Anwendungen möglichen Zeitskalen hängen stark von der Problemformulierung ab, und variieren von beispielsweise 10 Millisekunden CPU Zeit pro Optimierungsproblem bei nichtlinearen gewöhnlichen Differentialgleichungsmodellen (ODE) der Ordnung 5 [18;42], über z.B. 5 Sekunden für mittelgroße differential-algebraische Prozessmodelle (DAE) der Ordnung 250 [39], bis hin zu Rechenzeiten von Minuten bei der Prädiktion mit komplexen partiellen Differentialgleichungsmodellen (PDE) [43]. Eine gute Übersicht der gebräuchlichsten numerischen Verfahren zur Optimierung auf bewegtem Horizont wird beispielsweise in Binder et al. [7] gegeben. Exzellente Zusammenfassungen des Standes der Forschung und der Anwendungen von NMPC werden auch in den Monographien [1; 20] gegeben.

In allen NMPC Anwendungen gibt es neben der effizienten Lösung der Optimalsteuerungsprobleme auf dem in der Zukunft liegenden Horizont noch das Problem der Zustandsund Parameterschätzung aus vergangenen Messdaten. Hierfür gibt es unterschiedliche Methoden. Klassischerweise wird das erweiterte Kalman-Filter genannt [5], denkbar sind auch Normalformbeobachter [6] und so genannte High-Gain-Beobachter [44], um nur einige zu nennen. Eine weitere Methode ist der "Moving Horizon Estimator (MHE)", der eine große Ähnlichkeit zum NMPC-Ansatz aufweist, da ein Ausgleichsproblem auf einem sich mitbewegenden, in der Vergangenheit liegenden Zeithorizont gelöst wird, um eine Schätzung für die aktuellen Systemzustände und Parameter zu erhalten [37]. Den Hauptvorteil eines nichtlinearen MHE sehen wir in der Tatsache, dass neben den Zuständen auch Störungen in Form von unbekannten, aber sich nur langsam ändernden Parametern mitgeschätzt werden können [33]. Dies ist der aus unserer Sicht konsistenteste Ansatz, um in NMPC-Reglern z. B. einen Integralanteil zur Ausregelung bleibender Regelabweichungen zu realisieren. Ohne gute Modellkalibrierung in Echtzeit kann man keine guten Regelergebnisse von einem nichtlinearen modellbasierten prädiktiven Regler erwarten. Da aber selbst leistungsfähige Rechner an ihre Grenzen stoßen, wenn gleich *zwei* dynamische Optimierungsprobleme on-line gelöst werden müssen, ist die schnelle und effiziente Lösung des On-line-Schätzproblems – ebenso wie die des Optimalsteuerungsproblems – eine Herausforderung für die numerische Mathematik. Sie steht im Zentrum dieses Artikels.

1.1 Aufbau des Artikels

Wir stellen in Abschnitt 2 zunächst die von uns favorisierte Formulierung des MHE Zustands- und Parameterschätzproblems dar. Zur numerischen Lösung dieser Optimierungsprobleme hat sich bereits für den Off-line-Fall - ohne Echtzeitanforderungen - das Mehrzielverfahren [8; 10; 32] als ein besonders robuster und effizienter Ansatz erwiesen, das in Abschnitt 3 skizziert wird. Ein MHE Algorithmus, der die Feedback-Verzögerung zwischen Eintreffen einer neuen Messung und Ausgabe der Schätzantwort minimiert, wird in Abschnitt 4 vorgestellt. Zur Illustration wird der beschriebene MHE Algorithmus in Abschnitt 5 auf das Beispiel zweier thermisch gekoppelter Destillationskolonnen zur Trennung eines Dreistoffgemisches mit unbekannten Betriebsparametern angewandt und mit einem erweiterten Kalman-Filter (EKF) bei gleichen Tuningparametern verglichen. Im letzten Abschnitt 6 geben wir eine kurze Zusammenfassung.

2 MHE Zustands- und Parameterschätzproblem

Im gesamten Artikel gehen wir davon aus, dass ein Prozessmodell in differentiell-algebraischer (DAE) Form vorliegt, das durch den folgenden Satz von Gleichungen beschrieben ist,

$$\dot{x}(t) = f(x(t), z(t), u(t), p),$$

 $0 = g(x(t), z(t), u(t), p),$

mit differentiellen Zuständen $x \in \mathbb{R}^{n_x}$, algebraischen Zuständen $z \in \mathbb{R}^{n_z}$, Stellgrößen $u \in \mathbb{R}^{n_u}$, und sich möglicherweise langsam verändernden Parametern $p \in \mathbb{R}^{n_p}$. Dabei setzen wir voraus, dass die Matrix $\frac{\partial g}{\partial z}$ überall invertierbar ist, die DAE also den Index 1 hat. Wir bemerken, dass DAE mit höherem Index durch Indexreduktion auf diese Form gebracht werden können, und verweisen für genauere Angaben zur Behandlung von DAE auf klassische Lehrbücher [11].

Die Messdaten werden durch eine Funktion $h(x(t), z(t), u(t), p) \in \mathbb{R}^{n_h}$ generiert. Zur aktuellen Zeit, t_K , seien M+1 Punktmessungen $\eta_{K-M}, \ldots, \eta_K \in \mathbb{R}^{n_h}$ zu in der Vergangenheit liegenden Zeitpunkten $t_{K-M} < \ldots < t_K$ gegeben. Dabei ist $T_E = t_K - t_{K-M}$ die Länge des Schätzhorizontes, und wir setzen K-M=:L. Wir nehmen die Messfehler als

zeitlich unabhängig und normalverteilt an, mit Erwartungswert Null und bekannter Kovarianzmatrix $V_k^{-2} \in \mathbb{R}^{n_h \times n_h}$, sodass eine Minimierung der Summe der quadrierten Abweichungen $\|\eta_k - h(x(t_k), z(t_k), u(t_k), p)\|_{V_k}^2$ als Maximum Likelihood Schätzer interpretiert werden kann. Wir verwenden den Ausdruck $\|\cdot\|_V^2$ in diesem Artikel leicht abweichend vom Standard, als $\|x\|_V^2 := \|Vx\|_2^2 = x^T V^T Vx$. Um auch mit dem Nichtvorhandensein bestimmter Messungen zu bestimmten Zeitpunkten flexibel umgehen zu können, nehmen wir die inverse Kovarianzmatrix V_k^2 als messzeitpunktabhängig an. Diese Matrix ist positiv semidefinit und kann auch als Gewichtsmatrix interpretiert werden. Ein Gewicht von Null wird dann vergeben, wenn (noch) keine Messung vorliegt.

Neben den Punktmessungen η_k und Gewichtsmatrizen V_k und den bekannten Stellgrößen u(t) auf dem Schätzhorizont $t \in [t_L, t_K]$ nehmen wir an, dass noch zwei wichtige Formen von Vorwissen vorliegen. Zum ersten einfache Ober- und Untergrenzen der Form $x_{\min} \le x(t) \le x_{\max}$ und $p_{\min} \le p \le p_{\max}$, und zum zweiten eine gemeinsame a-priori Verteilung für $x(t_L)$ und p, die als normalverteilt mit Erwartungswert (\bar{x}_L, \bar{p}_L) und Kovarianzmatrix $P_L^{-2} \in \mathbb{R}^{(n_x+n_p)\times(n_x+n_p)}$ angenommen wird. Eine von uns verwendete Heuristik zur Wahl von (\bar{x}_L, \bar{p}_L) und P_L aus vor dem Schätzhorizont liegenden Messdaten, die so in approximierter Form weiterhin berücksichtigt werden können, stellen wir weiter unten vor. Das Zustands- und Parameterschätzproblem zur Zeit t_K , das wir zu gegebenen Daten η_k , V_k für $k = L, L + 1, \ldots, K$, und u(t) für $t \in [t_L, t_K]$ sowie für (\bar{x}_L, \bar{p}_L) und P_L lösen möchten, hat die folgende Form:

$$\min_{x(t),z(t),p} \left\{ \left\| \begin{matrix} x(t_L) - \bar{x}_L \\ p - \bar{p}_L \end{matrix} \right\|_{P_L}^2 + \sum_{k=L}^K \|\eta_k - h(x(t_k), z(t_k), u(t_k), p)\|_{V_k}^2 \right\}$$
(1)

für $t \in [t_L, t_K]$ unter den Nebenbedingungen

$$\begin{split} \dot{x}(t) &= f(x(t), z(t), u(t), p) \,, \quad t \in [t_L, t_K] \,, \\ 0 &= g(x(t), z(t), u(t), p) \,, \\ x_{\min} &\leq x(t) \leq x_{\max} \,, \end{split}$$

 $p_{\min} \leq p \leq p_{\max}$.

Man beachte, dass einige in der Praxis wichtige Aspekte in dieser Formulierung der Einfachheit halber nicht berücksichtigt werden, z.B. kontinuierlich vorliegende Messungen, nichtlineare Nebenbedingungen, oder auch Prozessrauschen. Diese Aspekte könnten jedoch ohne Probleme mit den in diesem Artikel vorgestellten Algorithmen behandelt werden.

2.1 Anfangsgewichtsbestimmung

Der Grund für die Einführung des Anfangsgewichtes

$$\left\| \begin{array}{c} x(t_L) - \bar{x}_L \\ p - \bar{p}_L \end{array} \right\|_{P_I}^2$$

ist der Wunsch, Information aus der noch weiter als t_L zurückliegenden Vergangenheit im Moving-Horizon Schätzproblem zu berücksichtigen. Aus theoretischer Sicht wäre es ideal, als Anfangsgewicht die so gennante "arrival cost" $\min \sum_{k=-\infty}^L \|\eta_k - h(x(t_k), z(t_k), u(t_k), p)\|_{V_k}^2$ zu nehmen, die jedoch nicht (bzw. nur im Falle linearer Systeme) exakt berechnet werden kann. Zudem wollen wir vermeiden, dass die vergangene und auf unendlich vielen Messungen beruhende Information deshalb ein unendlich hohes Gewicht bekommt.

Die Aufgabe, die wir in jedem Zeitschritt addressieren müssen, ist die folgende. Gegeben seien die Lösung $x^*(\cdot), z^*(\cdot), p^*$ des MHE Problems auf einem Horizont $[t_L, t_K]$ sowie die Daten des alten MHE Problems. Wie bestimmen wir für ein neues MHE Problem auf dem Horizont $[t_{L+1}, t_{K+1}]$ die neuen Anfangsgewichtsdaten $(\bar{x}_{L+1}, \bar{p}_{L+1})$ und P_{L+1} ? Ein manchmal in der Praxis verwendeter Ansatz lässt die Gewichtsmatrix unverändert, $P_{L+1} = P_L$, und setzt die so genannte "geglättete Schätzung" für Anfangswert und Parameter ein, $\bar{x}_{L+1} = x^*(t_{L+1})$ und $\bar{p}_{L+1} = p^*$. Dadurch soll der Anfangswert geeignet regularisiert werden. Wir raten von diesem Vorgehen strikt ab, da es keinerlei Motivation für diesen Ansatz gibt, und, schlimmer noch, es zu einer impliziten Doppelgewichtung der auf dem Intervall $[t_{L+1}, t_K]$ liegenden Messungen im neuen Problem führt.

Stattdessen schlagen wir das folgende Vorgehen vor, das sich an der dynamischen Programmierung – ähnlich dem Kalman-Filter - orientiert und auf einer Linearisierung eines Teilproblems auf dem Intervall $[t_L, t_{L+1}]$ beruht. Zur Motivation des dadurch vollzogenen Überganges vom Horizont $[t_L, t_K]$ auf den Horizont $[t_{L+1}, t_{K+1}]$ betrachten wir zunächst ein hypothetisches Problem auf dem gemeinsamen Horizont $[t_L, t_{K+1}]$, das exakt die gleiche Form wie das Problem (1) hat, aber mit einen um ein Intervall verlängerten Horizont. Idealerweise würden wir gerne dieses verlängerte Problem als nächstes lösen, aber um einen anwachsenden Schätzhorizont zu vermeiden, ist das Ziel, ein diesem möglichst ähnliches Problem auf dem neuen, vorgeschobenen Horizont $[t_{L+1}, t_{K+1}]$ zu formulieren und zu lösen. Die auf dem ersten Intervall $[t_L, t_{L+1}]$ liegende Information, das Anfangsgewicht und die erste Messung, soll also in einem neuen Anfangsgewicht zusammengefasst werden, das am Zeitpunkt t_{L+1} ausgewertet wird. Tatsächlich gäbe es eine Möglichkeit, ein dem verlängerten Problem voll äquivalentes Problem auf dem Horizont $[t_{L+1}, t_{K+1}]$ zu formulieren, wenn wir als neues Anfangsgewicht eine spezielle nichtlineare Funktion $F(x_{L+1}, p_{L+1})$ zulassen, die wir das "ideale Anfangsgewicht" nennen könnten. Der klareren Notation halber benennen wir hier und im Folgenden die Werte $x(t_{L+1})$ und p des neuen Problems – auf dem vorgeschobenen Horizont $[t_{L+1}, t_{K+1}]$ – mit x_{L+1} und p_{L+1} . Das ideale Anfangsgewicht kann mit Hilfe der dynamischen Programmierung wie folgt definiert werden. Bezeichnen wir die Lösung der DAE auf dem Intervall $t \in [t_L, t_{L+1}]$ mit Anfangswert x_L und Parameter p_L als $x(t; x_L, p_L)$, $z(t; x_L, p_L)$, so ergäbe sich das ideale An-

at 12/2006

fangsgewicht als die Lösung

$$F(x_{L+1}, p_{L+1}) = \min_{x_L, p_L} \left\{ \left\| \begin{array}{l} x_L - \bar{x}_L \\ p_L - \bar{p}_L \end{array} \right\|_{P_L}^2 + \left\| \eta_L - h(x_L, z(t_L; x_L, p_L), u(t_L), p_L) \right\|_{V_L}^2 \right\} \quad \text{s.t.}$$

$$0 = x_{L+1} - x(t_{L+1}; x_L, p_L), \quad 0 = p_{L+1} - p_L.$$

Es ist leicht zu sehen, dass diese Formulierung im Laufe vieler Übergänge zu einem immerwährenden Anwachsen des Anfangsgewichtes führen würde, wie bereits oben für die ideale "arrival-cost" bemerkt. Dies ist auf die Abwesenheit von Zustandsrauschen in unserer MHE Formulierung zurückzuführen. Wir schlagen deshalb im ersten Schritt vor, die Gleichungsnebenbedingungen dieses Problems durch einen quadratischen Strafterm zu ersetzen, der einem gaussverteilten Zustands- und Parameterrauschen mit Kovarianzmatrix $W^{-2} \in \mathbb{R}^{(n_x+n_p)\times (n_x+n_p)}$ entspricht. Dadurch ergibt sich ein "modifiziertes ideales Anfangsgewicht" F':

$$F'(x_{L+1}, p_{L+1}) = \min_{x_L, p_L} \left\{ \left\| \begin{array}{l} x_L - \bar{x}_L \\ p_L - \bar{p}_L \end{array} \right\|_{P_L}^2 + \left\| \eta_L - h(x_L, z(t_L; x_L, p_L), u(t_L), p_L) \right\|_{V_L}^2 + \left\| \begin{array}{l} x_{L+1} - x(t_{L+1}; x_L, p_L) \\ p_{L+1} - p_L \end{array} \right\|_{W}^2 \right\}.$$
(2)

Es ist leicht zu zeigen, dass punktweise $F' \leq F$ gilt. Das Zustandsrauschen führt also zu einem Abgewichten des Anfangsgewichtes. Dies ist gewünscht, da wir unserem Modell nicht für unendlich lange Zeithorizonte trauen wollen. Das Ziel der jetzt folgenden Heuristik ist, das "modifizierte ideale Anfangsgewicht" F' durch eine quadratische Funktion der Form

$$F''(x_{L+1}, p_{L+1}) = \text{const} + \left\| \begin{matrix} x(t_{L+1}) - \bar{x}_{L+1} \\ p - \bar{p}_{L+1} \end{matrix} \right\|_{P_{L+1}}^{2}$$

zu approximieren. Da der konstante Term für das MHE Optimierungsergebnis irrelevant ist, wäre damit bereits das angestrebte Ziel erreicht, die Vorinformation des Intervalls $[t_L, t_{L+1}]$ in einem quadratischen Anfangsgewicht zusammenzufassen, und ein neues Problem der Form (1) kann aufgesetzt werden.

Da das modifizierte ideale Anfangsgewicht F' nur aufgrund der beiden nichtlinearen Funktionen $x(t_{L+1}; x_L, p_L)$ und $h(x_L, z(t_L; x_L, p_L), u(t_L), p_L)$ nicht analytisch darstellbar ist, vereinfachen wir diese durch Linearisierung. Als Linearisierungspunkt wählen wir die beste verfügbare Schätzung $(x^*(t_L), p^*)$. Wir approximieren also

$$\begin{split} x(t_{L+1};x_L,\,p_L) &\approx \tilde{x} + X_x x_L + X_p p_L &\quad \text{mit} \\ X_x &:= \frac{dx(t_{L+1};\,x^*(t_L),\,p^*)}{d^*(t_L)} \,, \quad X_p := \frac{dx(t_{L+1};\,x^*(t_L),\,p^*)}{dp} \\ \text{und} \quad \tilde{x} &:= x^*(t_{L+1};\,x^*(t_L),\,p^*) - X_x x^*(t_L) - X_p p^* \end{split}$$

sowie analog

$$h(x_L, z(t_L; x_L, p_L), u(t_L), p_L) \approx \tilde{h} + H_x x_L + H_p p_L.$$

Durch diese Approximation wird das Problem (2) tatsächlich in ein analytisch lösbares lineares Ausgleichsproblem der Form

$$\min_{\substack{x_L,p_L\\ \\ \\ W}} \left\| \begin{aligned} P_L \begin{pmatrix} x_L - \bar{x}_L\\ p_L - \bar{p}_L \end{pmatrix} \\ V_L (\eta_L - \tilde{h} - H_x x_L - H_p p_L) \\ W \begin{pmatrix} x_{L+1} - \tilde{x} - X_x x_L - X_p p_L\\ p_{L+1} - p_L \end{pmatrix} \right\|_2^2$$

umgewandelt. Dieses quadratische Problem in den Variablen $(x_L, p_L | x_{L+1}, p_{L+1})$ kann durch eine QR Zerlegung der Matrix

$$\begin{pmatrix} P_L & & 0 \\ -(V_L H_x | V_L H_p) & 0 \\ -W \begin{pmatrix} X_x & X_p \\ 0 & \mathbb{I} \end{pmatrix} & W \end{pmatrix} = Q \begin{pmatrix} R_1 & R_{12} \\ 0 & R_2 \\ 0 & 0 \end{pmatrix}$$

in ein äquivalentes Problem

$$\min_{x_{L}, p_{L}} \left\| \begin{pmatrix} r_{1} \\ r_{2} \\ r_{3} \end{pmatrix} + \begin{pmatrix} R_{1} & R_{12} \\ 0 & R_{2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{L} \\ p_{L} \\ x_{L+1} \\ p_{L+1} \end{pmatrix} \right\|_{2}^{2}$$

transformiert werden. Dies hat die analytische Lösung

$$F''(x_{L+1}, p_{L+1}) = ||r_3||_2^2 + ||r_2 + R_2 {x_{L+1} \choose p_{L+1}}||_2^2$$

und für die Wahl des neuen Anfangsgewichtes ergibt sich

$$P_{L+1} := R_2, \quad \begin{pmatrix} \bar{x}_{L+1} \\ \bar{p}_{L+1} \end{pmatrix} := -R_2^{-1} r_2.$$

Bemerkung 1: Für alle Vektoren $v \in \mathbb{R}^{n_X+n_P}$ gilt $\|v\|_{P_{L+1}}^2 \leq \|v\|_W^2$ (dies folgt leicht aus der Identität $R_{12}^TR_{12} + R_2^TR_2 = W^TW$). Der a-priori Anfangswert \bar{x}_{L+1} und Parameterwert \bar{p}_{L+1} können also nicht stärker bestimmt sein, als es das zur Abschwächung von Vorinformation eingeführte Rauschen mit Kovarianz W^{-2} erlaubt.

Bemerkung 2: Die beschriebene Heuristik ist für lineare Systeme exakt äquivalent zum zeitdiskreten Kalman-Filter und erbt somit dessen theoretische und praktische Eigenschaften. Man beachte, dass der dadurch erhaltene MHE Schätzer jedoch selbst im Falle linearer Systeme und der Abwesenheit von Schranken verschieden vom Kalman-Filter ist, da wir auf dem Schätzhorizont kein Zustandsrauschen annehmen. In unserer Formulierung ist die Länge $T_{\rm E}$ des Schätzhorizontes somit ein wichtiger Designparameter.

Bemerkung 3: Für nichtlineare Systeme ist die Anfangsgewichtsheuristik ähnlich zum klassischen zeitdiskreten "Erweiterten Kalman-Filter" (EKF), aber nicht gleich. Sie unterscheidet sich vom EKF in der Wahl des Linearisierungspunktes, der sich beim EKF direkt aus dem alten Anfangsgewicht als (\bar{x}_L, \bar{p}_L) bestimmt, während bei unserer Heuristik der "geglättete Schätzwert" $(x^*(t_L), p^*)$ als Aufpunkt der Linearisierung verwendet wird. Die Linearisierung um einen geglättete Schätzwert wird auch in [38] behandelt.

Bemerkung 4: Für den Sonderfall M = 0, wenn also nur die aktuellste Messung im Schätzproblem berücksichtigt wird, lässt sich zeigen, dass der vorgeschlagene MHE-Algorithmus mit dem klassischen EKF identisch ist.

3 Das Mehrzielverfahren

Unser On-line-Ansatz zur numerischen Lösung des MHE Schätzproblems (1) beruht auf dem Mehrzielverfahren für die Parameterschätzung [8;9;32]. Die Grundidee ist, den Zeithorizont $[t_L, t_K]$ in Teilintervalle zu zerlegen, und auf jedem dieser so genannten Mehrzielintervalle unabhängig ein DAE Anfangswertproblem zu formulieren, das dann wiederum durch Stetigkeitsbedingungen an seine Nachbarintervalle gekoppelt ist. Das dadurch resultierende endlichdimensionale Problem wird dann mit einem verallgemeinerten Gauss-Newton-Verfahren (VGN) gelöst. Im Gegensatz zu einem konzeptionell einfacheren ähnlichen und wesentlich weiter verbreiteten Ansatz, dem einfachen Schießverfahren [25; 28; 45], ergeben sich durch Einführung der Teilintervalle günstigere lokale und globale Konvergenzeigenschaften. Dadurch wird eine Parameterschätzung sogar bei chaotischen dynamischen Systemen möglich [2; 24].

Der einfacheren Notation halber wählen wir in diesem Artikel die M Samplingintervalle $[t_k, t_{k+1}], k = L, \ldots, M-1$ als Mehrzielintervalle zur Unterteilung des Horizontes $[t_L, t_K]$. Wir bemerken jedoch, dass dies in keiner Weise zwingend ist, und insbesondere auch mehrere Messzeitpunkte auf einem Mehrzielintervall leicht behandelt werden können.

Als Anfangswerte für differentielle und algebraische Variablen x und z auf dem Intervall $[t_k, t_{k+1}]$ führen wir die Variablen s_k^x und s_k^z ein. Sodann betrachten wir das relaxierte Anfangswertproblem

$$\dot{x_k}(t) = f(x_k(t), z_k(t), u(t), p) \quad t \in [t_k, t_{k+1}]$$
 (3a)

 $0 = g(x_k(t), z_k(t), u(t), p)$

$$-\alpha_k(t)g\left(s_k^x, s_k^z, u(t_k), p\right) \tag{3b}$$

$$x_k(t_k) = s_k^x, \ z_k(t_k) = s_k^z.$$
 (3c)

Der Subtrahend in (3b) ist absichtlich eingeführt, um trotz möglicherweise algebraisch inkonsistenten Anfangswerten s_k^x , s_k^z einen DAE Löser konsistent starten zu können. Dies erlaubt es, teuere Konsistenziterationen zu Beginn der DAE Lösung zu vermeiden. Dabei ist der skalare Dämpfungsfaktor α_k so gewählt, dass $\alpha_k(t_k) = 1$, und dass α_k auf dem Intervall $[t_k, t_{k+1}]$ entweder konstant bleibt, oder, günstiger noch, exponentiell auf etwa ein Zehntel seines Wertes abfällt [27]. Wir bezeichnen die Lösung des Anfangswertproblems (3) mit $x_k(t; s_k^x, s_k^z, p)$ und $z_k(t; s_k^x, s_k^z, p)$ um die Abhängigkeit von den Anfangswerten und Parametern auszudrücken. Durch die Reformulierung des Mehrzielverfahrens ergibt sich ein dem ursprünglichen Problem (1) fast

äquivalentes Problem in endlich vielen Variablen:

$$\min_{\substack{s_L^x, \dots, s_K^x, \\ z_L^z, \dots, s_K^z, p}} \left\| s_L^x - \bar{x}_L \right\|_{P_L}^2 + \sum_{k=L}^K \| \eta_k - h \left(s_k^x, s_k^z, u(t_k), p \right) \|_{V_k}^2$$
(4a)

unter den Nebenbedingungen

$$s_{k+1}^{x} = x_k (t_{k+1}; s_k^{x}, s_k^{z}, p), \quad k = L, \dots, K-1,$$
 (4b)

$$0 = g(s_k^x, s_k^z, u(t_k), p), \qquad k = L, \dots, K,$$
 (4c)

$$x_{\min} \le s_k^x \le x_{\max}$$
, $k = L, \dots, K$, (4d)

$$p_{\min} \le p \le p_{\max}$$
 (4e)

Der einzige Unterschied im Vergleich zum ursprünglichen Problem (1) ist die Diskretisierung der Zustandsbeschränkungen (4d). Fassen wir alle Variablen dieses Problems der kleinsten Quadrate in dem Vektor

$$w = (s_L^x, s_L^z, s_{L+1}^x, s_{L+1}^z, \dots, s_K^x, s_K^z, p)$$

zusammen, und bedenken wir die Definition $||x||_V^2 = ||Vx||_2^2$, so kann das Problem (4) kompakt geschrieben werden als

$$\min_{w} \|F(w; D)\|_{2}^{2} \quad \text{s.t.} \quad G(w; D) = 0 \ , \ H(w; D) \ge 0 \ . \tag{5}$$

Das zusätzliche Argument *D* drückt dabei die Abhängigkeit von den On-line-Problemdaten

$$D = (\bar{x}_L, \bar{p}_L, P_L, \eta_L, V_L, \dots, \eta_K, V_K; u(t), t \in [t_L, t_K])$$

aus, die insbesondere für den folgenden Abschnitt 4 benötigt werden.

Das verallgemeinerte Gauss-Newton-Verfahren zur Lösung von (5) arbeitet iterativ und verwendet in jeder Iteration eine Linearisierung aller Problemfunktionen am aktuellen Iterationswert w_i . Mit Hilfe dieser Linearisierung wird ein beschränktes lineares Ausgleichsproblem formuliert und gelöst, um ein Inkrement Δw_i zu bestimmen. Sodann wird die nächste Iteration durch $w_{i+1} = w_i + \Delta w_i$ vollzogen. Das lineare Ausgleichsproblem ist gegeben durch

$$\min_{\Delta w} \left\| F(w_i; D) + \nabla_w F(w_i; D)^T \Delta w \right\|_2^2$$
 (6a)

s.t.
$$G(w_i; D) + \nabla_w G(w_i; D)^T \Delta w = 0$$
, (6b)

$$H(w_i; D) + \nabla_w H(w_i; D)^T \Delta w \ge 0.$$
 (6c)

Wir bemerken, dass dieses lineare Ausgleichsproblem dünnbesetzt und hochstrukturiert ist und deshalb aus Gründen der Effizienz und numerischen Stabilität mit strukturausnutzenden Methoden gelöst werden sollte [9; 14; 40]. Da die Berechnung der Ableitungsmatrix $\nabla_w G(w_i; D)^T$ eine Sensitivitätsberechung entlang aller Trajektorienstücke auf den Mehrzielintervallen verlangt, sollten die DAE Anfangswertprobleme (3) mit einem DAE Löser behandelt werden, der gleichzeitig auch Sensitivitäten berechnet. Strikt abzuraten ist von der Idee, die Ableitungsmatrizen durch mehrmaliges Aufrufen eines adaptiven DAE Lösers mit leicht gestörten Trajektorien und folgender Berechnung finiter Differenzen zu bestimmen. Dieses Vorgehen,

dass wir "Externe Numerische Differentiation (END)" nennen [9], versucht, die nichtdifferenzierbare Gittergenerierung des adaptiven DAE Lösers mit abzuleiten, und führt nur dann zu brauchbaren Ergebnissen, wenn die Genauigkeit des DAE Lösers deutlich höher gesetzt wird als dies für die DAE Simulation selbst nötig wäre. Dieses Vorgehen führt zu exzessiv hohen Rechenzeiten oder, schlimmer noch, zu gänzlich unbrauchbaren Ableitungswerten. Stattdessen sollten Variationsdifferentialgleichungen oder variierte Trajektorien gemeinsam mit der nominellen DAE Lösung auf einem gleichen Gitter berechnet werden. Dies führt direkt auf die Idee der "Internen Numerischen Differentiation (IND)" [9], die eng verwandt mit der Automatischen Differentiation [22] eines DAE Lösers ist. Die Gittergenerierung wird für die Ableitungsgenerierung eingefroren, die simultan mit der Vorwärtslösung erfolgt. Die Wahl eines geeigneten DAE Lösers mit Ableitungsgenerierung beeinflusst die Effizienz des Mehrzielverfahrens wesentlich. Für unseren Echtzeit-Algorithmus verwenden wir den BDF Code DAESOL[3;4]. Zur weiteren Effizienzerhöhung können die algebraischen Bedingungen (4c) genutzt werden, um viele Richtungsableitungen im Rahmen eines partiell reduzierten VGN Verfahrens einzusparen [27; 40].

Wir bemerken, dass die Konvergenz des verallgemeinerten Gauss-Newton-Verfahrens durch geeignete Liniensuche oder Trust-Region Strategien globalisiert werden kann, insbesondere durch Verwendung natürlicher Niveaufunktionen [9; 13]. Jedoch ist die lineare lokale Konvergenz des hier skizzierten ungedämpften Vollschritt-Verfahrens gegen im statistischen Sinne stabile lokale Minima immer gewährleistet [9].

4 MHE Echtzeititeration

Zur On-line-Lösung eines Zustands- und Parameterschätzproblems der Form (1) bzw. seiner Mehrziel-Reformulierung (5) mit sich von Problem zu Problem ändernden Daten *D* schlagen wir in diesem Artikel ein Echtzeititerationsverfahren vor, das sich an die Idee des Echtzeititerationsverfahrens für die optimale Steuerung [14; 15; 17] anlehnt.

Bezeichnen wir die Problemdaten des MHE Problems (1) auf dem Horizont $[t_L, t_K]$ mit D_K , und bezeichnen wir die optimale Lösung des entsprechend dem Mehrzielverfahren parametrisierten Problems (5) mit $w^*(D_K)$, so stellt sich die On-line-Optimierungsaufgabe wie folgt: berechne, sobald alle Problemdaten D_K bekannt sind, so schnell wie irgend möglich die Lösung $w^*(D_K)$. Bei diesem in vielen theoretischen Arbeiten [19;29;35] angenommenen Vorgehen wird jedoch die Tatsache ignoriert, dass die exakte Lösung $w^*(D_K)$ eines nichtlinearen Ausgleichsproblems niemals in endlicher Zeit berechnet werden kann, da das unterliegende Lösungsverfahren iterativ arbeiten muss und nur approximative Lösungen liefert. Zudem werden fast immer die für die numerische Lösung notwendigen Rechenzeiten vernachlässigt.

Wir schlagen deshalb in diesem Artikel einen Algorithmus vor, der zum ersten die Antwortzeiten zwischen Eintreffen einer neuen Messung und Ausgabe der neuen Zustandsund Parameterschätzung minimiert, und zum zweiten den Rechenaufwand pro Optimierungsproblem auf eine einzige verallgemeinerte Gauss-Newton-Iteration beschränkt. Dadurch können die durch die Rechenleistung nach unten beschränkten, möglichen Samplingzeiten deutlich reduziert werden.

Verschiedene Ansätze zur Echtzeitoptimierung von MHE Problemen, die eine gewisse Ähnlichkeit zu unserem Ein-Iterations-Ansatz haben, wurden bereits in der Literatur vorgeschlagen, [28; 30; 31], leiden jedoch an den schlechteren Kontraktionseigenschaften des bei allen verwendeten einfachen Schießverfahrens.

Das von uns vorgeschlagene Verfahren führt wie gesagt zu jedem Problem nur eine einzige VGN Iteration durch, und führt dann eine Initialisierung des folgenden Problems durch, die das Voranschreiten der Zeit berücksichtigt. Da sie gleichzeitig zur Zeit t_K auftreten, geben wir den VGN Iterationen w_K und den Problemdaten D_K den gleichen Index K. Wir betrachten im Folgenden den Übergang von einem Problem mit Daten D_K auf dem Horizont $[t_L, t_K]$ zum neuen Problem mit Daten D_{K+1} auf dem Horizont $[t_{L+1}, t_{K+1}]$.

Aus der vorhergehenden Iteration ist eine Approximation w_K der Lösung $w^*(D_K)$ des alten Problems bekannt. Neu zu den Daten hinzu kommt nun typischerweise nur die allerneueste Messung η_{K+1} und die Stellgrößen u(t) auf dem letzten Intervall $t \in [t_K, t_{K+1}]$, und durch das Voranschieben des Horizontes geht die erste Messung η_L verloren bzw. wird in dem neuen Anfangsgewicht zusammengefasst, wie in Abschnitt 2.1 beschrieben.

Jede MHE Echtzeititeration ist in eine lange Vorbereitungsphase und eine sehr kurze Schätzphase unterteilt. Die Vorbereitungsphase startet kurz nach der Zeit t_K , sagen wir zur Zeit $t_K + \delta$ und endet zur Zeit t_{K+1} . Die Schätzphase dauert typischerweise nur einen Bruchteil der Vorbereitungsphase (1%), und startet mit dem Eintreffen der neuen Messung η_{K+1} zur Zeit t_{K+1} , und endet zur Zeit $t_{K+1} + \delta$ mit der Ausgabe der neuen Iteration, w_{K+1} , einer Approximation der Lösung $w^*(D_{K+1})$ des neuen Problems auf dem Horizont $[t_{L+1}, t_{K+1}]$.

4.1 Vorbereitungsphase

Die Vorbereitungsphase dient dem Aufsetzen des neuen Anfangsgewichtes, der effizienten Initialisierung des neuen Problems aus der approximativen Lösung des alten Problems sowie der Vorbereitung eines linearen Ausgleichsproblems an der gegebenen Initialisierung.

Schritt 1: Nutze die alte Iteration w_K zur Berechnung des neuen Anfangsgewichtes nach der in Abschnitt 2.1 beschriebenen Methode. Als Linearisierungspunkt dieser dem Erweiterten Kalman-Filter ähnlichen Methode wird hierbei die aktuell beste Schätzung der Anfangswerte s_L^x und s_L^z

verwendet. Danach können die zeitlich ältesten Messwerte η_L gelöscht werden. Ebenso werden die ersten zwei Komponenten des Vektors

$$w_K = (s_L^x, s_L^z, s_{L+1}^x, s_{L+1}^z, \dots, s_K^x, s_K^z, p)$$

von nun an nicht mehr benötigt. Von den neuen Problemdaten

$$D_{K+1} = (\bar{x}_{L+1}, \bar{p}_{L+1}, P_{L+1}, \eta_{L+1}, V_{L+1}, \dots, \eta_K, V_K, \eta_{K+1}, V_{K+1}, \vdots u(t), t \in [t_{L+1}, t_{K+1}])$$

fehlt nun nur noch der letzte Messwert η_{K+1} , der erst zur Zeit t_{K+1} bekannt wird.

Schritt 2: Zur Initialisierung der nächsten VGN Iteration möchten wir aus w_K einen Vektor w_{K+1}^- berechnen, der zu dem um ein Intervall vorgeschobenen Horizont passt. Die Parameter p bleiben zu diesem Zweck auf ihren alten Werten, während die ersten M Komponenten einfach durch einen zeitlichen Shift initialisiert werden. Wir setzen also

$$w_{K+1}^- = (s_{L+1}^x, s_{L+1}^z, \dots, s_K^x, s_K^z, s_{K+1}^z, s_{K+1}^z, p)$$

wobei die Endwerte s_{K+1}^x , s_{K+1}^z noch unbekannt sind. Zu deren Ermittlung nutzen wir nun die in der NMPC Praxis bereits zur Zeit t_K vorliegende Information über die Steuerungen u(t) auf dem hinzukommenden Intervall $t \in [t_K, t_{K+1}]$. Dies erlaubt, die DAE startend mit der bisherigen Schätzung des aktuellen Zustands auf diesem Intervall zu lösen, und wir setzen dann einfach $s_{K+1}^x := x_K(t_{K+1}; s_K^x, s_K^z)$ und $s_{K+1}^z := z_K(t_{K+1}; s_K^x, s_K^z)$. Da die Ableitungen dieser DAE Lösung in der kommenden VGN Iteration noch gebraucht werden, berechnen wir diese gleich mit.

Schritt 3: Nachdem der Vektor w_{K+1}^- nun vorliegt, und von den neuen Problemdaten D_{K+1} nur noch eine einzige Messung η_{K+1} , fehlt, können wir nun fast alle Komponenten des neu zu lösenden linearen Ausgleichsproblems

$$\min_{\Delta w} \left\| F\left(w_{K+1}^{-}; D_{K+1}\right) + \nabla_{w} F\left(w_{K+1}^{-}; D_{K+1}\right)^{T} \Delta w \right\|_{2}^{2} \\
\text{s.t.}$$

$$G\left(w_{K+1}^{-}; D_{K+1}\right) + \nabla_{w} G\left(w_{K+1}^{-}; D_{K+1}\right)^{T} \Delta w = 0, \\
H\left(w_{K+1}^{-}; D_{K+1}\right) + \nabla_{w} H\left(w_{K+1}^{-}; D_{K+1}\right)^{T} \Delta w \geq 0,$$
(7)

berechnen. Die einzige fehlende Komponente ist die letzte Komponente der Funktion $F(w_{K+1}^-;\,D_{K+1}),$

$$V_{K+1} \left(\eta_{K+1} - h \left(s_{K+1}^x, s_{K+1}^z, u(t_{K+1}), p \right) \right).$$

Da aber der fehlende Wert η_{K+1} linear in diesen Term eingeht, können trotzdem bereits **alle** Ableitungsmatrizen des Problems vorberechnet werden. Zudem kann die Lösung des linearen Ausgleichsproblems weitestgehend vorbereitet werden, insbesondere durch Ausführen des Kondensieralgorithmus wie in [9] beschrieben. Bei Bedarf kann sogar Information über die Güte der zukünftigen Schätzung in Form einer Kovarianzmatrix berechnet werden, die glücklicherweise nicht vom konkreten Wert der fehlenden Messung η_{K+1} abhängt.

4.2 Schätzphase

Die Schätzphase beginnt mit dem Eintreffen der Messung η_{K+1} zum Zeitpunkt t_{K+1} . Nachdem in der Vorbereitungsphase bereits fast alles berechnet wurde, was zum Aufstellen und Lösen des linearen Ausgleichsproblems (7) nötig ist, kann dessen Lösung nun sehr rasch erfolgen. Das Ergebnis ist ein Korrekturschritt Δw_{K+1} in allen Variablen, der zu dem bisherigen Schätzwert w_{K+1}^- hinzuaddiert das Ergebnis w_{K+1} der (K+1)ten Zustandsschätzung liefert:

$$w_{K+1} = w_{K+1}^- + \Delta w_{K+1} .$$

Dieser Wert, bzw. seine Endwerte s_{K+1}^x , s_{K+1}^z und die Parameter p, wird nun als aktuelle Schätzung des Zustands zum Zeitpunkt t_{K+1} ausgegeben. Da die Berechnungen der Schätzphase eine kurze Zeit $\delta \ll t_{K+2} - t_{K+1}$ benötigen, liegt diese Schätzung erst zum Zeitpunkt $t_{K+1} + \delta$ vor. Nach Beendigung der Schätzphase beginnen wir sofort mit der Vorbereitungsphase für die nächste Iteration, die auf dem aktuellen Schätzwert w_{K+1} aufbaut.

4.3 Konvergenzeigenschaften

Eine mathematisch rigorose Konvergenzanalyse des beschriebenen MHE Echtzeititerationsverfahrens liegt noch nicht vor, aber könnte in ihren Grundzügen der Analyse der nominellen Stabilität der NMPC Echtzeititeration folgen [16]. Bei einer Konvergenzuntersuchung muss die nominelle Stabilitätstheorie des MHE Schätzverfahrens [37] (mit als exakt angenommener Lösung der MHE Probleme) verbunden werden mit der Konvergenztheorie verallgemeinerter Gauss-Newton-Verfahren [9]. Zwei einfache Aussagen lassen sich ohne weitere Analyse bereits jetzt angeben.

Eigenschaft 1: Der angegebene MHE Echtzeititerationsalgorithmus liefert im Falle linearer Systeme die exakten Lösungen der MHE Schätzprobleme, und erbt dementsprechend die theoretischen und praktischen Vorteile einer MHE Schätzformulierung, wie sie z.B. in [36] untersucht wird

Eigenschaft 2: Falls die Messdaten η_k durch das nominelle System erzeugt werden, also kein Model-Plant-Mismatch vorliegt, und falls der MHE Echtzeititerationsalgorithmus samt Anfangsgewicht in einer Iteration das Residuum $\|F(w_K, D_K)\|_2^2 = 0$ und die korrekte Zustandsschätzung gefunden hatte, bleibt der Algorithmus immer in der richtigen Lösung, und es gilt in jeder Iteration $\Delta w_K = 0$. Unter geeigneten Regularitätsannahmen an die Schätzprobleme sollte sich aus der Kontraktivität des VGN Verfahrens auch die Stabilität dieser Lösung bei kleinen Störungen leicht zeigen lassen.

In der Praxis zeigt der Algorithmus eine erstaunliche Robustheit und bei keinem unserer bisher durchgeführten Tests ist Divergenz des Schätzalgorithmus aufgetreten. Das gutartige Konvergenzverhalten ist dabei der Mehrzielformulierung und ihrer günstigen Behandlung der Nichtlinearitäten im Schätzproblem zu verdanken. Im folgenden Beispiel wollen wir illustrieren, wie der Algorithmus sich bei der Anwendung auf ein realitätsnahes Problem verhält.

5 Anwendung auf thermisch gekoppelte Destillationskolonnen

Zur Illustration wenden wir den beschriebenen MHE Algorithmus auf ein mittelgroßes DAE Modell zweier gekoppelter Destillationskolonnen an, mit 106 differentiellen und 371 algebraischen Variablen x, z sowie mit drei unbekannten und zu schätzenden Parametern p. Die Kolonnenkonfiguration ist in Bild 1 gezeigt, sie dient der Trennung eines Dreistoffgemisches aus Methanol, Ethanol und 1-Propanol. Methanol (A) und Ethanol (B) werden am Kopf jeweils der Haupt und der Nebenkolonne abgezogen, 1-Propanol (C) im Sumpf.

Das Energiesparpotenzial thermisch gekoppelter Destillationskolonnen macht diese sowohl für die verfahrenstechnische Industrie als auch für die Forschung interessant, siehe z.B. [21] oder [41] und die Zitate darin. Viele Studien konzentrieren sich auf den Entwurf und die Regelung [12]. Wenn gekoppelte Destillationskolonnen mit NMPC geregelt werden sollen, siehe z.B. [39], dann stellt sich sofort die Frage nach einer effizienten Zustands- und gleichzeitigen Parameterschätzung, die wir hier addressieren.

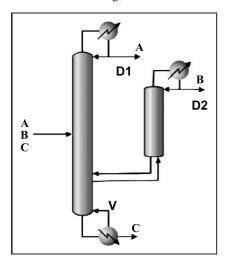


Bild 1: Gekoppelte Destillationskolonnen mit Seitenabzug für die Dreistofftrennung der Komponenten A, B, and C [41]. Stellgrößen sind der Verdampferfluss V und die Destillatflüsse D_1 and D_2 , die das System verlassen.

5.1 Das Kolonnenmodell

Das von uns verwendete Modell der gekoppelten Destillationskolonnen basiert auf der Arbeit von Lang [26]. Ein ähnliches Modell wurde jüngst auch zur Demonstration eines hierarchischen Reglers für Anfahrmanöver verwendet [23] sowie für eine NMPC Simulation ohne Schätzung [39].

Die Hauptkolonne besteht aus 42 Böden (einschließlich Sumpf und Kopf 1). Der Seitenabzug befindet sich auf Boden 11, der Zulauf auf Boden 21. Die Rektifizierkolonne besteht aus 11 Böden einschließlich Kopf 2. Das Modell beruht auf oft verwendeten Annahmen:

- chemisches und thermisches Gleichgewicht auf allen Böden,
- konstanter molarer Holdup auf allen Böden,

Tabelle 1: Liste einiger Prozessparameter.

| Parameter | Wert |
|---|--|
| Bodenholdup in Säule 1 Holdup Kopf 1 Holdup des Sumpf Bodenholdup in Säule 2 Holdup Kopf 2 Seitenabzugsrate Zulaufrate F Zulaufkomposition Methanol $x_{A,F}$ Zulaufkomposition Ethanol $x_{B,F}$ Nominalwert D_1 Nominalwert D_2 Nominalwert V | 1,0 mol 4,17 mol 114 mol 0,2 mol 3,08 mol 0,0265 mol/s 0,0555 mol/s 0,3 0,1 0,016840 mol/s 0,004732 mol/s 0,12700 mol/s |

- Vernachlässigung der Dampfmenge,
- perfekte Mischung, Ideale Gasphase,
- Phasengleichgewicht durch Antoine-Gleichung beschrieben,
- konstanter Druck in beiden Kolonnen,
- Totalkondensation in den Köpfen,
- thermodynamisch gesättigte Zu- und Rücklaufströme.

Das nichtlineare differential-algebraische Gleichungssystem (DAE) besteht aus den molaren Konzentrationen zweier der drei Komponenten auf den 42+11=53 Böden als den 106 differentiellen Variablen, während die dritte Konzentration, alle Temperaturen und die internen Stoffströme in Form von 371 algebraischen Variablen modelliert sind. Das gekoppelte Kolonnensystem wird in einer D-V-Konfiguration betrachtet, mit den Destillationsabzugsraten D_1 und D_2 in den Köpfen 1 und 2, und der Verdampfungsrate V im Sumpf als Stellgrößen, der Vektor u ist also gegeben durch $u = (D_1, D_2, V)$. Einige der (fixen) Modellparameter sind in Tabelle 1 aufgelistet.

5.2 Auslegung des Zustandsund Parameterschätzers

Für unsere numerischen Studien zur gleichzeitigen Zustands- und Parameterschätzung nehmen wir an, dass auf jedem der 53 Böden ein Temperatursensor vorhanden ist, und dass die Verläufe der Stellgrößen D_1 , D_2 und V bekannt sind. Als Samplingzeit nehmen wir 200 Sekunden an. Als Schätzhorizont betrachten wir $T_{\rm E}=1000$ Sekunden, die sich in M=5 Samplingintervalle aufteilen. Es liegen also im Schätzhorizont M+1=6 mal die 53 Temperaturmessungen vor. Diese Temperaturmessungen sind verrauscht mit einer Standardabweichung von 0,5 Kelvin.

Geschätzt werden, gleichzeitig mit den 106 differentiellen und 371 algebraischen Zuständen, auch drei dem Schätzer unbekannte und sich von Zeit zu Zeit ändernde Betriebsparameter: die Zuflussrate F sowie die Zuflusskonzentrationen $x_{A,F}$ und $x_{B,F}$.

Als Grenzen für alle differentiellen Zustände (relative molare Konzentrationen von A und B) geben wir im MHE-Schätzproblem $x_{\min} = 0$ und $x_{\max} = 1$ an. Für die drei Parameter $p = (F, x_{A,F}, x_{B,F})$ geben wir $p_{\min} = 0$

 $(4 \cdot 10^5 \, \mathrm{kmol/s}, 0, 0)$ und $p_{\mathrm{max}} = (6 \cdot 10^{-5} \, \mathrm{kmol/s}, 0, 5, 0, 3)$ an. Beim klassischen EKF entfällt die Möglichkeit der Angabe von oberen und unteren Schranken a-priori.

Die Gewichtungsmatrix $V \in \mathbb{R}^{53 \times 53}$ wird passend zur Standardabweichung des (unkorrelierten) Messrauschens von 0,5 Kelvin als Diagonalmatrix $V = \operatorname{diag}(0,5,\dots)^{-1}$ gewählt. Die Gewichtungsmatrix $W \in \mathbb{R}^{(106+3) \times (106+3)}$, die nur für das Update des Anfangsgewichtes für Zustände und Parameter verwendet wird, wählen wir entsprechend einer Störgrößenordnung von 0,1 für die Anfangswerte der 106 differentiellen Zustände $x(t_L)$, und von 0,001 für die drei Parameter $p = (F, x_{A,F}, x_{B,F})$ als

$$W = diag(0,1,\ldots,0,1,0,001,0,001,0,001)^{-1}$$

und setzen $P_0 = W$.

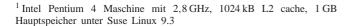
Das EKF ist gemäß [5] implementiert und verwendet die gleichen Kovarianz- und daher Gewichtsmatritzen, die auch dem MHE zu Grunde liegen. Auch die numerischen Methoden zur Berechnung der nötigen Jacobimatrix, zur Prädiktion der Zustände und der Berechnung konsistenter algebraischer Schätzwerte unterscheiden sich nicht von denen, die beim MHE verwendet wurden.

5.3 Numerische Ergebnisse

Das simulierte Schätz-Szenario wird durch die in den Bildern 2 und 3 gezeigten Anregungen erzeugt, wobei die (open-loop) Stellgrößen in Bild 2 auch dem Schätzer bekannt waren, während die in Bild 3 gezeigten und sich von Zeit zu Zeit ändernden Zulaufparameter von MHE und EKF gleichzeitig mit den Zuständen geschätzt werden mussten. Zu Beginn des Szenarios waren die Schätzer mit um 10% zu niedrigen Werten für die differentiellen und algebraischen Zustände sowie um 0,1% zu niedrigen Parametern initialisiert. Drei der geschätzen Konzentrationen sind in Bild 4 gezeigt und mit den wahren Werten verglichen. Die Grafik zeigt von oben nach unten die Konzentration der jeweils dominierenden Komponente in Kopf 1 und 2 sowie im Sumpf. Die entsprechenden geschätzten und gemessenen Temperaturen in den Köpfen 1 und 2 sowie dem Sumpf sind in Bild 5 gezeigt. Die Rechenzeiten pro MHE Optimierungsproblem lagen während des gesamten Szenarios bei rund 5 CPU-Sekunden auf einem Standard-PC1 (siehe Bild 6). Einzig im ersten Samplingintervall ist die Rechenzeit mit etwa 33 Sekunden deutlich höher, da die Zustände zum einen deutlich gestört waren und zudem inkonsistente algebraische Zustände vorlagen.

Bemerkung 1: Es ist deutlich zu sehen, dass der MHE weniger sensibel auf das Messrauschen reagiert und daher die besseren Schätzungen liefert. Dies liegt im Wesentlichen an der Nutzung mehrerer Messzeitpunkte für jede Schätzung und die daraus resultierende Glättung.

Bemerkung 2: In Bild 4 kommt es bei allen drei Konzentrationen zu physikalisch unzulässigen Schätzwerten durch



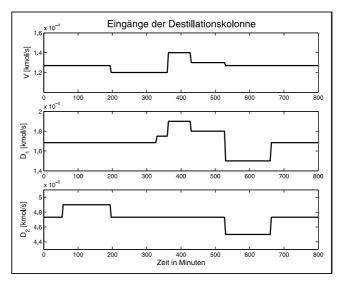


Bild 2: Zeitverläufe der drei Stellgrößen $u = (D_1, D_2, V)$, die der Anregung des Systems dienen.

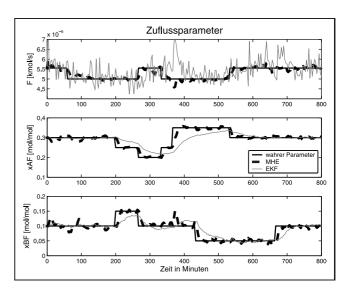


Bild 3: Wahre und geschätzte Werte für die drei den Schätzern nicht genau bekannten Parameter $p = (F, x_{A,F}, x_{B,F})$.

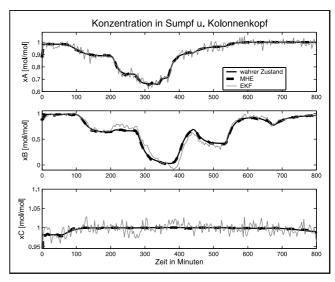


Bild 4: Wahre und geschätzte Werte für drei der Konzentrationen (in Kopf 1, Kopf 2 und Sumpf).

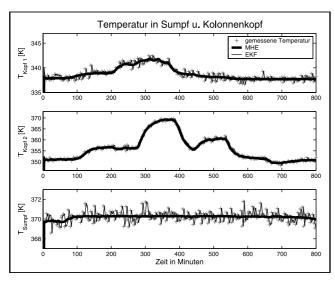


Bild 5: Gemessene und geschätzte Werte für drei der 53 Temperaturen (in Kopf 1, Kopf 2 und Sumpf).

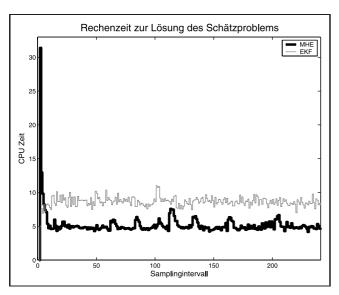


Bild 6: Gemessene Rechenzeiten für die Lösung des Schätzproblems zu einem Zeitpunkt von MHE und EKF.

das EKF, die entweder über 1 hinausgehen (x_A, x_C) oder negativ werden (x_B) . Dies lässt sich nur durch langwieriges Tunen und Testen verhindern, oder aber durch simples Abfragen und Wegschneiden. Dadurch verliert das EKF allerdings seine wahrscheinlichkeitstheoretische Bedeutung als minimalvarianter Schätzer. Diese Probleme tauchen beim MHE durch die andersartige Problemformulierung als beschränktes Optimierungsproblem gar nicht erst auf. Darin liegt wohl der größte Vorteil dieses Schätzverfahrens.

Bemerkung 3: Die Tatsache, dass die Rechenzeiten für das EKF recht deutlich über denen des MHE-Algorithmus liegen, mag zunächst überraschen. Die meiste Zeit benötigt der EKF-Algorithmus für die Integration der Systemgleichungen und die Ableitungsgenerierung, die ebenso beim MHE stattfindet. In der Wahl der Linearisierungspunkte unterscheiden sie sich jedoch, denn der MHE verwendet dafür bereits die "geglättete" Schätzung, während das EKF die Systemintegration jeweils bei den aus dem EKF Update

resultierenden Anfangswerten beginnt. Dadurch werden in jeder EKF Iteration erneut schnell abfallende Moden der Systemgleichungen angeregt, was zu mehr Integrationsschritten im adaptiven DAE Löser führt.

6 Zusammenfassung

In diesem Artikel haben wir die Frage nach effizienten Methoden für die gleichzeitige Zustands- und Parameterschätzung auf bewegtem Horizont (engl. Moving Horizon Estimation, MHE) untersucht. Wir haben eine Formulierung des MHE Schätzproblems für differential-algebraische Systemmodelle vorgeschlagen, die ein deterministisches Systemverhalten auf dem Schätzhorizont annimmt. Zur Hinzunahme geeigneter Vorinformation wird ein quadratisches Anfangsgewicht verwendet, das durch eine im Artikel vorgeschlagene Heuristik ähnlich dem erweiterten Kalman-Filter von einem Problem zum nächsten ermittelt werden kann. Zur numerischen Lösung des MHE Problems haben wir ein Echtzeititerationsschema vorgeschlagen, das auf dem Mehrzielverfahren für die Parameterschätzung in Verbindung mit dem verallgemeinerten Gauss-Newton-Verfahren (VGN) beruht. In diesem Algorithmus wird zu jedem MHE Optimierungsproblem nur eine einzige VGN Iteration durchgeführt, und der Übergang von einem Problem zum nächsten wird so gestaltet, dass die guten Kontraktivitätseigenschaften des Mehrzielverfahrens trotz der sich ständig ändernden Problemdaten erhalten bleiben. Wir haben die praktische Anwendbarkeit der Methode anhand eines Testproblems demonstriert, der gleichzeitigen Schätzung von Systemzustand und dreier Zuflussparameter für zwei thermisch gekoppelte Destillationskolonnen. Das verwendete DAE Modell weist 106 differentielle und 371 algebraische Zustände auf, und der Schätzhorizont umfasst 5 Mehrzielintervalle. Die Rechenzeiten pro Optimierungsproblem liegen bei etwa 5 CPU Sekunden. Ein Vergleich mit dem klassischen erweiterten Kalman-Filter (EKF) zeigt eine bessere Güte der Zustands- und Parameterschätzung. Die Verwendung des Echtzeititerationsschemas führt darüber hinaus im Vergleich mit dem EKF auch zu kürzeren Rechenzeiten.

Danksagung

Die vorgestellte Arbeit wurde im Rahmen des Projektes "Numerische Methoden für die optimierungsbasierte Regelung: Kopplung von On-line-Schätzung und robuster Prozessoptimierung" (BO-864/10) durch die *Deutsche Forschungsgemeinschaft (DFG)* finanziell unterstützt. Wir danken Leonard Wirsching für Hilfe bei der Implementierung des vorgestellten Zustandsschätzalgorithmus' sowie den anonymen Gutachtern für hilfreiche Kommentare.

Literatur

- [1] F. Allgöwer and A. Zheng. *Nonlinear Predictive Control*, volume 26 of *Progress in Systems Theory*. Birkhäuser, Basel Boston Berlin, 2000.
- [2] E. Baake, M. Baake, H.G. Bock, and K.M. Briggs. Fitting ordinary differential equations to chaotic data. *Phys. Rev. A*, 45:5524–5529, 1992.

- [3] I. Bauer. Numerische Verfahren zur Lösung von Anfangswertaufgaben und zur Generierung von ersten und zweiten Ableitungen mit Anwendungen bei Optimierungsaufgaben in Chemie und Verfahrenstechnik. PhD thesis, Universität Heidelberg, 1999.
- [4] I. Bauer, H.G. Bock, and J.P. Schlöder. DAESOL a BDF-code for the numerical solution of differential algebraic equations. Internal report, IWR, SFB 359, Universität Heidelberg, 1999.
- [5] V.M. Becerra, P.D. Roberts, and G.W. Griffiths. Applying the extended kalman filter to systems described by nonlinear di!erential-algebraic equations. *Control Engineering Prac*tice, 9:267–281, 2001.
- [6] D. Bestle and M. Zeitz. Canonical form observer design for nonlinear time-variable systems. *International Journal of Control*, 38:419–431, 1983.
- [7] T. Binder, L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder, and O.v. Stryk. Introduction to model based optimization of chemical processes on moving horizons. In M. Grötschel, S.O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems: State of the Art*, pages 295–340. Springer, 2001.
- [8] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Ebert, P. Deuflhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, volume 18 of *Springer Series in Chemical Physics*, pages 102–125. Springer, Heidelberg, 1981.
- [9] H.G. Bock. Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen, volume 183 of Bonner Mathematische Schriften. Universität Bonn, Bonn, 1987.
- [10] H.G. Bock and K.J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings* 9th IFAC World Congress Budapest, pages 243–247. Pergamon Press, 1984.
- [11] K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*. SIAM, Philadelphia, 1996. Classics in Applied Mathematics 14.
- [12] D. Demicoli and J. Stichlmair. Separation of ternary mixtures in a batch distillation column with side withdrawal. *Computers and Chemical Engineering*, 28(1):643–650, 2004.
- [13] P. Deuflhard. A Modified Newton Method for the Solution of Ill-conditioned Systems of Nonlinear Equations with Applications to Multiple Shooting. *Numerische Mathematik*, 22:289–311, 1974.
- [14] M. Diehl. Real-Time Optimization for Large Scale Nonlinear Processes, volume 920 of Fortschr.-Ber. VDI Reihe 8, Mess-, Steuerungs- und Regelungstechnik. VDI Verlag, Düsseldorf, 2002. Download also at: http://www.ub.uniheidelberg.de/archiv/1659/.
- [15] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. SIAM Journal on Control and Optimization, 43(5):1714–1736, 2005.
- [16] M. Diehl, R. Findeisen, F. Allgöwer, H.G. Bock, and J.P. Schlöder. Nominal stability of the real-time iteration scheme for nonlinear model predictive control. *IEE Proc.-Control Theory Appl.*, 152(3):296–308, 2005.
- [17] M. Diehl, R. Findeisen, S. Schwarzkopf, I. Uslu, F. Allgöwer, H.G. Bock, E.D. Gilles, and J.P. Schlöder. An efficient algorithm for nonlinear model predictive control of large-scale systems. Part I: Description of the method. *Automatisie*rungstechnik, 50(12):557–567, 2002.
- [18] H.J. Ferreau, G. Lorini, and M. Diehl. Fast nonlinear model predictive control of gasoline engines. In *CCA Conference Munich*, 2006. (accepted).

- [19] P.K. Findeisen. Moving horizon estimation of discrete time systems. Master's thesis, University of Wisconsin-Madison, 1997.
- [20] R. Findeisen, F. Allgöwer, and L. Biegler, editors. Assessment and Future Directions of Nonlinear Model Predictive Control. Lecture Notes in Control and Information Sciences. Springer, 2006.
- [21] A.J. Finn. Rapid assessment of thermally coupled side columns. *Gas. Sep. Purif.*, 10(3):169–175, 1996.
- [22] A. Griewank. Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation. Number 19 in Frontiers in Appl. Math. SIAM, Philadelphia, 2000.
- [23] A. Itigin, J. Raisch, T. Moor, and A. Kienle. A two-level hybrid control strategy for the start-up of a coupled distillation plant. In *European Control Conference*, Cambridge, UK, September 1–4 2003.
- [24] J. Kallrath, H.G. Bock, and J.P. Schlöder. Least squares parameter estimation in chaotic differential equations. Celestial Mechanics and Dynamical Astronomy, 56:353–371, 1993.
- [25] D. Kraft. On converting optimal control problems into nonlinear programming problems. In K. Schittkowski, editor, *Computational Mathematical Programming*, volume F15 of NATO ASI, pages 261–280. Springer, 1985.
- [26] L. Lang. Prozeβführung gekoppelter Mehrstoffkolonnen am Beispiel einer Destillationsanlage mit Seitenabzug. PhD thesis, Universität Stuttgart, 1991.
- [27] D.B. Leineweber. Efficient reduced SQP methods for the optimization of chemical processes described by large sparse DAE models, volume 613 of Fortschritt-Berichte VDI Reihe 3, Verfahrenstechnik. VDI Verlag, Düsseldorf, 1999.
- [28] A. M'hamdi, A. Helbig, O. Abel, and W. Marquardt. Newtontype receding horizon control and state estimation. In *Proc.* 13rd IFAC World Congress, pages 121–126, San Francisco, 1996.
- [29] K.R. Muske, J.B. Rawlings, and J.H. Lee. Receding horizon recursive state estimation. In *Proc. Amer. Contr. Conf.*, pages 900–904, San Francisco, 1993.
- [30] T. Ohtsuka. Nonlinear receding-horizon state estimation with unknown disturbances. *Trans. of the Society of Instrument* and Control Engineers, 35(10):1253–1260, 1999.
- [31] T. Ohtsuka and H.A. Fujii. Nonlinear receding-horizon state estimation by real-time optimization technique. *Journal of Guidance, Control, and Dynamics*, 19(4), 1996.
- [32] M.R. Osborne. On shooting methods for boundary value problems. *Journal of Mathematical Analysis and Applicati*ons, 27:417–433, 1969.
- [33] M. Perrier, S. Feyo de Azevedo, E.C. Ferreira, and D. Dochain. Tuning of observer-based estimators: theory and application to the on-line estimation of kinetic parameters. *Control Engineering Practice*, 8:377–388, 2000.
- [34] S.J. Qin and T.A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11:733–764, 2003.
- [35] C.V. Rao. Moving Horizon Estimation of Constrained and Nonlinear Systems. PhD thesis, University of Wisconsin– Madison, 1999.
- [36] C.V. Rao, J.B. Rawlings, and J.H. Lee. Constrained linear state estimation - a moving horizon approach. *Automatica*, 37(2):1619–1628, 2001.
- [37] C.V. Rao, J.B. Rawlings, and D.Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.
- [38] D.G. Robertson and J.H. Lee. A least squares formulation for state estimation. *Journal of Process Control*, 5(4):291– 299, 1995.

at 12/2006

- [39] A. Schäfer, P. Kühl, M. Diehl, J.P. Schlöder, and H.G. Bock. Fast reduced multiple shooting methods for nonlinear model predictive control. *Chemical Engineering and Processing*, 2006. (accepted).
- [40] J.P. Schlöder. Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung, volume 187 of Bonner Mathematische Schriften. Universität Bonn, Bonn, 1988.
- [41] J.G. Segovia-Hernandez, S. Hernandez, V. Rico-Ramirez, and A. Jimenez. A comparison of the feedback control behavior between thermally coupled and conventional distillation columns. *Computers and Chemical Engineering*, 28:811–819, 2004.
- [42] H. Seguchi and T. Ohtsuka. Nonlinear receding horizon control of an underactuated hovercraft. *International Journal of Robust and Nonlinear Control*, 13(3–4):381–398, 2003.
- [43] A. Toumi, M. Diehl, S. Engell, H.G. Bock, and J.P. Schlöder. Finite horizon optimizing control of advanced SMB chromatographic processes. In *IFAC World Congress*, Prague, 2005. (CD ROM).
- [44] A. Tournambe. High-gain obserbers for non-linear systems. *Int. Journal of Systems Science*, 23:1475–1489, 1992.
- [45] V.S. Vassiliadis, R.W.H. Sargent, and C.C. Pantelides. Solution of a class of multistage dynamic optimization problems.
 2. problems with path constraints. *Industrial and Engineering Chemistry Research*, 10(33):2122–2133, 1994.

Manuskripteingang: 3. März 2006.

Prof. Dr. Moritz Diehl ist Principal Investigator im neugegründeten "Optimization in Engineering Center" der K.U. Leuven. Bis Oktober

2006 war er C1 Assistent am Interdisziplinären Zentrum für wissenschaftliches Rechnen (IWR) der Universität Heidelberg, wo auch die vorgestellte Arbeit entstand. Sein Hauptarbeitsgebiet sind schnelle Algorithmen für dynamische Optimierungsprobleme in der nichtlinearen prädiktiven Regelung, mit Anwendungen u.a. in der Verfahrenstechnik, Automobiltechnik, Robotik und Energietechnik.

Adresse: Center of Excellence "Optimization in Engineering", Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, BELGIEN, E-Mail: moritz.diehl@esat.kuleuven.be

Dipl.-Ing. Peter Kühl ist Promotionsstipendiat der DFG am IWR der Universität Heidelberg. Seine Hauptarbeitsgebiete sind robuste nichtlineare prädiktive Regelung und optimierungsbasierte Zustandsschätzung komplexer technischer Prozesse.

Adresse: IWR, Universität Heidelberg, INF 368, 69120 Heidelberg, Tel.: +49-(0)6221-54-8830, E-Mail: peter.kuehl@iwr.uni-heidelberg.de

Prof. Dr. Dr. h.c. Hans Georg Bock ist Direktor am IWR der Universität Heidelberg und Inhaber des Lehrstuhls für Softwareinformatik. Hauptarbeitsgebiete: Algorithmen für hochdimensionale nichtlineare Optimierungsprobleme, Prozessoptimierung, Anwendungen u.a. in der Robotik, Fluiddynamik, Reaktionskinetik und Biotechnologie.

Adresse: IWR, Universität Heidelberg, Tel.: +49-(0)6221-54-8237, E-Mail: scicom@iwr.uni-heidelberg.de,

Homepage: http://www.iwr.uni-heidelberg.de/~agbock/

Dr. Johannes P. Schlöder ist Akademischer Oberrat am IWR der Universität Heidelberg. Hauptarbeitsgebiete: Numerische Methoden zur Parameterschätzung, optimalen Versuchsplanung und Optimierung nichtlinearer Prozesse, Anwendungen in Chemie, chemischer Verfahrenstechnik, Systembiologie und Maschinenbau.

Adresse: IWR, Universität Heidelberg, Tel.: +49-(0)6221-54-8239, E-Mail: johannes.schloeder@iwr.uni-heidelberg.de

