

Parallele versus klassisch-iterative Parameteridentifikation in dynamischen Systemen

Florian Judex, Markus Wawra*
Technische Universität Wien
efelo@osiris.tuwien.ac.at

Siegfried Wassertheurer†
ARCS Seibersdorf Research GmbH

12. August 2005

Zusammenfassung

Die Parameteridentifizierung in dynamischen Systemen wird beinahe ausschließlich mit Methoden der mathematischen Optimierung betrieben, die nicht darauf eingehen dass das System selbst nur näherungsweise gelöst wird. Die näherungsweise Lösung des Systems und die Anpassung der Parameter sind von einander getrennte Vorgänge, es erfolgt abwechselnd eine Lösung des Systems und eine Iteration des Optimierungsverfahrens. In dieser Arbeit wird ein Versuch eines neuen Ansatzes, der beide Probleme vereint lösen soll, vorgestellt. Leider führt das Verfahren - insbesondere bei für die Parameteridentifizierung besonders interessanten Fällen - nicht zum Ziel.

1 Motivation

Das Ziel von Modellbildung und Simulation ist, ein Modell eines dynamischen Systems zu erstellen - die Modellbildung - und dann aus Experimenten am Modell Rückschlüsse auf das ursprüngliche System ziehen zu können - dieser Teil wird als Simulation bezeichnet. Ein sehr wichtiger Abschnitt im Prozess der Modellbildung und Simulation ist die Identifikation von Parametern, die ein individuelles System in seiner Klasse von Modellen festlegen. Dafür werden Experimente, die auch am Modell möglich sind, am System durchgeführt. Die Reaktion des Systems wird gemessen, und die Parameter des Modells so lange variiert, bis die Reaktion des Modells den vom realen System erhaltenen Messwerten möglichst gleicht. Bei Strukturmodellen, die mit zumindest teilweise Wissen über das dynamische System erstellt werden, und deren Parameter zumindest ungefähre Gegenstücke im ursprünglichen System haben, ist oft die Identifizierung bestimmter Parameter das Ziel der Simulation.

*Institut für Analysis und Scientific Computing, Wiedner Hauptstraße 8-10, 1040 Wien, Austria

†Div. Biomedical Engineering, 2444 Seibersdorf, Austria

2 Das Verfahren

In dieser Arbeit wird zu diesem Zweck ein neuer Ansatz für die Identifikation bei Differentialgleichungsmodellen untersucht. Anstatt wie bisher die Optimierungsroutinen für die Identifikation (wie z.B. der Levenberg-Marquardt Algorithmus) und Numerischen Algorithmen zur Differentialgleichungslösung iterativ zu verwenden, d.h. je nach Algorithmus eine oder mehrere Lösungen des Differentialgleichungsmodells und der damit verbundenen Auswertung des Gütefunktionsals und danach eine Iteration des Optimierungsalgorithmus durchzuführen, sollen Identifikation und Modellberechnung parallel erfolgen.

Sei das Modell durch das n dimensionale Differentialgleichungssystem erster Ordnung

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{p}) \\ \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{pmatrix} &= \begin{pmatrix} f_1(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{p}) \\ f_2(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{p}) \\ \vdots \\ f_n(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{p}) \end{pmatrix} \end{aligned}$$

festgelegt, und p die Anzahl der zu identifizierenden Parameter. Ableiten aller Gleichungen nach den Parametern liefert die $n \cdot p$ Empfindlichkeitsgleichungen der Form

$$\dot{x}_k|_{p_i} = f_k|_{p_i}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{x}|_{p_i}, \dot{\mathbf{x}}|_{p_i}, t, \mathbf{p}).$$

Das System der ursprünglichen Differentialgleichungen und der Empfindlichkeitsgleichungen soll als

$$\dot{\mathbf{y}} = \tilde{\mathbf{f}}(\mathbf{y}, \dot{\mathbf{y}}, t, \mathbf{p})$$

zusammengefasst werden und hat die Dimension $n \cdot (p + 1)$. Weiters gebe es zu m Zeitpunkten t_1, \dots, t_m Messwerte M_{t_i} (o.B.d.A.) der ersten Koordinate. Zu minimieren ist dann das Gütefunktional

$$F(\mathbf{p}) = \sum_{i=0}^m (x_1(t_i, \mathbf{p}) - M_{t_i})^2$$

Bei der Erhebung der Messwerte soll das dynamische Verhalten des zu identifizierenden Systems so weit bekannt gewesen sein, dass mit einem geeigneten Integrationsalgorithmus für das Intervall $[t_i, t_{i+1}]$ eine einzige Iteration ausreicht d.h.

$$\mathbf{y}^{(i+1)} = \mathbf{g}(\mathbf{y}^{(i)}, \dot{\mathbf{y}}^{(i)}, \mathbf{p}, t, \tilde{\mathbf{f}})$$

für $y^{(i+1)} \approx y(t_{i+1})$. Diese Iteration läßt sich immer in Form eines Nullstellenproblems

$$\tilde{\mathbf{g}}(\mathbf{y}^{(i)}, \dot{\mathbf{y}}^{(i)}, \mathbf{p}, t, \tilde{\mathbf{f}}) = 0$$

darstellen.

Ein Extremum des Gütefunktional $F(\mathbf{p})$ zu finden ist ebenfalls ein Nullstellenproblem:

$$\sum_{i=0}^m (x_1(t_i, \mathbf{p}) - M_{t_i}) \cdot x_1|_{p_k}(t_i) = 0 \quad \text{für } k = 1..p$$

bzw.

$$\sum_{i=0}^m (x_1^{(i)} - M_{t_i}) \cdot x_1|_{p_k}^{(i)} = 0 \quad \text{für } k = 1..p$$

wenn die entsprechenden Komponenten der Näherungslösung eingesetzt werden. Diese beiden Nullstellenprobleme können nun zu einem einzigen zusammengefasst werden, das dann die Dimension

$$n \cdot (p + 1) \cdot m + p$$

besitzt. Eine Iteration des mehrdimensionalen Newton Verfahrens hat hier den Aufwand

$$(n \cdot (p + 1) \cdot m + p)^3,$$

also ca. $n^3 p^3 m^3$.

Stellt man diesem Verfahren die klassische Vorgangsweise gegenüber, bei der eine Iteration des Optimierungsverfahrens zumeist einen Aufwand von p^3 hat, da ein Gleichungssystem mit der Dimension des Parametersatzes zu lösen ist, ist der Aufwand des Verfahrens im wesentlichen durch die numerische Lösung des Systems bestimmt. Daher kann mit $O(p^3 n^3)$ pro Zeitschritt, also insgesamt $O(p^3 n^3 m)$, abgeschätzt werden. Damit stellt sich dieser neue, parallele Ansatz zumindest bei Anfangswertproblemen gewöhnlicher Differentialgleichungen wegen des vergleichsweise hohen Aufwands nicht als zielführend heraus.

3 Auftretende Probleme

Nebenbei treten bei Systemen, deren Lösungen abklingenden Exponentialfunktionen entsprechen, zusätzliche Probleme auf. Denn gerade bei Identifikationsproblemen hat man oft mit diesen Funktionen zu tun, da sich das System dann ab einem gewissen Zeitpunkt nicht mehr relevant verändert, was eine beinahe notwendige Bedingung für die Überprüfung der Ergebnisse darstellt.

Wenn man den oben beschriebenen Ansatz tatsächlich implementiert, sind die Ergebnisse deutlich anders als erwartet. Zwar wird fast immer ein Minimum erreicht - die gewählte Vorgangsweise garantiert nur ein Extremum, aber kein Maximum - aber trotz der Tatsache, dass die Ableitungen des Gütefunktional fast 0 ergeben, sind bei einer Überprüfung der Ergebnisse die berechneten Funktionswerte von den Messdaten signifikant verschieden, vorausgesetzt, man startet nicht gerade bei Parameterwerten die

den tatsächlichen Parametern so nahe sind, dass sich im Anwendungsfall die Frage nach einer Parameteridentifizierung erst gar nicht stellt.

Eine erneute Betrachtung der Ableitungen des Gütefunktional

$$\sum_{i=1}^m (M_{t_i} - x_1^{(i)}) x_1^{(i)} |_{p_k} \quad (1)$$

zeigt folgendes: da der Algorithmus keine exakte Nullstelle berechnet, sondern nur einen Wert erreicht, der nahe 0 ist, kommt es nun auch auf die Größenordnungen der beiden Faktoren in 1 an. Denn während die ersten Terme $(M_{t_i} - x_1^{(i)})$ immer nur in der Größenordnung der Messwerte schwanken, können die Normalgleichungen oft beliebig klein werden. An einem einfachen

3.1 Beispiel

wird das klarer: Sei die Lösungsfunktion $e^{-t \cdot a}$ und damit die Normalgleichung $-t \cdot e^{-t \cdot a}$ gegeben. Die Identifizierung soll das Intervall $t \in [0, 1]$ betreffen, mit Messwerten zu den Zeitpunkten $t_i = \frac{i}{10}, i = 1 \dots 10$. Der zu identifizierende Parameter a hat den Wert 10, und als weitere Vereinfachung soll die Annahme gelten, zu den Zeitpunkten t_i exakte Messwerte zu haben. Nun ein Vergleich der Ergebnisse, wenn die Identifizierung $a = 11$ bzw. $a = 1000$ ergibt, also einen relativen Fehler von 10^{-1} bzw. 10^2 :

$$\sum_{j=1}^{10} (e^{-\frac{j}{10} \cdot 10} - e^{-\frac{j}{10} \cdot a}) \cdot -\frac{j}{10} \cdot e^{-\frac{j}{10} \cdot a} = \begin{array}{ll} 0.188 \cdot 10^{-2} & \text{für } a = 11 \\ 0.136 \cdot 10^{-44} & \text{für } a = 1000 \end{array}$$

im Vergleich zu

$$\sum_{j=1}^{10} (e^{-\frac{j}{10} \cdot 10} - e^{-\frac{j}{10} \cdot a})^2 = \begin{array}{ll} 0.203 \cdot 10^{-2} & \text{für } a = 11 \\ 0.156 & \text{für } a = 1000 \end{array}$$

Eine Analyse der dazugehörigen Fehlerkurven und Lösungskurven zeigt, warum der Algorithmus versagt. Bild 1 zeigt die Funktionen, die für $a = 11$ (durchgezogene Linie) und $a = 1000$ (gestrichelte Linie) ausgewertet werden. Man sieht, dass die größten Unterschiede zwischen den beiden Funktionen schon vor dem ersten Messzeitpunkt $\frac{1}{10}$ auftreten. Während die Sensitivitätsgleichung für $a = 1000$ in Abbildung 2 schon bei $t = 10^{-3}$ ihr Maximum erreicht und dann gegen 0 strebt, ist das für $a=10$ erst bei $t = \frac{1}{10}$ der Fall. Das Produkt der beiden nimmt so die in Abbildung 3 gezeigten Werte an. e^{-at} ist die Lösung der Gleichung $\dot{x} = -ax$, die Änderung der Zustandsgröße ist proportional zu ihr. Wenn die Kurve schnell genug gegen die Null geht, ändert sich bei Parameteränderungen ab einem gewissen Zeitpunkt nicht mehr sehr viel, was nichts anderes bedeutet, als dass die Normalgleichung auch nur mehr sehr kleine Werte annimmt. Parameter und Eigenwert des Systems sind hier äquivalent. Bis auf die trivialsten Beispiele werden immer ein oder mehrere Eigenwerte direkt von den Parametern abhängen.

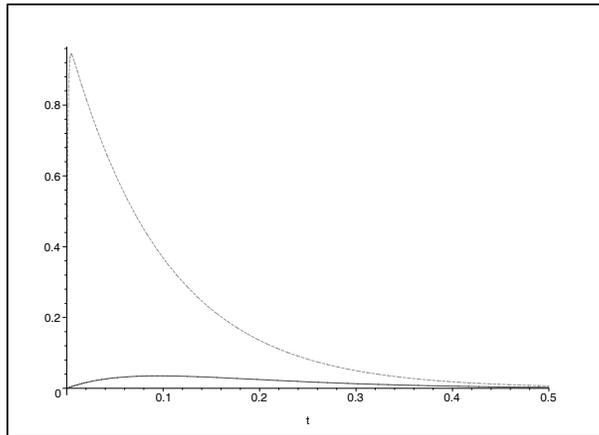


Abbildung 1: $e^{-\frac{t}{10} \cdot 10} - e^{-\frac{t}{10} \cdot 1000}$ und $e^{-\frac{t}{10} \cdot 10} - e^{-\frac{t}{10} \cdot 1000}$

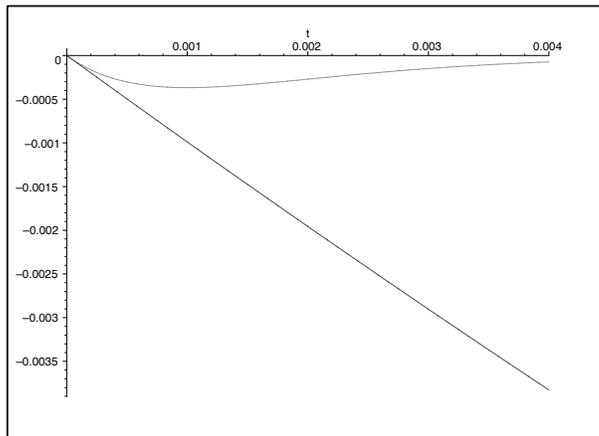


Abbildung 2: Parameterabelitung von e^{-10t} und e^{-1000t}

4 Konklusio

Durch eine passende - wenn auch falsche - Kombination können so die Eigenwerte vom Näherungsverfahren in beliebige Höhen gebracht werden. So spielt sich das ganze dynamische Verhalten der Lösung in einem sehr knappen Zeitbereich ganz am Anfang ab. Mit steigender Größe der Parameter ist es daher fast immer möglich, dass die gesamte Dynamik des Systems zwischen dem Anfangswert und dem ersten Messwert abläuft, und die Lösung sich dann in einen relativ stationären Zustand einschwingt, in dem die Normalgleichungen betragsmäßig fast 0 sind. Zur Verdeutlichung sind in Bild 4 die

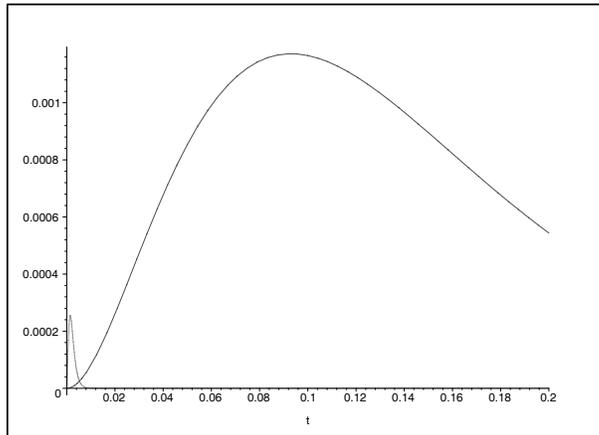


Abbildung 3: Fehler des Kompletten Terms

Fehlerterme für $a = 25$, $a = 100$ und $a = 1000$ graphisch dargestellt.

Im Bezug auf die ursprüngliche Minimierungsaufgabe stellt dieses Extremum ein Randextremum dar, das zumeist erreicht wird. Das Verfahren hat sich also insgesamt als nicht zielführend herausgestellt.

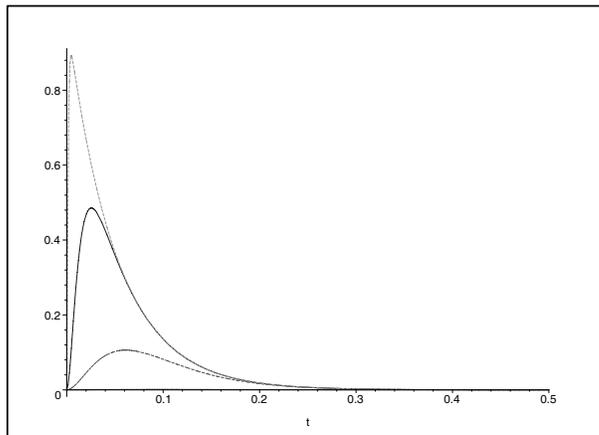


Abbildung 4: Fehlerquadrate für $a=1000, a=100, a=25$