Problem Numerik: Rundungs- und Verfahrensfehler

Ein Erfahrungsbericht aus dem Praktikum zur Lehrveranstaltung "Simulation"

Problem

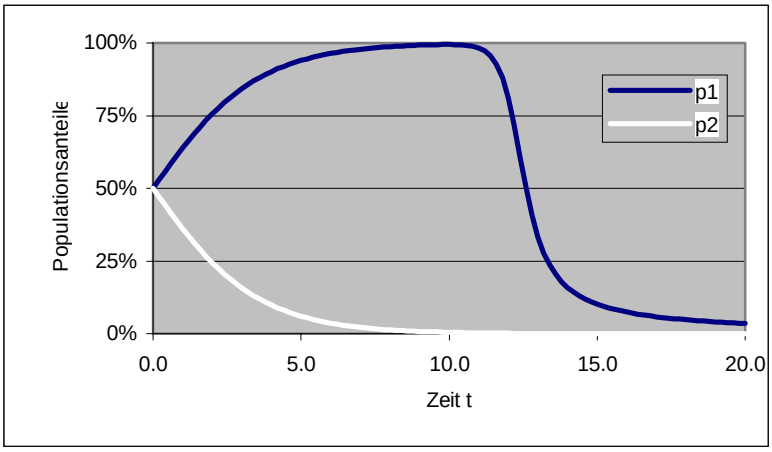
Im Praktikum zur Lehrveranstaltung "Simulation" fällt einer Arbeitsgruppe auf, dass das Arbeitsblatt

`http://www.fh-fulda.de/~grams/OekoSimSpiele/MemoLessStrategies.xls`

zur Ökologischen Simulation des Falke-Taube-Spiels bei bestimmten Parameterkonstellationen sichtlich falsche Ergebnisse liefert[1]. Beispielsweise ist das der Fall, wenn man für die ökologische Simulation die unten angegebene Spielmatrix des Falke-Taube-Spiels zu Grunde legt, die Strategien x_1= 25% und x_2=20% sowie die Schrittweite h=0.2 wählt.

Die Spielmatrix des Falke-Taube-Spiels		
(aTT, aTF)=	15	0
(aFT, aFF)=	50	-25

Es ergibt sich folgender Verlauf der Populationsanteile:



Anfangs, bis etwa t=11, sieht die Sache recht vernünftig aus und entspricht den Erwartungen. Danach wird die Bedingung $p_1+p_2 = 1$ verletzt. Das Ergebnis muss falsch sein. Was ist passiert? Woran könnte es liegen?

---

[1] Die Aufgabenstellung, die Definitionen der Variablen der ökologischen Simulation sowie die Modellgleichungen findet man im Kurs "Umweltsimulation mit Tabellenkalkulation":

`http://www.fh-fulda.de/~grams/oekosim.htm`

Fehlerhypothesen

Es werden Hypothesen zur Fehlerursache aufgestellt und erwogen:

(H1) Programmierfehler

(H2) Integrationsfehler (Verfahrensfehler des Euler-Cauchy-Verfahrens)

(H3) Rundungsfehler (Computerarithmetik)

Die erste Hypothese (H1) wird nach einer genauen Durchsicht der Formeln ausgeschlossen. Das Fehlerbild spricht auch gegen die zweite Hypothese. Eine Variation der Schrittweite zeigt, dass bei Wahl einer höheren Schrittweite der Fehler nicht etwa größer, sondern gar vollständig unsichtbar wird - beispielsweise im Fall h=0.8. Auch gegen die Rundungsfehlerhypothese (H3) spricht einiges: Die dem Arbeitsblatt zu Grunde liegende Differentialgleichung ist zwar nichtlinear, aber - im Sinne der Analysis - grundanständig: sie enthält nur Polynomausdrücke der Systemgrößen. Wie kann sich ein Rundungsfehler nach nur etwa 50 Rechenschritten bemerkbar machen?

Nun wird eine explorative Phase eingeschoben mit dem Ziel, das *Fehlerbild* klarer herauszuarbeiten.

Exploration

Das Arbeitsblatt wird zu Versuchszwecken erweitert: Die Tabelle wird auf 2000 Zeilen verlängert und die Zahlendarstellung in der Kontrollspalte des Arbeitsblattes wird so gewählt, dass 20 Stellen nach dem Komma sichtbar werden. So lassen sich sämtliche signifikanten Stellen der Zahlendarstellung überwachen. Die erste Abweichung in den 15 signifikanten Stellen tritt im obigen Beispiel ab dem 4 Zeitschritt auf. Jetzt stimmt die letzte Stelle nicht mehr. Ab dem 8 Schritt ist auch die vorletzte Stelle falsch. Nach weiteren drei Schritten die drittletzte, und so weiter. Etwa alle vier Zeitschritte geht eine weitere signifikante Stelle "verloren".

Beobachtung 1: Der Fehler betrifft zuerst nur die letzte signifikante Stelle.

Beobachtung 2: Der Fehler vergrößert sich mit fortschreitender Zeit exponentiell.

Die Beobachtung 1 spricht für einen Rundungsfehler. Nun wird in einer Reihe von Experimenten untersucht, wie sich der Parameter h und wie sich die Anfangsbedingungen auf die Fehlerausbreitung auswirken. Der Fehler ist weitgehend unabhängig von der Schrittweite und tritt immer etwa im Zeitintervall zwischen 5 und 15 ein (bei Schrittweiten von 0.005 bis 0.5).

Beobachtung 3: Die Fehlerausbreitung ist in erster Linie von der Zeit und nicht von der Anzahl der Rechenschritte abhängig.

Beobachtung 4: Der Fehler wirkt sich nicht immer in der gleichen Weise aus. Manchmal sinken alle Systemgrößen auf null ab, manchmal werden sie sehr groß.

Beobachtung 3 lässt die Vermutung zu, dass die Systemdynamik bei der Ausbreitung des Fehlers eine entscheidende Rolle spielt. Bei größeren Schrittweiten gilt das nicht mehr. Dann gewinnt der Integrationsfehler allmählich die Überhand. Die Beobachtung 4 spricht für die Rundungsfehlerhypothese (H3).

Analyse

Eine Analyse soll Aufschluss über die Fehlerfortpflanzung geben. Die im Arbeitsblatt verwendeten Variablen und Relationen werden zu diesem Zweck kurz zusammengestellt.

Mit den Elementen a_{ij} der Gewinnmatrix sind die Produktionsraten der Strategien 1 und 2 gegeben durch

$$r_1 = a_{11}\cdot p_1 + a_{12}\cdot p_2$$

$$r_2 = a_{21}\cdot p_1 + a_{22}\cdot p_2$$

Die Überschussproduktionsrate ist

$$u = r_1\cdot p_1 + r_2\cdot p_2$$

Die diskretisierten Systemgleichungen sind

$$p_1^+ = (1+h\cdot(r_1-u))\cdot p_1$$

$$p_2^+ = (1+h\cdot(r_2-u))\cdot p_2$$

Die Analyse soll vorerst auf die Kontrollvariable beschränkt werden. Sie wird mit p bezeichnet und ist die Summe der Wahrscheinlichkeiten aller Populationen:

$$p=p_1+p_2 \text{ bzw. } p^+=p_1^+ + p_2^+$$

Im fehlerfreien Fall ist stets $p= p^+=1$. Die diskrete Übergangsbeziehung ist gegeben durch

$$p^+ = p_1^+ + p_2^+ = (1+h\cdot(r_1-u))\cdot p_1 + (1+h\cdot(r_2-u))\cdot p_2$$

$$= p_1+ p_2 + h\cdot(r_1\cdot p_1 + r_2\cdot p_2 - u\cdot(p_1+ p_2)) = p + h\cdot(u - u\cdot p)$$

Schließlich haben wir also die Rekursionsbeziehung

$$p^+ = (1- h\cdot u)\cdot p + h\cdot u \qquad (*)$$

Die Überschussproduktionsrate u hängt von den p_i ab. Für die Abschätzung der Fehlerfortpflanzung können wir sie einmal als konstant ansehen. Wir lösen die Differenzengleichung (*). Es gilt $p(k\cdot h) = A\cdot(1- h\cdot u)^k + 1$. Bei hinreichend kleiner Schrittweite h ist näherungsweise $p(k\cdot h) = A\cdot e^{-u\cdot h\cdot k} + 1$ oder

$$p(t) = 1 + A\cdot e^{-u\cdot t} \qquad (**)$$

Hat sich irgendwo ein Fehler ε eingeschlichen, beispielsweise zum Zeitpunkt $t = 0$, dann ist

$$p(0) = 1 + \varepsilon$$

Mit dieser Anfangsbedingung geht die Gleichung (**) über in

$$p(t) = 1 + \varepsilon\cdot e^{-u\cdot t}$$

Ein einmal eingeschleppter Rundungsfehler pflanzt sich also gemäß dem Gesetz $\varepsilon\cdot e^{-u\cdot t}$ fort. Nur bei negativem u wächst dieser Fehler mit der Zeit an. Eine positive Überschussproduktionsrate u wirkt sich auf eingeschleppte Fehler dämpfend aus.

Bei der Tabellenkalkulation mit Excel ist der relative Rundungsfehler ungefähr gleich $\varepsilon = \pm 10^{-15}$. Zum Zeitpunkt

$$t = \ln(|\varepsilon|)/u \approx 35/(-u)$$

erreicht der Fehler die Größenordnung 1, dann ist er nicht mehr zu übersehen.

Was passiert, wenn anstelle des Euler-Cauchy-Verfahrens ein anderes Integrationsverfahren gewählt wird? Versuchsweise wird das Arbeitsblatt auf das Verfahren von Heun umgestellt. Es ergibt sich grundsätzlich dieselbe Fehlerdynamik - was ja auch zu erwarten ist. Die Sache

wird sogar noch schlimmer, denn nun lässt sich der Fehler sogar bei positiver Überschussproduktionsrate blicken, und zwar bei recht großer Schrittweite h. Rechnet man nach, dann ist in der Formel $p^+=(1-h\cdot u)\cdot p+h\cdot u$ die Überschussproduktionsrate u zu ersetzen durch den Ausdruck $(u+u^*-h\cdot u\cdot u^*)/2$. Hierin ist u die Überschussproduktionsrate im Zustand (p_1, p_2) und u^* die Überschussproduktionsrate im Zustand (p_1^*, p_2^*), der sich als erster Näherungswert für (p_1^+, p_2^+) nach der Euler-Cauchy-Formel ergibt.

Ergebnisse

Ob ein Rundungsfehler wesentlichen Einfluss gewinnen kann, hängt von der Überschussproduktionsrate u ab. Analyse und Experimente zeigen folgende Ergebnisse:

1. Eine positive Überschussproduktionsrate ist - zumindest beim Euler-Cauchy-Verfahren - harmlos. Eingeschleppte Fehler werden gedämpft.
2. Bei negativer Überschussproduktionsrate u wächst ein einmal eingeschleppter Rundungsfehler exponentiell mit der Zuwachsrate $|u|$ an.
3. Bei einer Zahlendarstellung mit 15-stelliger Mantisse muss man im Falle einer negativen Überschussproduktionsrate u etwa ab der Zeit $35/|u|$ mit einem völlig verfälschten Simulationsergebnis rechnen.

Abhilfemaßnahmen

Es werden zwei Vorschläge unterbreitet, die das Problem der Rundungsfehler lösen. Im ersten Fall wird das mathematische Verfahren geändert, im zweiten nur die Eingabe.

1. Die Systemgleichungen werden direkt für die Populationsgrößen N_i formuliert. Erst für die Ausgabe geht man zu den Anteilen $p_i = N_i/N$ mit $N=N_1+ N_2$ über.
2. Die Spielmatrix wird - unter Wahrung der Systemdynamik - so abgeändert, dass eine negative Überschussproduktion nicht mehr auftritt. Das geschieht durch Addition einer Konstanten c zu jedem Element der Spielmatrix. Dadurch ändern sich die Elemente der Auszahlungsmatrix entsprechend: $a_{ij} \rightarrow a_{ij} + c$. Dasselbe gilt für die Produktionsraten: $r_i \rightarrow r_i + c$, $u \rightarrow u + c$. Die Wachstumsraten in den Systemgleichungen bleiben durch Übergang auf die neue Spielmatrix unverändert: $(r_i + c) - (u + c) = r_i - u$. Durch geeignete Wahl von c lässt sich die Überschussproduktionsrate in den positiven Bereich transformieren. Dann werden eingeschleppte Fehler nicht mehr verstärkt, sondern gedämpft.

Zum Beispiel verschwinden die Rundungsfehler, wenn man zu jedem Element der anfangs gezeigten Spielmatrix den Wert c=25 addiert.

FH Fulda, 1. November 2000 (ergänzt am 03.12.00) Timm Grams