

# Algebraische Modellierung mit Methoden der approximativen Computeralgebra und Anwendungen in der Ölindustrie

Daniel Heldt

Fachbereich Mathematik  
[daniel.heldt@mathematik.uni-dortmund.de](mailto:daniel.heldt@mathematik.uni-dortmund.de)  
Universität Dortmund

Martin Kreuzer

Fachbereich Mathematik  
[martin.kreuzer@uni-dortmund.de](mailto:martin.kreuzer@uni-dortmund.de)  
Universität Dortmund

Sebastian Pokutta

Operations Research Center  
[pokutta@mit.edu](mailto:pokutta@mit.edu)  
Massachusetts Institute of  
Technology

Hennie Poulisse

Exploratory Research  
[hennie.poulisse@shell.com](mailto:hennie.poulisse@shell.com)  
Shell International  
Exploration & Production

**Zusammenfassung – Möchte man auf der Basis von Messdaten Vorhersagen über das weitere Verhalten von komplexen Systemen machen, so benötigt man üblicherweise ein mathematisches Modell, das die Daten annäherungsweise interpoliert. Oftmals sind geeignete Modelle nicht bekannt oder wegen ihrer hohen Komplexität unzugänglich, so dass man eine auf statistische Beschreibung zurückgreifen muss. Der hier kurz skizzierte Ansatz erlaubt es, einfache, approximativ geltende Relationen und somit Modelle direkt aus den Daten abzuleiten. Er basiert auf einem Algorithmus aus der approximativen Computeralgebra. Zwei mögliche Anwendungen des Verfahrens werden anhand aktueller Probleme in der Ölindustrie beschrieben. Zum einen handelt es sich dabei um ein Problem aus der Ölförderung, zum anderen geht es um die Kompensation von Messfehlern defekter Messinstrumente.**

**Stichworte – approximate computeralgebra, data mining, approximation theory, modeling**

## 1 Einführung

Verwendet man Methoden des Operations Research, um Prozesse oder Planungsabläufe zu optimieren, so basiert diese Optimierung üblicherweise auf einem geeigneten mathematischen Modell. Wie geht man aber vor, wenn ein solches Modell nicht vorhanden ist, sondern nur approximative Daten, z.B. Messdaten oder Schätzungen, zur Verfügung stehen? Im Folgenden stellen wir eine Methode aus der approximativen Computeralgebra vor, mit der man in gegebenen Daten bei vorgegebener Toleranz nach polynomialen Relationen suchen kann, von denen dann Modelle abgeleitet werden können. Diese Methode unterscheidet sich von den klassischen Verfahren des Data Minings, da sie beispielsweise eine vollständige Beschreibung der approximativ geltenden polynomialen Relationen erlaubt. Insbesondere ermöglicht unsere Methode, zwei Modelle auf Äquivalenz bezüglich der Eingangsdaten zu prüfen. Die gewonnenen Modelle

bilden dann eine Basis für weitere Untersuchungen. Ebenso können sie als Ansatzpunkt für Optimierungsverfahren dienen.

In Abschnitt 2 geben wir eine kurze Einführung in das übliche Verfahren zur Bestimmung von Verschwindungsidealen. Danach fahren wir mit der Darstellung der approximativen Methoden fort. Insbesondere erklären wir den Zusammenhang zwischen der Berechnung eines approximativen Verschwindungsideals und dem ursprünglichen Problem der Bestimmung eines Modells. In Abschnitt 3 behandeln wir Anwendungen dieser Methoden in der Ölförderung bzw. dem *Reservoir Engineering*. In Abschnitt 4 stellen wir einige Ergebnisse unserer Berechnungen vor und geben einen Ausblick auf mögliche Anwendungen in anderen Bereichen.

## 2 Approximative Verschwindungsideale

Zunächst beginnen wir mit der klassischen Situation bei approximativen Interpolationsaufgaben. Gegeben sei eine Menge von (exakten) Punkten  $X = \{m_i : 1 \leq i \leq s\} \subseteq \mathbb{R}^n$ . Wir suchen alle Polynome  $p \in \mathbb{R}[x_1, \dots, x_n]$  mit der Eigenschaft  $p(m) = 0$  für alle  $m \in X$ . Jedes dieser Polynome  $p$  spiegelt eine polynomiale Relation wider, die auf der Punktmenge gültig ist. Die Menge der Polynome mit dieser Eigenschaft bildet ein Ideal  $I(X) \subseteq \mathbb{R}[x_1, \dots, x_n]$ , welches als *Verschwindungsideal von X* bezeichnet wird. Die Elemente des Verschwindungsideals von  $X$  können als polynomiale Modelle aufgefasst werden, d.h. als Gleichungen, die polynomiale Relationen unter den Input-Daten darstellen. Modelliert man eine Output-Datenreihe ebenfalls durch ein polynomiales Modell, so kann man die Äquivalenz zweier solcher Modelle bzgl. der Input-Daten überprüfen. Man braucht dazu nur zu testen, ob das Differenzpolynom der beiden Modelle im Verschwindungsideal liegt, was z.B. mit

Hilfe einer Gröbner-Basis von  $I(X)$  leicht durchzuführen ist.

Liegt die Punktmenge  $X$  exakt vor, können diese Verschwindungsideale mit Hilfe des Buchberger-Möller Algorithmus [BM82] leicht berechnet werden. Für eine Berechnung bei gestörten Daten ist dieser Algorithmus aber nicht geeignet. Der in [HKPP06] vorgestellte approximative Buchberger-Möller Algorithmus (ABM) erlaubt hingegen eine

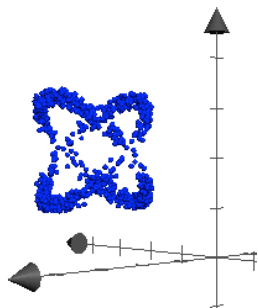


Abb. 1: Punktwolke im  $\mathbb{R}^3$

Berechnung von approximativen Verschwindungsidealen (appVI) für gestörte Daten. Hierbei wird die eigentliche Berechnung der Relationen im klassischen Buchberger-Möller Algorithmus modifiziert. In der ursprünglichen Form wird jeder Punkt als exakt gegeben angenommen und als Nullstelle der polynomialen Relationen realisiert. Mit Hilfe geeigneter Singulärwertberechnungen und -abschneidungen kann man eine Version dieses Algorithmus konstruieren, bei der die berechneten Funktionen auf der berechneten Punktmenge nahezu verschwinden. Unter anderem werden dadurch eng benachbarte Punkte durch eine eigene gemeinsame approximative Nullstelle repräsentiert. Die Details dieses neuen Algorithmus werden in [HKPP06] präsentiert.

Zur Illustration betrachten wir das folgende Beispiel. Es wurden 1000 Punkte erzeugt, die wie zwei Gürtel stark gestört (bis zu 30%) in der Nähe eines Ellipsoids liegen. Die resultierende Punktmenge  $X \subseteq \mathbb{R}^3$  ist in Abbildung 1 dargestellt. Wir suchen nun die polynomialen Relationen, die auf diesen Punkten approximativ verschwinden. Insbesondere sollte die definierende Gleichung des Ellipsoids als Relation auftreten, da die Daten so erzeugt wurden und somit diese Relation approximativ gültig ist. An dieser Stelle erkennt man bereits, dass über die Gestalt der berechneten Modelle nur geringe Annahmen gemacht wurden: die in den gesuchten Relationen vorkommenden Variablen seien bekannt, und die zu berechnenden Relationen seien polyno-

mialer Natur. Ein Element der Gröbner-Basis des approximativen Verschwindungsideals von  $X$  ist z.B. das Polynom

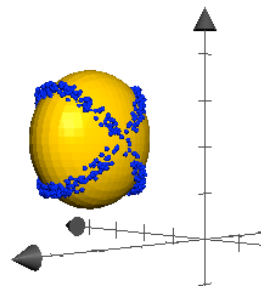


Abb. 2: Punktwolke im  $\mathbb{R}^3$  und gefundene Relation

$$p(x, y, z) = -0.071009z^2 - 0.002006yz - 0.073202y^2 - 0.0046655xz - 0.0056742xy - 0.1511x^2 - 0.78093 + 0.31644z + 0.32854y + 0.38572x,$$

welches das Ellipsoid beschreibt, das in Abbildung 2 zu sehen ist. Der klassische Buchberger-Möller Algorithmus würde hier nicht die gewünschten Ergebnisse liefern, da er jeden Punkt als Nullstelle der Polynome realisieren würde. Die approximative Variante hingegen greift viel mehr die eigentliche Information und weniger die vorgegebenen Punkte auf und liefert so auch bei gestörten Daten verwertbare Ergebnisse.

An dieser Stelle können wir bereits zwei weitere Vorteile der neuen Methode vermerken. Zum einen ist es möglich, auch nicht-polynomiale Modelle zu finden, zum Beispiel Wurzeln oder rationale Funktionen der Variablen. Dazu genügt es, diese als zusätzliche Unbestimmte der Berechnung hinzuzufügen. Bei der weiter unten vorgestellten Anwendung in der Ölindustrie wurden etwa Differenzen von Druckmessungen verwendet, da diese typische Bestandteil von physikalischen Gesetzen der Flüssigkeitsdynamik sind. Ein solcher Ansatz ist nicht gleichbedeutend mit der Vorgabe einer Gestalt des gesuchten Modells, da unser Verfahren noch einen weiteren großen Vorteil hat: es ist stabil gegenüber Erweiterungen des Modellhorizonts. Fügt man also unnötige Variablen der Berechnung hinzu, so werden diese automatisch ignoriert und die ursprüngliche Lösung wird gefunden. Diese Stabilität gegenüber Modellerweiterungen erlaubt es uns somit, dem Input neue Datenreihen hinzuzufügen, die aus den vorhandenen mittels fester Funktionen berechnet wurde, sofern diese Funktionen genügend unabhängig voneinander sind.

Der entscheidende Vorteil dieser Art der automatischen Modellierung ist der, dass nicht nur eine Approximation oder ein „best-fit“ zu den Daten bestimmt wird, sondern polynomiale Gleichungen in möglichst einfacher Form, die die approximativ geltenden Relationen erzeugen. Diese Behauptung wird durch die erfolgreiche Anwendung in der Ölförderung (vgl. Abschnitt 3) gestützt, in der tatsächlich sowohl bereits bekannte als auch vollständig neue, durch weitere Messungen verifizierte physikalische Gesetzmäßigkeiten abgeleitet wurden. Die so gewonnenen Ergebnisse erlauben Vorhersagen

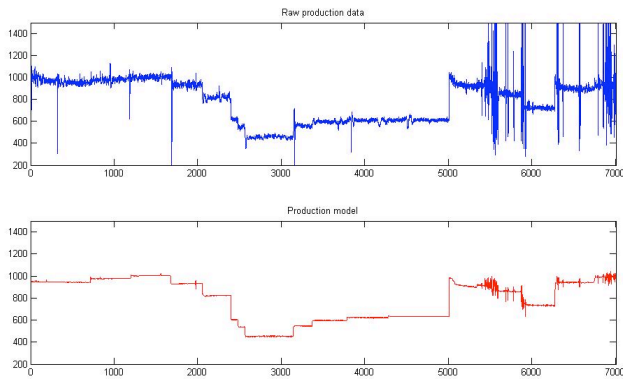


Abb. 3: Vorhersage der Gesamtproduktion. Die untere Linie beschreibt die Vorhersage, bei der oberen Linie handelt es sich um die tatsächliche Produktion

mit einer hohen Güte (vgl. Abbildung 3). Darüber hinaus können hierbei auch Relationen nachgewiesen werden, die zwar von gesuchten Funktionen erfüllt werden, aber die nicht explizit nach jenen Funktionen aufgelöst werden können.

### 3 Anwendungen in der Ölindustrie

Im Folgenden skizzieren wir kurz zwei Anwendungen in der Ölindustrie. Die Tatsache, dass es bis heute keine verlässlichen physikalischen Modelle für die Förderung von Öl oder Gas aus Reservoiren gibt, liefert einen starken Anreiz für die Entwicklung neuer Methoden zur Herleitung solcher Modelle. Nahezu alle bekannten Modelle der „Flüssigkeitsdynamik in porösen Medien“ sind für Vorhersagen nicht geeignet, so dass bisher auf Simulationen zurückgegriffen werden musste (siehe Abbildung 4). Dabei ist der Vorhersagefehler teilweise so groß, dass die berechneten Ergebnisse praktisch wertlos sind. In fast allen Fällen (bis auf wenige Ausnahmen) werden **lediglich 30% oder weniger** des vorhandenen Öls aus einem Reservoir gefördert, bevor eine weitere Förderung unmöglich wird.

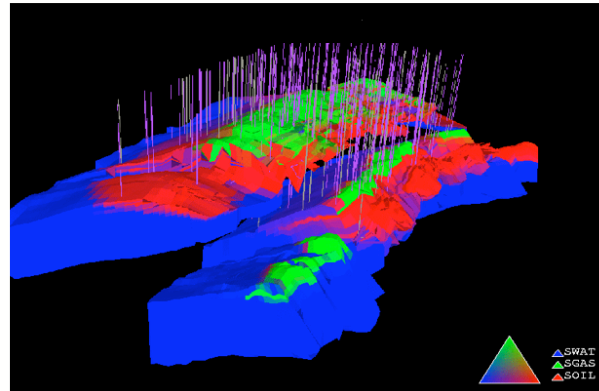


Abb. 4: Simulation zur Bestimmung von Öl-, Gas- und Wasservorkommen: Das Bild zeigt die Simulation eines Ölfeld. Bei den Linien handelt es sich um die Bohrungen.

Eine typische Ursache dafür ist die Bildung von sogenannten „dry spots“, d.h. das Reservoir zerfällt in viele kleinere Reservoire, aus denen nicht mehr wirtschaftlich gefördert werden kann. Ferner kann es zum sogenannten „water breakthrough“ kommen, d.h. ein falsches Reservoir-Management führt zu einem frühzeitigen Wassereintritt, und ab diesem Zeitpunkt wird nur noch Wasser gefördert. Oder aber zu Beginn wird zuviel Gas gefördert, so dass der Reservoir-Druck zu schnell abfällt und eine weitere Förderung wegen des Druckverlusts unmöglich wird. All diese Prozesse sind nach heutigem Kenntnisstand irreversibel, so dass in diesen Fällen eine weitere Bewirtung des Feldes ausgeschlossen ist und große Mengen Öl (über 70%) im Boden verbleiben müssen.

Um das Eintreten dieser unerwünschten Effekte zu

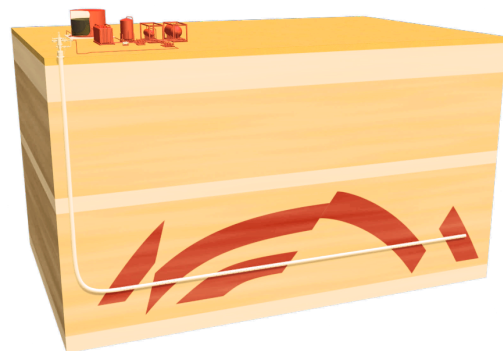


Abb. 5: Multi-Zone Well. Förderung aus mehreren, teilweise verbundenen Taschen

verhindern, ist es wichtig, die (physikalischen) Interaktionen innerhalb des Reservoirs zu verstehen (siehe Abbildung 5). Es ist durchaus denkbar, dass aus einer feldspezifischen Modellierung des Reservoirs frühzeitige Indikatoren für die oben beschriebenen Phänomene gewonnen werden können. Dann könnte man frühzeitig adäquate Gegenmaßnahmen

einleiten, z.B. eine Variation der Förderanteile. Entsprechende Untersuchungen dazu finden derzeit im Rahmen des *Algebraic Oil* Projektes [AO06] statt. Es ist sogar vorstellbar, dass ein umfassendes Verständnis der Interaktionen innerhalb eines spezifischen Reservoirs dazu führt, dass man durch geschicktes Ausnutzen derselben ein beschädigtes Feld wiederherstellen kann, was die Gesamtausbeute (*ultimate recovery*) erhöhen würde. Die Berechnung von Interaktionen zwischen verschiedenen „Taschen“ (Bereiche, die Öl beinhalten), kann auch dazu genutzt werden, Rückschlüsse über die tatsächliche Struktur des Feldes zu ziehen. Diese Information wird mit herkömmlichen seismischen oder geophysischen Methoden nur äußerst ungenügend bereitgestellt, ist aber wichtig für eine optimale Bewirtung eines Feldes; insbesondere die Kenntnis über die Verbundenheit bzw. Unabhängigkeit von Taschen spielt hier eine entscheidende Rolle. Ebenso tragen die Interaktionen zwischen verschiedenen Bohrlöchern zum Verständnis des Produktionssystems bei. Somit ist eine Modellierung der Interaktionen sehr wertvoll, um eine Produktionsoptimierung zu gestalten. Mit Hilfe einer expliziten Darstellung der Relationen kann man das Verhalten des Feldes vorhersagen (vgl. Abbildung 3) und somit das Produktionssystem wesentlich gezielter zu steuern. Zum Beispiel kann eine hohe Gasproduktion einer Bohrung genutzt werden, um das Öl einer anderen Bohrung leichter zu machen und es somit effizienter fördern zu können, allerdings nur dann, wenn das Mischverhältnis korrekt ist (vgl. Abbildung 6). Auch bauen Teile des Reservoirs den eigenen Druck wieder auf, wenn sie für eine Weile vom Produktionssystem abgetrennt werden. Ferner gibt es Felder, bei denen das erneute Starten der Produktion nach einem *Shutdown* einer komplexen Abfolge von Operationen bedarf und teilweise Tage dauert. Beispielsweise kann die Reihenfolge, in der die Bohrlöcher geöffnet werden, sehr wichtig sein.

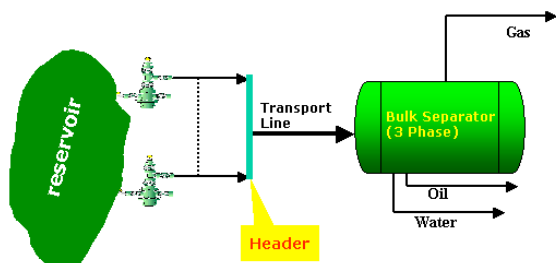


Abb. 6: Schematischer Aufbau

Das Verständnis der Interaktionen zwischen den Bohrlöchern kann man nun dazu nutzen, diese Operationen zu optimieren, so dass das System nach einem Shutdown schneller (und unter Umständen besser) neu gestartet werden kann.

Eine weitere, direkt abgeleitete Anwendung ist die Möglichkeit, fehlerhaftes Equipment zu kompensieren. Dies bezieht sich insbesondere auf Messgeräte im *sub-surface* Bereich. Hier können mit Hilfe des berechneten Reservoir-Modells fehlende Messwerte abgeleitet werden. In Bohrlöchern existieren Druck- und Temperatur-Sensoren, die benutzt werden um Einblick in den inneren Zustand des Produktionssystems zu erhalten und zum Beispiel Ströme von Gas, Öl und Wasser innerhalb der Leitungen zu ermitteln. Aufgrund der rauen Umgebung neigen diese Sensoren allerdings dazu auszufallen. Gleichzeitig ist das Ersetzen oder Reparieren der Sensoren sehr kostspielig oder gar unmöglich. Jedoch können die Relationen, die zuvor aus den vor dem Ausfall gemessenen Daten bestimmt wurden, benutzt werden um den aktuellen Wert der defekten Sensoren mit Hilfe von nicht defekten Sensoren zu bestimmen. Im Normalfall ist dies durch die erhöhte Redundanz, insbesondere durch die Korrelation von Temperatur und Druck möglich. Zusätzlich kann überprüft werden, inwieweit ein Erzeugendensystem des Verschwindungsidents weiterhin auf den aktuellen Messdaten ungefähr verschwindet. So zeigt sich, ob das Modell noch gültig ist bzw. ob die dem Modell zu Grunde liegenden Annahmen noch zutreffen, oder aber ob sich das Reservoir verändert

	$\epsilon$	#GB	Time
1	1	15	0.77 s
2	0.5	19	0.84 s
3	0.1	24	0.82 s
4	0.01	39	1.43 s
5	0.001	59	2.34 s
6	0.0001	82	4.04 s
7	0.00001	125	7.31 s

Tabelle 1: Timings zur Berechnung einer DegLex Gröbner Basis für 2445 Punkte in  $\mathbb{R}^9$

hat. Auf diese Weise erhält man zum einen ein Maß für die Güte der vorhergesagten Werte, zum anderen lassen sich auch Veränderungen im Reservoir diagnostizieren (siehe Abbildung 7).

## 4 Ergebnisse und Ausblick

Mit der skizzierten Berechnung des approximativen Verschwindungsideals konnten feldspezifische Gesetzmäßigkeiten nachgewiesen werden, die durch physikalische und statistische Untersuchungen bestätigt wurden, aber durch ihre unerwartete Form höchstwahrscheinlich unentdeckt geblieben wären. Dabei ist zu bedenken, dass insbesondere die hohe Komplexität der Modelle und die vielen Einfluss-

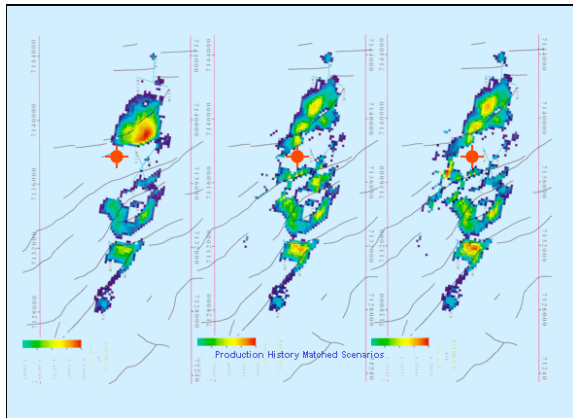


Abb. 7: Veränderung des Ölfeldes über die Zeit

größen eine klassische Deduktion der geltenden Gesetzmäßigkeiten unmöglich machen. Die Geschwindigkeit der eigentlichen Berechnung ist hoch genug um auch komplexe Datensätze bearbeiten zu können. Typische Datensatzgrößen des Algebraic Oil Projektes liegen hier zwischen 2000 und 10000 Punkten mit einem zu erwartenden Fehler bis 30.0% in bis zu 15 Unbestimmten. Die vorliegende (prototypische) Implementierung in CoCoA [CoCoA] berechnet wahlweise eine Gröbner-, Rand- oder Macaulay-Basis des approximativen Verschwindungsideals. Hierbei ist die Geschwindigkeit beider Implementierungen nahezu identisch. Einige Timings sind in Tabelle 1 aufgeführt.

Darüber hinaus erschließen sich auch in anderen Industriebereichen völlig neue Anwendungsfelder. Der entscheidende Vorteil zum klassischen Vorgehen ist der, dass kein Modell vorgegeben werden muss. Der übliche Ansatz, im Vorfeld ein Modell herzuleiten und durch Regression anzupassen, entfällt somit. Vielmehr erlaubt die Verwendung des approximativen Verschwindungsideals, bei keinerlei Kenntnis der zugrundeliegenden Relationen und Gesetzmäßigkeiten direkt die gesuchten Modellgleichungen aus den Daten zu gewinnen. Die vielversprechenden Ergebnisse der Anwendung in der

Ölindustrie machen Hoffnung auf ähnliche Ergebnisse in anderen Bereichen.

## 5 Anmerkungen

Die hier vorgestellten Methoden sind Gegenstand aktueller Forschung und wurden im Rahmen des von *Shell International Exploration & Production* finanzierten *Algebraic Oil* Projektes [AO06] entwickelt und erforscht [HKPP06]. Der dritte Autor wurde durch ein Stipendium im Rahmen der Postdoc Programme des Deutschen Akademischen Auslandsdienst (DAAD) unterstützt.

## Literaturverzeichnis

[AO06] Algebraic Oil Project, <http://www.mathematik.uni-dortmund.de/CoCoA/>

[BM82] B. Buchberger and H.M. Möller. The construction of multivariate polynomials with preassigned zeros, in: J. Calmet (ed.), Proceedings of EUROCAM'82, Lect. Notes in Comp. Sci. **144**, Springer, Heidelberg 1982, 24--31.

[CoCoA] CoCoA – a system for doing Computations in Commutative Algebra, available at <http://cocoa.dima.unige.it>

[HKPP06] D. Heldt, M. Kreuzer, S. Pokutta, H. Poulisse. *Approximate computation of zero-dimensional polynomial ideals*, submitted (2006), Preprint available upon request.