Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

Department
MATHEMATIK

Vorlesungsskript

Reaktiver Transport in porösen Medien

Sommersemester 2010, 2011, 2016, 2017, 2019, 2021, 2022, 2023

Prof. Dr. Serge Kräutle

am

Lehrstuhl für Angewandte Mathematik (Modellierung und Numerik)

AM1

— Version vom 29.04.2023 —

Friedrich–Alexander–Universität Erlangen–Nürnberg

Inhaltsverzeichnis

In theory, there is no difference between theory and practice.
In practice, there is.

(Autorschaft unbekannt bzw. umstritten)

1 Einführung: Was ist ein poröses Medium, Beispiele, Anwendungen

Ein *poröses Medium* besteht aus einem *Feststoffskelett* (auch: Feststoff- oder Bodenmatrix) und einem *Porenraum*. Die Struktur ist sehr kleinskalig im Vergleich zur Gesamtgröße des porösen Mediums. In den Poren befindet sich und fließt ein (oder auch mehrere unmischbare) Fluid(e). Unter einem Fluid verstehen wir sowohl Flüssigkeiten als auch Gase. Wichtigstes Beispiel für ein poröses Medium ist der Erdboden bzw. Gestein. Der Porenraum ist dort meist mit Wasser und/oder Luft gefüllt. Aber auch z.B. die Kombinationen Öl/Wasser ($\rightarrow$ Erdölförderung) oder auch CO_2/Wasser ($\rightarrow$ Lagerung von CO_2 im Boden zur Reduktion des Treibhauseffekts) werden betrachtet. Der Porenraum sollte zusammenhängend sein, um Bewegung des/der Fluid(e) zu erlauben.

Bitte beachten Sie: Ich habe einen Foliensatz (den Vorlesungsteilnehmer z.B. auf StudOn finden können) vorbereitet, der Bilder verschiedener poröser Medien beinhaltet, und in dem verschiedene Fragestellungen motiviert werden. Diesen Teil habe ich nicht in das vorliegende Vorlesungsskript aufgenommen.

Beispiele

poröses Medium	Anwendung
(1) Erdboden/Gestein/Sand, mit Grund- oder Tiefenwasser	(a) Altlasten (Verunreinigungen des Bodens/Grund- wassers u.a. durch Industriebetriebe, Unfälle, Müllkippen) (b) Erdöl/-gasgewinnung (c) CO_2-Lagerung im Erdboden (d) nukleare Endlager (e) Versalzung von Böden/Grundwasserleitern
(2) (historische) Gebäude	Verwitterung: Schwefel in Luft/Regenwasser kann Umwandlung des Steins ins Gips bewirken
(3) (Stahl-)Beton	Eindringen von Wasser und Sauerstoff bewirkt Korrosion des Stahls
(4) Porenbrenner	Wärmeerzeugung durch Gasverbrennung ohne offene Flamme
(5) Filteranlagen (Aktivkohle, Sandschüttung)	Reaktion oder Ablagerung von Verunreinigungen an Porenwänden
(6) landwirtschaftlich genutze Böden	Erforschung der komplizierten miteinander wechselwirkenden Prozessen, die zur Forderung der Bodenbildung und -erhalt oder zur Bodenverschlechterung beitragen

Im Forschungsschwerpunkt eines Teils des Lehrstuhls *Angewandte Mathematik - Modellierung und Numerik* (vormals: *AM1*) standen und stehen (1) und (6) aus obiger Tabelle im Fokus.

Charakteristisch für poröse Medien ist, dass verschiedene (räumliche) Skalen (=Größenordnungen) eine Rolle spielen: Da ist zunächst die *Poren-* oder *Mikroskala* (etwa $10^{-4}\text{m} = 10^{-1}\text{mm}$ oder kleiner), dann eine Skala, auf der die Bodeneigenschaften[1] 'einigermaßen konstant' sind, in etwa 10^{-2}–10^{2}m.[2] Dann gibt es noch die *Makroskala* (In den Geo-/Hydrologie genannt *Feldskala*), die die Größe des betrachteten Gebietes beschreibt, und die $10^2 - 10^4\text{m}$ oder sogar mehr beträgt.

2 Die Modellierung des Fließgeschehens

Wollten wir, im Sinne einer numerischen Simulation (FDM,FEM) die Porenskala eines $10^3 \times 10^3 \times 10^1\text{m}^3$ großen Gebietes $\Omega \subset \mathbb{R}^3$ auflösen, so müssten wir, bei einer Gitterweite $h = 10^{-5}\text{m} = 10^{-2}\text{mm}$, $10^8 \cdot 10^8 \cdot 10^6 = 10^{22}$ verwenden, was unmöglich ist (ganz zu Schweigen von den Schwierigkeiten, ein dreidimensiones, für numerische Zwecke verwendbares FEM-Gitter im Porenraum überhaupt erst zu generieren). Wir brauchen daher eine *makroskopische* Beschreibung der Vorgänge, eine Beschreibung durch *'effektive'* ('gemittelte') Gleichungen. Der einfachste Zugang ist die volumetrische *Mittelung*. Diese wollen wir im Folgenden verwenden. (Daneben gibt es noch als anderen Zugang die sog. *Homogenisierung*, auch genannt *Upscaling*; diese ist mathematisch präziser, macht allerdings die Annahme, dass der Porenraum periodisch ist, siehe Übungsaufgabe.)

Wir wählen ein sog. Repräsentatives Elementarvolumen (REV), das ist ein Würfel oder eine Kugel mit Durchmesser, der deutlich größer als die Porenskala, jedoch kleiner als die Korrelationslänge ist. Wir mitteln Größen über REVs um jeden Punkt x:

Porosität $\omega(x) := \frac{V_{por}(x)}{V_{tot}}$, wobei V_{tot} das totale Volumen des REVs ist und $V_{por}(x)$ das Volumen des Porenraums geschnitten mit dem REV. Es ist also $\omega \in (0, 1)$.
Bemerkung: In manchen Situationen kann ω auch eine Funktion von x und t sein, falls nämlich Mineralienauflösung oder -ausfällung die Große des Porenraums verändert oder falls sich Biofilme auf den Oberflächen der Poren bilden, oder falls sich Drücke

[1] Als Bodeneigenschaften kann man hier an Porosität und Leitfähigkeit denken, siehe später

[2] Es ist möglich, diese Größe präziser zu fassen im Rahmen von *stochastischen* Modellen des Bodens. Die Bodeneigenschaft φ an jedem Punkt x des Gebietes ist dann eine Zufallsvariable mit einer gewissen Zufallsverteilung, und zusätzlich wird gefordert, dass die *Korrelation* der Zufallsvariablen $\varphi(x)$ und $\varphi(y)$ eine (i.a. monoton fallende) Funktion von $|x-y|$ ist, d.h. je weiter Punkte voneinander entfernt sind, desto unkorrelierter ist die Bodeneigenschaften an den Punkten. Ein Parameter, der angibt, wie schnell diese Korrelations-Funktion abfällt, wird Korrelationslänge genannt. Als 'Länge auf der die Bodeneigenschaften einigermaßen konstant sind' kann man diese Korrelationslänge verwenden.

so stark ändern, dass dies Einfluss auf den Porenraum haben kann.

(Volumetrischer) Wassergehalt $\theta(t,x) := \frac{V_W(t,x)}{V_{tot}}$, wobei $V_W(t,x)$ das Volumen des Wassers innerhalb des REVs bezeichnet. Es ist also $V_W(t,x) \leq V_{por}(x)$, also $0 \leq \theta(t,x) \leq \omega(x)$.
Falls $\theta(t,x) = \omega(x)$, so nennen wir das poröse Medium an der Stelle x zur Zeit t *gesättigt*, andernfalls *ungesättigt*.

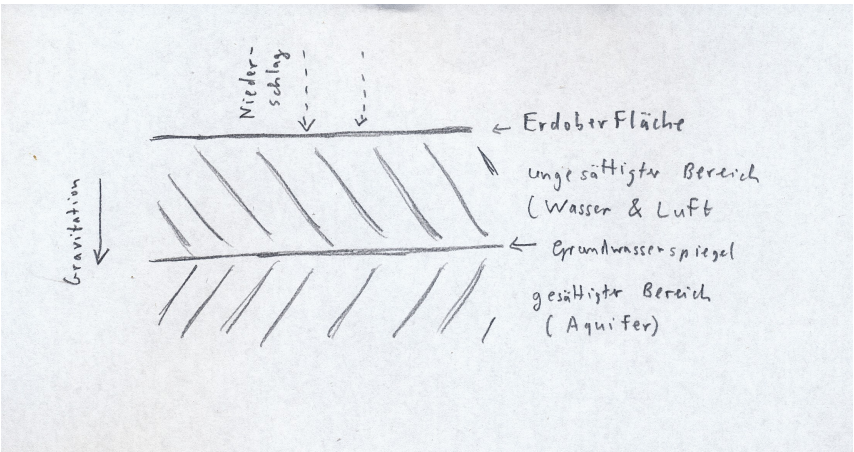


Abbildung 1: Typischer Verlauf des Wassergehalts in den obersten Bodenschichten in gemäßigten Zonen, schematisch.

Eine Mittlung der Geschwindigkeit der Fluidphase über das REV kann man auf zweierlei Artehn durchführen: Zum einen kann man die sog. *(Sicker-)Geschwindigkeit* des Wassers (engl.: *seepage velocity* oder *pore velocity*) $\vec{u}(t,\vec{x})$ betrachten; dies ist die mittlere Geschwindigkeit des Wassers, und zwar gemittelt (nur) über das $V_W(t,\vec{x})$-Volumen des REVs. Daneben kann gibt es noch die sog. *Darcy-Geschwindigkeit*[3] $\vec{v}(t,\vec{x})$; bei dieser ist die Geschwindigkeit über das totale Volumen des REVs (also Porenraum und Feststoffmatrix) gemittelt. Es gilt somit

$$\vec{v} = \theta\,\vec{u} + (1-\theta)\,\vec{0} = \theta\,\vec{u}\,.$$

Die Sickergeschwindigkeit ist beispielsweise relevant, wenn wir die Geschwindigkeit, mit der die Front einer Schadstoffverteilung voranschreitet (zumindest wenn wir Diffusion vernachlässigen und es keine Sorptionsprozesse gibt). die Darcy-Geschwindigkeit dagegen wird gebraucht, um den *Volumenfluss* durch eine makroskopische Fläche F (also das Wasservolumen, das sich pro Zeiteinheit durch die Fläche bewegt) zu beschreiben: Der Volumenfluss wird beschrieben durch das Oberflächenintegral ('zweiter Art', d.h. für vektorwertige Funktionen) über $\vec{v}$ über F; die Darcy-Geschwindigkeit stellt somit die *Volumenflussdichte* dar. Sucht man den *Massen*fluss bzw. die *Massen*flussdichte, muss man $\vec{v}$ noch mit der Dichte ρ multiplizieren, also $\rho\,\vec{v}$ bzw $\int_F \rho\,\vec{v}\cdot\vec{n}\,do$ betrachten.

[3] Henry Darcy, 1803-1858, Wasserbauingenieur, Dijon 1856.

Für jedes makroskopische Volumen $V \subseteq \Omega$ fordern wir den *Massenerhalt*. Die Änderung der Wassermasse in V pro Zeit sollte gleich dem Netto-Wasserfluss über den Rand von V pro Zeit sein. Somit

$$\frac{d}{dt} \int_V \rho\, \theta(t, \vec{x})\, d\vec{x} \stackrel{!}{=} - \int_{\partial V} \rho\, \vec{v} \cdot \vec{n}\, do\,,$$

wobei $\vec{n}$ der nach außen gerichtete Normaleneinheitsvektor auf ∂V ist. Der Leser mache sich klar, dass auf der rechten Seite das Vorzeichen nötig ist, um den Netto-Massenfluss nach *innen* zu beschreiben. Auf der rechten Seite benutzen wir den Satz von Gauß (engl.: the divergence theorem) und auf der linken Seite, eine gewisse Glattheit des Integranden vorausgesetzt, ziehen wir die Differenziation ins Integral. Wir bekommen

$$\int_V \frac{\partial}{\partial t}\, \rho\, \theta(t, \vec{x})\, d\vec{x} \stackrel{!}{=} - \int_V \nabla \cdot (\rho\, \vec{v})\, d\vec{x}\,.$$

Um die Integralzeichen loszuwerden, nehmen wir V als Kugel um einen Punkt $\vec{x} \in \Omega$ mit Radius r an, wir dividieren die obige Gleichung durch das Volumen der Kugel, und lassen dann r gegen null gehen. Im Grenzwert erhalten wir, sofern die Integranden hinreichend glatt sind (z.B. wenn wir die Integrale als Riemann-Integrale auffassen, ist dazu die Stetigkeit der Integranden hinreichend)

$$\boxed{\partial_t(\rho\, \theta) + \nabla \cdot (\rho\, \vec{v}) = 0}\,. \tag{2.1}$$

Sofern wir annehmen, dass die Dichte des Wassers konstant ist[4], vereinfacht sich die Gleichung des Massenerhalts zu

$$\partial_t(\theta) + \nabla \cdot \vec{v} = 0\,.$$

Dies ist *eine* Gleichung für $n+1$ Unbekannte $\vec{v}, \theta$. Wir benötigen also noch weitere Annahmen/Gleichungen.

2.1 Der gesättigte Fall

Im Fall, dass a priori bekannt ist, dass das Medium gesättigt ist, (oder wenn wir uns konzentrieren auf denjenigen Teil des Gebiets, der gesättigt ist), gilt $\theta = \omega$. Und die Porosität ω kann meist als zeitlich konstant angenommen werden. In diesem Fall bekommen wir also die Gleichung

$$\nabla \cdot \vec{v} = 0\,. \tag{2.2}$$

Wir benötigen nun noch ein konstitutives Gesetz. Henry Darcy fand 1856 experimentell für gesättigte poröse Medien das sog. *Darcy-Gesetz* (engl. *Darcy's law*)

$$\vec{v} = -K_{sat}\, \nabla p$$

[4]Die Dichte des Wassers kann ein wenig von Temperatur, eventuellen Verunreinigungen, sowie, falls extreme Drücke vorkommen, auch ein wenig vom Druck abhängen.

wobei p der *Druck* ist und K_{sat} als *hydraulische Leitfähigkeit* (engl. *hydraulic conductivity*) bezeichnet wird.
Kann man, neben experimentellen Befunden, noch weitere Begründungen oder Herleitungen für das Darcy-Gesetz finden?
Auf der Mikroskala (in der Pore) wird die Bewegung des Fluids beschrieben durch die Navier-Stokes-Gleichungen

$$\begin{aligned} \rho\,\partial_t\vec{w} + \rho\,(\vec{w}\cdot\nabla)\vec{w} - \eta\,\Delta\vec{w} + \nabla p &= \vec{f}\,, \\ \nabla\cdot w &= 0 \end{aligned}$$

(oder, da die Geschwindigkeiten gering sind, näherungsweise durch die Stokes-Gleichung) mit Null-Randbedingungen am Porenrand. η ist die dynamische Viskosität des Fluids. Die äußere Kraft(dichte) f wird hier hervorgerufen durch die Gravitation der Erde, also durch eine Potenzialkraft, d.h. es gibt zu diesem Kraftfeld ein Potential $\psi(x_1,...,x_n) = -\rho\,g\,x_n$, d.h. $\vec{f} = \nabla\psi = -\rho g\vec{e}_n$, wobei $\vec{e}_n$ der n-te Standardbasisvektor und $g = 9.81N/kg$ die *Erdbeschleunigung* ist. Indem wir $\hat{p} := p - \psi$ einführen, können wir die Navier-Stokes-Gleichung als

$$\begin{aligned} \rho\,\partial_t\vec{w} + \rho\,(\vec{w}\cdot\nabla)\vec{w} - \eta\,\Delta\vec{w} + \nabla\hat{p} &= \vec{0}\,, \\ \nabla\cdot w &= 0 \end{aligned}$$

schreiben, wobei man ψ als *hydrostatischen Druck* und $\hat{p}$ als *hydrodynamischen Druck* bezeichnen kann; $p = \psi + \hat{p}$.
Falls wir nun die extrem vereinfachende Annahme treffen, dass alle Poren gleichartige, in die gleiche Richtung zeigende Röhren mit konstantem Radius $R>0$ sind, so können wir die Navier-Stokes-Gleichung analytisch lösen. Unter der genannten Annahme an die Porengeometrie erhält man ein parabelförmiges Profil, das sog. *Poiseuille-Profil*. Es wird, wenn die Poren in x_1-Richtung zeigen und die Raumdimension $n = 3$ ist, beschrieben durch

$$w(r) = -\frac{\partial_{x_1}\hat{p}}{4\eta}\,(R^2 - r^2)\,, \quad r \in [-R, R]\,.$$

Durch Integration über eine Querschnittsfläche der Pore und unter Verwendung von Polarkoordinaten berechnet man den Volumenfluss durch die Pore:

$$\mathcal{F} = -\frac{\partial_{x_1}\hat{p}}{4\eta}\int_0^R\int_0^{2\pi} r\,(R^2 - r^2)\,d\varphi\,dr = -\frac{\partial_{x_1}\hat{p}}{4\eta}\,2\pi\left(\frac{1}{2}r^2R^2 - \frac{1}{4}r^4\right)\Big|_0^R = -\frac{\pi R^4}{8\eta}\,\partial_{x_1}\hat{p}\,.$$

Dies ist das sogenannte *Gesetz von Hagen-Poiseuille*. Der Fluss durch eine Pore mir Radius R (also Querschnittsfläche $\sim R^2$, hängt also, bei vorgegebenen Druckgradienten, mit der vierten Potenz von R ab! Durch Mittlung über die Querschnittsfäche (also durch Division durch πR^2) bekommen wir die makroskopische Sickergeschwindigkeit

$$\vec{u} = -\frac{R^2}{8\eta}(\partial_{x_1}\hat{p}, 0, 0)^T$$

und die Darcy-Geschwindigkeit

$$\vec{v} = \theta\,\vec{u} = -\theta\,\frac{R^2}{8\eta}(\partial_{x_1}\hat{p}, 0, 0)^T = -\theta\,\frac{R^2}{8\eta}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\nabla\hat{p}\,.$$

Wir können ablesen, dass die Darcy-Geschwindigkeit quadratisch vom Porenradius abhängt. Vor allem aber erkennen wir: Der hergeleitete Zusammenhang bestätigt das *Darcy-Gesetz*

$$\vec{v} = -K_{sat}\,\nabla\hat{p}\,, \tag{2.3}$$

wobei hier

$$K_{sat} = \theta\,\frac{R^2}{8\eta}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,. \tag{2.4}$$

Dieser Leitfähigkeitstensor spiegelt die extreme Anisotropie der angenommenen Porengeometrie wider; er ist dergestalt, dass das resultierende $\vec{v}$ – naturgemäß – immer in x_1-Richtung orientiert ist, selbst wenn der Druckgradient eine andere Richtung hat; als Antrieb wirkt nur die x_1-Komponente des Druckgradienten.
Setzen wir (2.3) in (2.2) ein, so bekommen wir das elliptische Problem

$$-\nabla\cdot(K_{sat}\,\nabla\hat{p}) = 0\,. \tag{2.5}$$

Ein reales poröses Medium wird isotroper sein als der oben durchgerechnete Fall. Im Fall völliger Isotropie (d.h. was die Porenausrichtung angeht, gibt es keinerlei ausgezeichnete Richtungen) würde man anstelle von (2.4 wohl eher

$$K_{sat} = \theta\,\frac{R^2}{8\eta}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.6}$$

annehmen. In diessem Fall kann man das elliptische Problem (2.5) auch mit *skalarem*

$$K_{sat} = \theta\,\frac{R^2}{8\eta}$$

schreiben. Bei einem moderat anisotropen porösen Medium kann die Matrix als voll besetzt angenommen werden. Drucken wir den hydrodynamischen Druck $\hat{p}$ wieder durch den physikalischen Druck aus, so lautet das Darcy-Gesetz (2.3)

$$\vec{v} = -K_{sat}\,\nabla(p+\rho g\vec{e}_3)\,. \tag{2.7}$$

Um diesen Zusatzterm im Druck besser zu verstehen folgende Erläuterung: In einem voll gesättigten porösen Medium, einem Grundwasserreservoir, das auf einer undurchlässigen Gesteinsschicht ruht, und auf das außer der Gravitation keine Kräfte wirken, wächst p nach unten (d.h. in $-\vec{e}_3$-Richtung) kontinuierlich an. Dennoch ist das Fluid in Ruhe,

d.h. $\vec{v}=\vec{0}$. Dies ist nur durch den kompensierenden Term $\rho g\vec{e}_3$ möglich. Dasselbe gilt in einem Swimming-Pool: Obwohl überall Kraft nach unten auf das Wasser wirkt, sorgt diese nicht dafür, dass sich das Wasser in Bewegung setzt. Lediglich der hydrodynamische Anteil des Drucks sorgt für Bewegungen des Wassers.
Abschließend beachte man auch noch, dass sich der Leitfähigkeitstensor (2.4) oder (2.6) schreiben lässt als

$$K_{sat} = \frac{\kappa_{sat}}{\eta}\,,$$

also als ein Zähler, der nur von der Geometrie des Porenraums abhängt, und einem Faktor, der vom Fluid abhängt, nämlich die Viskosität. Die Matrix oder der Skalar κ_{sat} wird als *Permeabilität* des porösen Mediums bezeichnet.
Eine weitere Möglichkeit, das Darcy-Gesetz herzuleiten, werden wir in der Übung kennenlernen, nämlich die asymptotische Entwicklung. Diese erlaubt eine recht komplexe Porengeometrie; sie nimmt jedoch an, dass der Porenraum periodisch ist.

2.2 Der ungesättigte Fall

Dieser wird in dieser Vorlesung nur sehr kurz abgehandelt.
Im zwanzigsten Jahrhundert (ab 1908) wurde das Darcy-Gesetz, das sich ursprünglich auf gesättigte poröse Medien bezog, auch auf ungesättigte übertragen:

$$\begin{aligned} \partial_t(\rho\,\theta) + \nabla\cdot(\rho\,\vec{v}) &= 0 \\ \vec{v} &= -K\,\nabla(p+\rho g x_n) \end{aligned} \tag{2.8}$$

Jedoch ändert sich dabei etwas beim Leitfähigkeitstensor. Das kann man leicht verstehen, wenn man sich vergegenwärtigt:

- In Kap. 2.1 haben wir herausgefunden, dass die mittlere Geschwindigkeit in großen Poren viel größer ist als in kleinen Poren.
- Bei einem ungesättigten Medium befindet sich aufgrund des Kapillareffekts das Wasser vorzugsweise in den kleinen Poren und Luft in den großen Poren.

In einem ungesättigten Medium wird somit die Sickergeschwindigkeit, d.h. die Leitfähigkeit, kleiner sein als im gesättigten Fall.[5] Es ist sinnvoll anzunehmen, dass die Leitfähigkeit monoton wachsend vom Wassergehalt abhängt, und dass bei einem sehr trockenen porösen Medium die Wasserphase nicht mehr zusammenhängend ist und somit gar kein Transport des Wassers stattfindet. Zusammengenommen modelliert man also

$$K = K(\theta) = k_{rel}(\theta)\,K_{sat}\,, \quad \text{mit } k_{rel}(0)=0\,,\ k_{rel}(\omega)=1,$$

$k_{rel} : [0,\omega] \to [0,1]$ monoton wachsend. Wenn wir dies in unsere Gleichungen (2.8) einsetzen, sind wir immer noch nicht fertig, denn anders als im gesättigten Fall enthält

[5] Als Gründe kann man anführen, dass in einem ungesattigten Medium weniger Volumen, in dem der Transport stattfindet, zur Verfügung steht, aber auch, dass die Variablilität von Porendurchmessern entlang einer Pore in Verbindung mit Kapillareffekten den Transport des Fluids behindern

unser Modell neben der Unbekannten p noch die Unbekannte θ. Man benötigt also ein weiteres konstitutives Gesetz. Hierzu postuliert man, dass sich der Wassergehalt als Funktion des Drucks schreiben lässt: $\theta = \theta(p)$
Wir erhalten das Modell (ρ=const angenommen)

$$\partial_t \, \theta(p) - \nabla \cdot [K(\theta(p)) \, \nabla(p + \rho g x_n)] \, . \tag{2.9}$$

Dies ist die sog. *Richards-Gleichung*[6]. Sie ist nichtlinear und von parabolischer Struktur. Für die (nichtlinearen) Isothermen $p \mapsto \theta$ und $p \mapsto K$ sind verschiedene Ansätze in Gebrauch, auf die wir hier nicht eingehen. Man beachte auch, dass i.a. a priori oft nicht klar ist, ob/wo gesättigte und ungesättigte Bedingungen vorherrschen. In Teilen des Rechengebiets gilt also die parabolische Richards-Gleichung, in anderen Teilen ist das Geschehen elliptisch (vgl. Kap. 2.1), und die Lage des Interfaces zwischen beiden Gebieten ist i.a. unbekannt und zeitlich variabel.

3 Herleitung der Transport-Reaktionsgleichung

Wir gehen in diesem Kurs davon aus, dass es in der Fluidphase eine 'dominante' chemische Spezies gibt (Wasser als dominante Spezies in der Wasserphase), und dass alle anderen in der Fluidphase gelösten Stoffe kleine Konzentrationen haben.
Die *Konzentration* einer im Wasser gelösten chemischen Spezies X bezeichnen wir mit $c(t, x)$, ist ≥ 0, und kann gemessen werden in Masse pro *Wasser*volumen $V_W(t, x)$, hat also die Einheit kg/l oder g/l. Wenn chemische Reaktionen modelliert werden, ist es sinnvoll stattdessen in Mol/l zu messen. Strenggenommen bezeichnet man diese Große als *Molarität*; im Folgenden verwende ich jedoch weiterhin den Begriff 'Konzentration', egal ob g/l oder mol/l. Ferner gibt es für die Molzahl von X bezogen auf die *Masse* des Wassers den Begriff *Molalität*.[7]
Ein Mol ist diejenige Menge eines Stoffs, die aus $6.022 \cdot 10^{23}$ Teilchen besteht. Diese Zahl wird auch Avogadro-Konstante genannt. Sie ist gerade so gewählt, dass gilt: Besteht ein Teilchen des Stoffs aus n Kernbausteinen (Protonen, Neutronen), so hat ein Mol dieses Stoffes die Masse n Gramm.[8]
Die Konzentration eines Stoffes bezogen auf das *Gesamt*volumen ist $\theta(t, x)c(t, x)$.

[6] Lorenzo Richards, 1931-93, USA, 1931

[7] Neben diesen beiden Möglichkeiten, also Masse des Stoffs pro Volumen und Mol des Stoffs pro Volumen, kann man auch noch Masse des Stoffs pro Masse der Fluidphase (=Massenbrüche) oder Mol des Stoffs pro Mol der Fluidphase (=Molbrüche) zur Messung von Konzentrationen angeben.

[8] Strenggenommen gilt diese Beziehung nur für Kohlenstoff C_{12}, für andere Stoffe gibt es minimale Abweichungen durch relativistische Masseneffekte.

Nun zur *Massenbilanz* eines im Fluid gelösten Stoffes im Volumen $\Omega' \subset \Omega$:

$$\begin{aligned} \frac{d}{dt}\int\limits_{\Omega'} \theta(t,\vec{\vec{x}})\, c(t,\vec{x})\, d\vec{x} &= \text{Netto-Zufluss durch Rand} + \text{Quellen (chem. Reakt.)} \\ &= -\int\limits_{\partial\Omega'} \vec{Q}\cdot\vec{n}\, d\vec{o} + \int\limits_{\Omega'} \theta(t,\vec{x})\, f(t,\vec{x})\, d\vec{x} \end{aligned}$$

Bei der Darstellung des Randintegrals wurde angenommen, dass sich der Massenfluss mittels eines Vektorfeldes $\vec{Q}$ darstellen lässt, auf dessen Normalkomponente $\vec{Q}\cdot\vec{n}$ es ankommt, wenn der Fluss über die Gebietsgrenzen zu berechnen ist.
Der Integrand auf der linken Seite hat die Dimension Masse pro (tot.) Volumen, somit hat die linke Seite die Dimension Masse pro Zeit. Es ergibt sich, dass das oben eingeführte Vektorfeld $\vec{Q}$ die Dimension Masse pro Fläche×Zeit hat.[9] Die Größe $\vec{Q}$ der Dimension

$$\frac{\text{Masse}}{\text{Fläche}\times\text{Zeit}} = \frac{\text{Masse}}{\text{Volumen}}\cdot\frac{\text{Strecke}}{\text{Zeit}}$$

wird *Flussdichte* des Stoffes X genannt. Man sieht an obiger Darstellung, dass sich eine Flussdichte als Produkt aus einer Konzentration (Masse/Volumen) und einer Geschwindigkeit (Strecke/Zeit) ergibt; diesen Sachverhalt werden wir gleich nochmal aufgreifen. Die Reaktionsdichte f hat die Dimension Masse pro Zeit×(Wasser-)Volumen. Für f werden wir später konkrete Modelle untersuchen. θf beschreibt die Reaktionsdichte als Masse pro Zeit×(tot.)Volumen.

Zum Massenfluss $\vec{Q}$ tragen in einem porösen Medium drei Anteile/Phänomene bei: Advektion, Diffusion und Dispersion:

$$\vec{Q} = \vec{Q}_{adv} + \vec{Q}_{diff} + \vec{Q}_{disp}$$

1. Advektion: Der Stoff X wird vom Strömungsfeld $\vec{u}$ mittransportiert:

$$\vec{Q}_{adv} = \theta\vec{u}c = c\vec{v}\,.$$

Man beachte, dass die Dimension von $\vec{Q}_{adv}$ zu der oben geforderten Dimension passt.
2. (molekulare) Diffusion: Die thermische Eigenbewegung der Teilchen (Brown'sche Bewegung) sorgt, makroskopisch betrachtet, für eine Netto-Wanderbewegung von Teilchen des Stoffes X, die in Bereichen mit hoher Konzentration sind, in Bereiche mit niedriger Konzentration. Es wird angenommen, dass der diffusive Fluss eine Funktion vom Konzentrationsgradienten ist, genauer, dass zwischen diesen ein linearer Zusammenhang besteht. Dies wird als *Fick'sches Gesetz* bezeichnet:

$$\vec{Q}_{diff} = -d_{diff}\,\theta\,\vec{\nabla}c$$

[9] Beachte, dass $\vec{Q}\cdot\vec{n}$ und $\vec{Q}$ die gleiche Dimension haben; denn $\vec{Q}\cdot\vec{n}$ ist einfach die Komponente von $\vec{Q}$ in Normalenrichtung, bzw. $\vec{n}$ wird als dimensionslos betrachtet, was dadurch zu motivieren ist, dass $\vec{n}$ mittels Division eines Normalenfeldes durch seine Norm erklärt ist.

Der Diffusionskoeffizient $d_{diff} > 0$ kann von anderen Parametern, insbesondere von der Temperatur, abhängen, er ist außer dem i.a. für jede Spezies ein anderer. Das Einfügen eines Faktors θ in obiges Gesetz ist plausibel, da die Größe des Porenraums (und nur dort kann ja Diffusion ablaufen) mittels θ gemessen wird.
Beachte: Advektion und Diffusion findet auch bei Strömungen *außerhalb* von porösen Medien statt. Diffusion findet auch dann statt, wenn das Fluid ruht ($\vec{u} = \vec{0}$).
3. (Kinematische) Dispersion: Dies ist ein Effekt, der nur speziell in porösen Medien vorkommt. (Nur) in porösen Medien gibt es mikroskopische Vorgänge, die wir bisher noch nicht in unserem makroskopischen Modell (mit den über REVs gemittelten Geschwindigkeiten) eingebaut haben (s. Abb. 2):
(a) In der Mitte der Poren ist die Geschwindigkeit größer als am Rand der Poren – siehe dazu auch unsere Überlegungen in Kap. 2.1, wo wir für homogene Röhrenförmige Poren ein Poiseuille-Profil hergeleitet hatten. Je nachdem, ob ein Teilchen des Stoffes X in Randnähe der Poren ist oder nicht, kommt es also schneller oder langsamer voran in Strömungsrichtung, als die Advektion angibt, dies führt zu einem zusätzlichen diffusionsartigen Effekt, vor allem in Strömungsrichtung.
(b) Teilchen, die sehr nahe beieinander sind, und die bei einem Fluid außerhalb eines porösen Mediums sich nur sehr langsam durch Diffusion voneinander entfernt hätten, können in einem porösen Medium gelegentlich in unterschiedliche Poren abbiegen und sich, dem Verlauf der Poren folgend, weiter voneinander entfernen, als es die molekulare Diffusion vorhersagt. Dieser Effekt wirkt sich makroskopisch aus wie eine zusätzliche Diffusion in Richtungen quer zur Strömungsrichtung $\vec{u}$. (c) Es gibt Poren, in denen Teilchen schneller vorankommen (dickere Poren erlauben größere Geschwindigkeiten, außerdem geradlinige Poren) als in anderen (schmalere Poren, gewundene Poren).

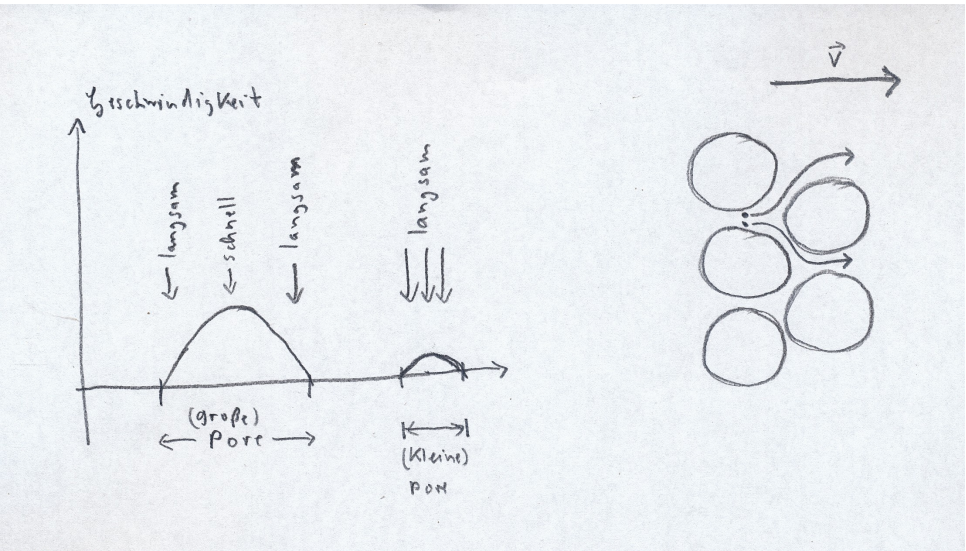


Abbildung 2: Links: Strömungsprofil innerhalb einer Pore. Rechts: Hauptursache der Querdispersion.

Die drei o.g. Effekte mit einfachen Formeln gut zu modellieren und zu quantifizieren ist schwierig, u.a., da sie stark von der Geometrie der Poren abhängt, außerdem von der Wahrscheinlichkeitsverteilung der Porengrößen, etc.[10] Ein weit verbreitetes Modell ist, die Dispersion analog zur molekularen Diffusion zu modellieren. Da allerdings die Effekte in Strömungsrichtung andere sind als quer zur Strömungsrichtung, ist es sinnvoll, längs und quer zur Strömung *unterschiedliche* Dispersionsparameter (Dispersionslängen) β_l, β_t (l=longitudinal, t=transversal) zu verwenden; eine solche Richtungsabängigkeit nennt man auch *Anisotropie*. Als Faustregel gilt, dass $\beta_l \approx 10\beta_t$. Im Fall, dass die Advektion $\vec{u}$ in x_1-Richtung geht, wäre also eine sog. tensorielle Diffusion

$$\vec{Q}_{disp} = -|\vec{v}|\, D_{disp,0}\, \vec{\nabla} c, \quad D_{disp,0} = \mathrm{diag}(\beta_l, \beta_t, \beta_t)$$

anzunehmen; der Faktor $|\vec{v}|$ erscheint plausibel, da die Stärke der o.g. Phänomene offenbar proportional zur Advektionsgeschwindigkeit ist (inbesondere verschwinden die Phänomene, wenn $\vec{v}=\vec{0}$ ist!). Dieses Gesetz zu übertragen auf den Fall, dass $\vec{u}$ nicht in x_1-Richtung zeigt, ist leicht, mittels einer Hauptachsentransformation: Im allgemeinen Fall ist

$$\vec{Q}_{disp} = -|\vec{v}|\, D_{disp}\, \vec{\nabla} c, \quad D_{disp} = X D_{disp,0} X^{-1},$$

wobei X die Orthogonalmatrix ist, die eine Drehung der Standard-Basis auf die Basis $\{\vec{v}, \vec{v}^{\perp}\}$ beschreibt (zur Vereinfachung nehme ich bei der folgenden Rechnung den 2-D Fall an; die Herleitung und das Ergebnis lässt sich auf den 3-D Fall übertragen):

$$X = \frac{1}{|\vec{v}|}\begin{pmatrix} v_1 & -v_2 \\ v_2 & v_1 \end{pmatrix}, \quad X^{-1} = \frac{1}{|\vec{v}|}\begin{pmatrix} v_1 & v_2 \\ -v_2 & v_1 \end{pmatrix}, \quad D_{disp,0} = \begin{pmatrix} \beta_l & 0 \\ 0 & \beta_t \end{pmatrix}$$

Wir spalten von $D_{disp,0}$ den Summanden $\beta_t\,\mathrm{Id}$ ab und erhalten

$$\begin{aligned} D_{disp} &= \beta_t X X^{-1} + \begin{pmatrix} v_1 & -v_2 \\ v_2 & v_1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 & v_2 \\ -v_2 & v_1 \end{pmatrix} \\ &= \beta_t\,\mathrm{Id} + \frac{\beta_l - \beta_t}{|\vec{v}|^2}\begin{pmatrix} v_1^2 & v_1 v_2 \\ v_1 v_2 & v_2^2 \end{pmatrix} = \boxed{\beta_t\,\mathrm{Id} + \tfrac{\beta_l-\beta_t}{|\vec{v}|^2}\,\vec{v}\vec{v}^T} \end{aligned}$$

Die Matrix D_{disp} wird *Bear-Scheidegger-Tensor*, das Dispersionsmodell das *Bear-Scheidegger-Modell* genannt (Scheidegger'61, Bear'72). In 3-D ergibt sich die gleiche Formel.

Insgesamt bekommen wir also die Flussdichte

$$\vec{Q} = -\Big(\underbrace{\theta\, d_{diff}\,\mathrm{Id} + \beta_t\,|\vec{v}|\,\mathrm{Id} + \frac{\beta_l - \beta_t}{|\vec{v}|}\,\vec{v}\vec{v}^T}_{=:D(\vec{v})}\Big)\,\vec{\nabla} c + c\vec{v}.$$

[10] Unter der idealisierenden Annahme, dass die Porenstruktur des Mediums *periodisch* ist, kann man mathematisch rigoros effektive Gleichungen herleiten, die sogar genauer sind als das von uns hier anvisierte sog. Bear-Scheidegger-Modell.

mit dem Diffusions-Dispersionstensor $D(\vec{v})$.

Dies in die Massenerhaltungsgleichung eingesetzt ergibt mit dem Satz von Gauß:

$$\begin{aligned}\frac{d}{dt}\int\limits_{\Omega'} \theta\, c\, d\vec{x} &= \int\limits_{\partial\Omega'} (D(\vec{v})\vec{\nabla}c - c\vec{v})\cdot\vec{n}\, d\vec{o} + \int\limits_{\Omega'} \theta\, f\, d\vec{x} \\ &= \int\limits_{\Omega'} \vec{\nabla}\cdot(D(\vec{v})\vec{\nabla}c - c\vec{v})\, d\vec{x} + \int\limits_{\Omega'} \theta\, f\, d\vec{x}\end{aligned}$$

Multiplikation dieser Gleichung mit $\frac{1}{|\Omega'|}$ und dann $|\Omega'| \to 0$ liefert, wenn die Integranden eine gewisse Regulariät haben (z.B. stetig in $\vec{x}$ sind), die PDE des Massenerhalts für die Spezies X

$$\boxed{\partial_t(\theta c) - \vec{\nabla}\cdot(D(\vec{v})\vec{\nabla}c - c\,\vec{v}) = \theta f},$$

die *Advektions-Diffusions-Dispersions-Reaktionsgleichung*, *Gleichung des reaktiven Transports*.

Zum Vergleich die entsprechende Gleichung für Transportvorgänge außerhalb von porösen Medien (d.h. $D_{disp}=0$, $\theta=1$):

$$\partial_t c - \vec{\nabla}\cdot(d_{diff}\vec{\nabla}c - c\vec{v}) = f$$

Falls d_{diff} konstant ist, kann man diese besonders leicht in Nicht-Divergenzform überführen:

$$\partial_t c - d_{diff}\Delta c + \vec{\nabla}\cdot(c\vec{v}) = f$$

Falls das Strömungsfeld divergenzfrei ist (inkompressibles Fluid) folgt

$$\partial_t c - d_{diff}\Delta c + \vec{v}\cdot\nabla c = f.$$

Bei Mehrspeziesproblemen ist die Diffusion speziesabhängig, die Dispersion jedoch nicht. Die Dispersion hängt von $|\vec{v}|$ ab, die Diffusion nicht. In der Praxis ist fast immer die Dispersion (deutlich) größer als die Diffusion (d.h. i.a. ist $|\vec{v}|$ hinreichend groß, so dass dies gilt), so dass häufig die Diffusion gegenüber der Dispersion vernachlässigt wird, so dass der Diffusions-Dispersionstensor als *spezies-unabhängig* angenommen wird. Welche Vorteile das haben kann, sehen wir später.

4 Allgemeine chemische Reaktionsraten

4.1 Einfache Beispiele

Eine chemische Reaktion wird mittels einer sog. *chemischen Gleichung* beschrieben, z.B.

$$\mathrm{X}_1 + \mathrm{X}_2 \longrightarrow \mathrm{X}_3.$$

Das bedeutet, dass ein Teilchen (Molekül, Ion,...) der Stoffs X_1 sich mit einem Teilchen des Stoffs X_2 verbindet, und sich daraus ein Teilchen des Stoffs X_3 ergibt. Es folgt sofort, dass sich jeweils 1 *Mol* von X_1 und von X_2 zu einem *Mol* X_3 reagieren.
Für die Quellterme f_1, f_2, f_3 der drei zugehörigen Konzentrationen ergibt sich also, dass f_1 und f_2 negativ und f_3 positiv ist; genauer, dass $\vec{f}$ die Form

$$\vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} -R \\ -R \\ +R \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ +1 \end{pmatrix} R$$

wobei $R \geq 0$ eine noch zu modellierende Reaktionsrate ist, und wobei Konzentrationen in Mol (nicht: Gramm) pro Liter gemessen werden.
Eine plausible Annahme ist, dass die Reaktionsrate proportional zur Begegnungswahrscheinlichkeit der miteinander reagierenden Teilchen ist, und dass diese wiederum proportional zum Produkt der beiden Konzentrationen ist:

$$R(\vec{c}) = k\, c_1\, c_2, \quad k > 0$$

Ein solches Modell ist recht einfach und zieht nicht in Betracht die genauen Auswirkungen von Kräften zwischen Teilchen (Van-der-Waals-Kräfte,...). Beachte, dass unsere PDEs des reaktiven Transports nun *gekoppelt* sind über den Quellterm, und dass dieser *nichtlinear* ist, wir es also mit *Systemen von nichtlinearen (genauer: semilinearen) partiellen Differentialgleichungen* zu tun haben.

Eine Übertragung auf etwas kompliziertere chemische Reaktionen kann wie folgt geschehen:
Bei der Reaktion

$$X_1 + 2\,X_2 \longrightarrow X_3 + X_4$$

verbinden sich ein Teilchen von X_1 mit zwei Teilchen von X_2, d.h. der Quellterm muss hier lauten

$$\vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} = \begin{pmatrix} -R \\ -2R \\ +R \\ +R \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ +1 \\ +1 \end{pmatrix} R.$$

Die Koeffizienten in der chemischen Gleichung heißen *stöchiometrische Koeffizienten*; sie treten als Vorfaktoren vor der Reaktionsrate auf, jedoch sind sie mit Vorzeichen versehen, je nachdem, auf welcher Seite des Reaktionspfeils sie stehen. Die Spezies auf der linken Seite einer Reaktionsgleichung heißen *Edukte*, die auf der rechten Seite heißen *Produkte*.
Was die Modellierung der Rate selbst angeht, so wird diese auch dann, wenn mehr als zwei Teilchen miteinander reagieren, häufig weiterhin über die Begegnungswahrscheinlichkeit aller beteiligten Teilchen modelliert, im Beispiel wäre das

$$R(\vec{c}) = k\, c_1\, c_2^2.$$

Sinnvoll ist es durchaus, solche Ratengesetze über Experimente abzusichern, was jedoch aufwändig ist – so hängt die in Experimenten beobachtete Reaktionsgeschwindigkeit auch davon ab, wie schnell und wie gut die Reaktanden gemischt werden. Außerdem sind Reaktionen, bei denen mehr als zwei Reaktanden miteinander reagieren, häufig in Wirklichkeit zusammengesetzt aus mehreren (teilweise sehr schnell) nacheinander ablaufenden Teilreaktionen, was zu anderen Ratenmodellen als dem obigen führen kann.

4.2 Mindestanforderungen an Reaktionsraten

Wir betrachten eine allgemeine chemische Reaktion; es sei I die Anzahl der Spezies:

$$s_1^e \,\mathrm{X}_1 + ... + s_I^e \,\mathrm{X}_\mathrm{I} \longrightarrow s_1^p \,\mathrm{X}_1 + ... + s_I^p \,\mathrm{X}_\mathrm{I}$$

Dabei seien $s_1^e, ..., s_I^e,\ s_1^p, ..., s_I^p \in \mathbb{R}_0^+$ (häufig: $\in \mathbb{N}_0$). Edukte sind diejenigen $\mathrm{X_i}$ für die $s_i^e > 0$ ist, und Produkte sind diejenigen $\mathrm{X_i}$, für die $s_i^p > 0$ ist. Beachte, dass wir hier zulassen, dass eine Spezies sowohl als Edukt als auch als Produkt in Erscheinung tritt ($\rightarrow$'Katalysator').
Als Mindestanforderung für Ratenfunktionen $R : (\mathbb{R}_0^+)^I \to \mathbb{R}$ legen wir fest: R sei stetig, und es gelte:

$$R(\vec{c}) = 0, \text{ falls } \exists\, i : \ s_i^e \neq 0 \ \wedge \ c_i = 0$$

(D.h. wenn eines der Edukte Konzentration=0 hat, kann die Reaktion nicht stattfinden.)

$$R(\vec{c}) > 0, \text{ falls } \forall\, i : \ s_i^e = 0 \ \vee \ c_i \neq 0$$

(D.h. wenn alle Edukte Konzentration>0 haben, dann läuft die Reaktion sicher ab.)
Wenn man den *Träger* eines Vektors

$$\begin{aligned} \mathrm{supp}\ \vec{c} \quad &:= \quad \{i \in \{1, ..., I\} \mid c_i \neq 0\}, \\ \mathrm{supp}\ \vec{s}^e \quad &:= \quad \{i \in \{1, ..., I\} \mid s_i^e \neq 0\} \end{aligned}$$

einführt, so kann man obige Bedingungen kurz schreiben als

$$\begin{aligned} R(\vec{c}) > 0 \quad &\Longleftrightarrow \mathrm{supp}\ \vec{s}^e \subseteq \mathrm{supp}\ \vec{c}, \\ R(\vec{c}) = 0 \quad &\text{sonst.} \end{aligned}$$

Für negative Konzentrationen haben wir die Ratenfunktion bisher nicht definiert, was physikalisch verständlich ist; aus mathematischen Gründen kann es jedoch sinnvoll sein, auch für negative Konzentrationen die Ratenfunktion fortzusetzen (z.B. indem man für $u \in \mathbb{R}^I \setminus (\mathbb{R}^+)^I$ setzt $R(u) := R(u^+)$, wobei $u^+ := \max\{u, 0\}$, wobei das Maximum komponentenweise gebildet werde; eine solche Fortsetzung ist stetig, da $x \mapsto \max\{x, 0\}$ stetig ist).[11]

[11] Weshalb sollte man dafür sorgen, dass die Raten auch für negative Konzentrationen wohldefiniert sind? Konzentrationen sind doch immer nichtnegativ, oder? Nun, *in der realen Wert* sind Konzentrationen immer nichtnegativ. Aber ob die Lösungen unserer mathematischen Modelle ebenfalls immer

4.3 Das Massenwirkungsgesetz

Das Massenwirkungsgesetz (MWG) ist nichts anderes als das Prinzip, das wir auf die Beispiele aus Kap. 4.1 angewendet haben:

$$\boxed{R(\vec{c}) = k \prod_{i=1}^{I} c_i^{s_i^e}}, \quad k>0,$$

mit dem Quellterm

$$\boxed{\vec{f} = (\vec{s}^p - \vec{s}^e)\, R(\vec{c}).}$$

Nun modellieren wir in gleicher Weise ein *System* von J chemischen Reaktionen:
Die chemischen Gleichungen sind

$$\begin{array}{ccc} s_{11}^e \,\mathrm{X}_1 + ... + s_{I1}^e \,\mathrm{X_I} & \longrightarrow & s_{11}^p \,\mathrm{X}_1 + ... + s_{I1}^p \,\mathrm{X_I} \\ \vdots & \vdots & \vdots \\ s_{1J}^e \,\mathrm{X}_1 + ... + s_{IJ}^e \,\mathrm{X_I} & \longrightarrow & s_{1J}^p \,\mathrm{X}_1 + ... + s_{IJ}^p \,\mathrm{X_I} \end{array}$$

Die stöchiometrischen Koeffizienten bilden nun Matrizen $S^e = (s_{ij}^e), S^p = (s_{ij}^p) \in (\mathbb{R}_0^+)^{I\times J}$. Die Konvention in dieser Vorlesung ist also: Zu jeder Reaktion gehört eine Spalte, zu jeder Spezies eine Zeile. Viele Autoren machen es genau umgekehrt. Der Quellterm lautet

$$\vec{f} = \sum_{j=1}^{J} \vec{f}_j = \sum_{j=1}^{J} R_j(\vec{c})\,(\vec{s}_j^e - \vec{s}_j^p) = \boxed{(S^p - S^e)\,\vec{R}(\vec{c})}$$

wobei $\vec{s}_j^e$ und $\vec{s}_j^p$ die j-te Spalte der betreffenden Matrix bedeutet und $\vec{R} = (R_j)_{j=1..J}$ der Vektor der Reaktionsraten ist. Dieser hat offensichtlich die Komponenten

$$\boxed{R_j(\vec{c}) = k_j \prod_{i=1}^{I} c_i^{s_{ij}^e}}, \quad k_j>0, \quad j = 1, ..., J.$$

4.4 Reversible Systeme

Oft wird die Annahme getroffen, dass alle Reaktionen sowohl 'vorwärts' als auch 'rückwärts' ablaufen können, d.h. in unserer Menge von Reaktionen $j = 1, ..., J$ treten

nichtnegativ sind, muss durch Beweise abgesichert werden. Und bevor so ein Beweis geführt ist, d.h. als Vorbedingung dafür dass so ein Beweis überhaupt geführt werden kann, ist es notwendig, die Raten auch für negative Konzentrationen zu definieren. Außerdem kann eine solche Definition notwendig sein, wenn wir an numerische Lösungsverfahren denken: Unsere Gleichungen sind nichtlinear, d.h. wir werden *iterative* Verfahren, z.B. das *Newton-Verfahren*, benutzen. Selbst wenn die Lösung immer nichtnegativ ist, so können aber Iterierte negativ sein. Wir wollen nicht, dass der Algorithmus in so einem Fall crasht. Siehe z.B. [Kr21], wo das Verhalten von Algorithmen beim Vorkommen von negativen Iterierten eine Rolle spielt.

dann alle Reaktionen paarweise auf; wir können so ein paar von Reaktionen durch die chemische Gleichung[12]

$$s^e_{1j}\,\mathrm{X}_1 + ... + s^e_{Ij}\,\mathrm{X}_\mathrm{I} \longleftrightarrow s^p_{1j}\,\mathrm{X}_1 + ... + s^p_{Ij}\,\mathrm{X}_\mathrm{I}, \quad j = 1, ...\tilde{J},$$

wobei j nun, um Konsistenz zu den früheren Kapiteln zu halten, von 1 bis $\tilde{J} := \frac{J}{2}$ läuft. Solche Reaktionen bzw. Systeme von Reaktionen heißen *reversibel*. Fassen wir die Raten ebenfalls paarweise zusammen, so bekommen wir

$$R_j(\vec{c}) = R^f_j(\vec{c}) - R^b_j(\vec{c}), \quad j = 1, ..., \tilde{J},$$

sowie den Quellterm

$$\begin{aligned}
\tilde{f} &= \sum_{j=1}^{\tilde{J}} R^f_j(\vec{c})\,(\vec{s}^{\,p}_j - \vec{s}^{\,e}_j) + R^b_j(\vec{c})\,(\vec{s}^{\,e}_j - \vec{s}^{\,b}_j) \\
&= \sum_{j=1}^{\tilde{J}} (R^f_j(\vec{c}) - R^b_j(\vec{c}))\,(\vec{s}^{\,p}_j - \vec{s}^{\,e}_j) = \boxed{(S^p - S^e)\,(\vec{R}^{\,f}(\vec{c}) - \vec{R}^{\,b}(\vec{c}))} \\
&= \boxed{S\,\vec{R}(\vec{c})}
\end{aligned}$$

wobei nun $S := S^p - S^e$ und $\vec{R}(\vec{c}) := \vec{R}^{\,f}(\vec{c}) - \vec{R}^{\,b}(\vec{c})$.

4.5 Weitere Ratengesetze

4.5.1 Massenwirkungsgesetz mit Aktivitätkorrektur

Genauere Untersuchungen zeigen, dass die Reaktionsraten nicht exakt proportional zur Konzentration sind, sondern eher zur sogenannten *Aktivität* der Spezies. Das MWG

$$R(\vec{c}) = k \prod_{i=1}^{I} c_i^{s_i}$$

wird somit korrigiert zu

$$R(\vec{c}) = k \prod_{i=1}^{I} a_i(\vec{c})^{s_i},$$

wobei $\vec{a} = (a_1, ..., a_I)^T$ der Vektor der Aktivitäten der Spezies $\mathrm{X}_1, ..., \mathrm{X}_\mathrm{I}$ ist. Jede Aktivität a_i wiederum ist Funktion des Vektors der Konzentrationen $\vec{c}$. Diesen Zusammenhang von Aktivität und Konzentration schreibt man oft in der Form

$$a_i(\vec{c}) = \gamma_i(\vec{c})\,c_i.$$

[12] Welchen Teil wir auf die rechte und welchen auf die linke Seite packen, d.h. welche Teilreaktion wir als 'vorwärts' und welche als 'rückwärts' bezeichnen, legen wir willkürlich fest.

Dabei heißt γ_i *Aktivitätskoeffizient* der Spezies X_i. Er ist eine Zahl, die meist näherungsweise eins ist, was das 'vereinfachte' MWG aus Kap. 4.3 rechtfertigt. Die Näherung aus Kap. 4.3, dass also $a_i \approx c_i$ ($\gamma_i \approx 1$) gilt umso besser, je kleiner die sog. Ionenstärke der Lösung ist; die Ionenstärke (s.u.) ergibt sich als gewichtete Summe der Konzentrationen von geladenen Teilchen. Eine in den Geowissenschaften vielfach verwendete *Aktivitätskorrektur* ist die *nach Debye-Hückel*:

$$\gamma_i = \exp\left(-\frac{Az_i^2\sqrt{H}}{1+r_iB\sqrt{H}}\right),$$

wobei $z_i \in \mathbb{Z}$ die Ladungszahl der Teilchen, $H = H(\vec{c}) = \frac{1}{2}\sum_{i=1}^{I} c_i z_i^2$ die Ionenstärke der Lösung, r_i der effektive Durchmesser der Teilchen ist (nachzulesen in Tabellen, in der Größenordnung von 10^{-10}m), sowie, bei 25 Grad Celsius, $A = 0.509\,(\mathrm{l/mol})^{1/2}$ und $B = 0.328 \cdot 10^{10}\,(\mathrm{m/mol})^{1/2}$.[13] Es ist also $\gamma_i \leq 1$, und $\gamma_i \longrightarrow 1$ für $\vec{c} \longrightarrow \vec{0}$.
Verwendet man, wie in Kap. 4.3, die Näherung $a_i = c_i$, so spricht man von *idealen Aktivitäten*.

4.5.2 Das Monod-Modell für biologischen Abbau

Der Stoffwechsel (Metabolismus) von Mikroorganismen führt zum Abbau gewisser organischer Substanzen, darunter auch 'Schadstoffe', weshalb man sich für diesen Abbau besonders interessiert. Der Metabolismus eines Organismus besteht aus Tausenden von chemischen Reaktionen, die weitgehend unerforscht sind. Es stellt sich jedoch heraus, dass, wenn man diese Reaktionen zusammenfasst, im wesentlichen eine Redoxreaktion abläuft, d.h. ein *Elektronenakzeptor=Oxidationsmittel A* (wie z.B. O_2, Fe(III), NO_3^- (=Nitrat), SO_4^{2-} (=Sulfat),...) und ein *Elektronendonator=Reduktionsmittel D* (z.B. ein organischer Schadstoffe wie der chlorierte Kohlenwasserstoff Chlorperethen=Tetrachlorethen C_2Cl_4) reagieren miteinander. Die per Saldo ablaufende Reaktion ist eine *Redoxreaktion*; das Reduktionsmittel wird oxidiert und das Oxidationsmittel wird reduziert. Unter Oxidation (Reduktion) versteht man eine Steigerung (Senkung) der sog. Elektronegativität, was durch Zu-Sich-Hinziehen oder Von-Sich-Wegdrücken von Elektronen geschieht. Eine Mikrobenpopulation X hat dabei eine Rolle, die mit einem Katalysator vergleichbar ist: Sie wird weder abgebaut noch aufgebaut durch die Reaktion. Jedoch profitiert die Mikrobenpopulation von der Reaktion dahingehend, dass ihre Vermehrungsrate von ihr abhängt. Das Abbauprodukt nennen wir P; ggf. wird für P gar keine Unbekannte und Differentialgleichung in das Modell aufgenommen, je nachdem, ob der Stoff als 'harmlos' eingestuft wird, oder der Stoff in weiteren interessanten Reaktionen eine Rolle spielt.

[13] Die Herleitung, bei der das Boltzmann-Modell der Thermodynamik (statistisches Modell zur Beschreibung der Wahrscheinlichkeitsverteilung der Geschwindigkeiten der Teilchen) mit der Elektrostatik (Poisson-Gleichung zur Beschreibung elektrostatischer Felder einer Punktladung) kombiniert werden, ergibt $B = \sqrt{2e^2N_A/(\epsilon k_B T)}$ und $A = e^2B/(8\pi\epsilon k_B T \ln 10)$, wobei k_B die Boltzmann-Konstante, e die elektrische Elementarladung, N_A die Avogadro-Konstante, ϵ die dielektrische Leitfähigkeit der Lösung und T die absolute Temperatur ist.

Der Metabolismus eines Mikroorganismus besteht aus Tausenden von Reaktionen; stark vereinfach stellt sich der Mechanismus dar als

$$s_A\,\mathrm{A} + s_B\,\mathrm{B} + \mathrm{X} \longrightarrow s_C\,\mathrm{P} + \mathrm{X}$$

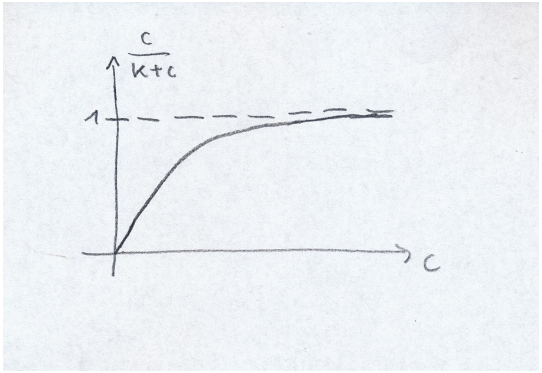
Beim *Monod-Model* geht man davon aus, dass die zugehörige Rate die Form

$$R(c_A, c_B, c_X) = k\,c_X\,\frac{c_A}{k_A + c_A}\,\frac{c_D}{k_D + c_D}, \quad \text{mit } k, k_A, k_D > 0,$$

hat. Die Quellterme des Dgl.-Systems lauten

$$\begin{aligned}
f_A &= -s_A\,R(c_A, c_B, c_X), \\
f_D &= -s_D\,R(c_A, c_B, c_X), \\
f_X &= \underbrace{\left(1 - \frac{c_X}{c_{X,max}}\right)}_{\text{Biomassenlimitierung}}\,R(c_A, c_B, c_X) - \underbrace{k_d\,c_X}_{\text{Sterbeterm}}, \\
f_P &= +s_P\,R(c_A, c_B, c_X) \quad \text{(falls } c_P \text{ ins Modell aufgenommen wird)}
\end{aligned}$$

Es gibt Varianten des Modells. So kann man z.B. den Sterbeterm weglassen oder die Biomassenlimitierung (die für $c_X < c_{X,max}$ sorgt), oder man kann in die Rate R weitere Faktoren, sogenannte Inhibitionsfaktoren der Form $k_Y/(c_Y + k_Y)$ einbauen um zu modellieren, dass ein Stoff Y den Prozess inhibiert/dem Mikroorganismus schadet. Der Stoff Y kann z.B. Sauerstoff sein, bei *anaeroben* Mikroorganismen.
Zur Motivation des Monod-Modells: Zunächst ist es plausibel, dass bei knappem Nahrungsangebot der Umsatz proportional zum Nahrungsangebot ist, bei Überversorgung allerdings ein – im wahrsten Sinne des Wortes – ein Sättigungseffekt eintritt, was Terme der Form $c/(c+k)$ plausibel macht.



Was die tiefere Begründung des Modells angeht, so ist diese etwas umstritten (es gibt ein Paper in dem 5 verschiedene Theorien zur Erklärung des Verhaltens diskutiert werden). Die am weitesten verbreitete Begründung des Monod-Modells ist folgende: Die Geschwindigkeit des Abbaus ist im wesentlichen bestimmt wird durch die am langsamsten ablaufende Teilreaktion, und diese ist eine Enzymreaktion, welche der

sog. *Michaelis-Menten-Kinetik* unterliegt. Eine Enzymreaktion nach Michaelis-Menten (1911) hat die Form

$$\mathrm{A} + \mathrm{E} \underset{k_2}{\overset{k_1}{\rightleftarrows}} K \xrightarrow{k_3} \mathrm{E} + \mathrm{P}.$$

Dabei ist E das (unkomplexierte) Enzym, K der Enzymkomplex, A das Edukt, P das Produkt. Man nimmt an, dass alle drei Reaktionen nach dem Massenwirkungsgesetz ablaufen:

$$\begin{aligned} c'_K &= k_1 c_A c_E - (k_2+k_3) c_K, \\ c'_E &= (k_2+k_3)\, c_K - k_1 c_A c_E, \\ c'_P &= k_3 c_K \end{aligned}$$

Per Addition folgt, dass $c'_K + c'_E = 0$, also $c_K + c_E = \text{const} =: c_{E0}$. Nun wird ferner die Annahme eines Gleichgewichtszustandes getroffen, d.h. es soll $c'_K = c'_E = c'_A = 0$ sein.[14] Wir erhalten also die Gleichungen

$$\begin{aligned} k_1 c_A c_E &= (k_2+k_3)\, c_K\,, \\ c_K + c_E &= c_{E0}\,, \\ c'_P &= k_3 c_K\,. \end{aligned}$$

Eliminieren wir c_E in der ersten Gleichung unter Verwendung der zweiten Gleichung, bekommen wir $k_1 c_A (c_{E0} - c_K) = (k_2+k_3) c_K$, was

$$c_K = c_{E0} \frac{c_A}{c_A + \frac{k_2+k_3}{k_1}}$$

ergibt, somit

$$c'_P = k_3 c_{E0} \frac{c_A}{c_A + \frac{k_2+k_3}{k_1}}.$$

Dies hat die Form eines Monod-Terms, mit Monod-Parameter $k = (k_2+k_3)/k_1$, was als Motivation für die Annahme solcher Terme im Monod-Modell dienen kann.

5 Das Batch-Problem/ODE-Modell

In diesem Kapitel beschäftigen wir uns mit einem allgemeinem Massenwirkungs-System unter Vernachlässigung von räumlichem Transport, d.h. unser Modell ist ein (nichtlineares) ODE-System. In den Geowissenschaften nennt man ein reaktives Problem, bei dem der Transport vernachlässigt wird, auch *Batch-Problem*.
Gerechtfertigt ist diese Betrachtung, wenn man an räumlich homogene (d.h. bzgl. x konstante) Situationen denkt. Der Grund für diese Vereinfachung ist, dass die Analysis

[14] was für $k_3 > 0$ nur dann möglich ist, wenn der Stoff A mit einer Rate $R_0 > 0$ nachgeliefert wird; es ist $c'_A = k_2 c_K - k_1 c_A c_E + R_0 = 0$

für ein ODE-System i.a. viel einfacher (weniger 'technisch') ist als für ein PDE-System, welches wir dann später in Kap. 7 ebenfalls noch untersuchen wollen. Viele der Eigenschaften, die man fürs ODE-System zeigen kann, lassen sich, wenn man den nötigen Aufwand betreibt, auch aufs PDE-Modell übertragen.

5.1 Positivität von Lösungen

Satz. Die Lösung des Batch-Problems

$$\vec{c}'(t) = S\,\vec{R}(\vec{c}(t))$$

mit positivem Anfangswert $\vec{c}(0) > \vec{0}$ und Massenwirkungskinetik (nicht notwendigerweise reversibel) und stöchiometrischen Koeffizienten $s^e_{ij}, s^p_{ij} \in \mathbb{N}$[15] ist auf jedem Existenzintervall echt positiv.

Beweis. Sei $[0,T)$ ein Existenzintervall der Lösung. Angenommen es gibt ein $i \in \{1,...,n\}$ und ein $t_1 \in [0,T)$ mit $c_i(t) \leq 0$. Da der Anfangswert strikt positiv ist, ist die kompakte Menge $\{t \in [0,t_1] \,|\, \exists j : c_j = 0\}$ nichtleer und hat somit ein kleinstes Element t_2, und $t_2 > 0$. Auf $[0,t_2)$ sind somit alle $c_i(t)$ strikt positiv, und es gibt ein i, so dass $c_i(t_2) = 0$.

Die i-te Komponente unseres ODE-Systems lautet

$$c_i'(t) = \sum_{j=1}^{J} (s^p_{ij} - s^e_{ij}) \underbrace{R_j(\vec{c}(t))}_{\geq 0}, \quad R_j(\vec{c}) = k_j \prod_{k=1}^{J} c_k^{s^e_{kj}}.$$

Auf $[0,t_2]$ ist $R_j(\vec{c}(t))$ offenbar nichtnegativ. Wir sortieren die Terme nach Vorzeichen. Ein Summand kann nur dann negativ sein, wenn $s^e_{ij} > s^p_{ij}$, also wenn $s^e_{ij} \geq 1$. Dann aber enthält das $R_j(\vec{c}(t))$ einen Faktor $c_i(t)^{s^e_{ij}}$ mit $s^e_{ij} \geq 1$. Die Terme mit positivem Vorzeichen fassen wir zusammen als Term $\alpha_0(t) \geq 0$ (dieser Term enthält c_i nicht, siehe Übungsaufgabe, was jedoch hier nicht essentiell ist). Die negativen Terme dagegen sind in polynomieller Form von c_i abhängig. Sie können somit als $\sum_{r=1}^{m} \alpha_r(t)\, c_i(t)^r$ geschrieben werden, wobei die $\alpha_r(t)$ die Abhängigkeiten von den c_j, $j \neq i$, beinhalten, und somit ebenfalls ≥ 0 sind auf $[0,t_2]$.[16]

Auf dem kompakten Intervall $[0,t_2]$ sind die stetigen Funktionen α_0, α_r beschränkt, also gibt es $C > 0$ mit $0 \leq \frac{c_i(t)}{C} \leq 1$, somit $0 \leq (\frac{c_i(t)}{C})^r \leq \frac{c_i(t)}{C} \forall\, r, i$, somit $c_i(t)^r \leq C^{r-1} c_i(t)$. Wir können somit die Potenzen $c_i(t)^r$ durch lineare Terme abschätzen:

$$c_i'(t) = \underbrace{\alpha_0(t)}_{\geq 0} - \sum_{r=1}^{m} \underbrace{\alpha_r(t)}_{\geq 0}\, c_i(t)^r \geq - \underbrace{\sum_{r=1}^{m} C^{r-1} \alpha_r(t)}_{\leq \tilde{C}}\, c_i(t)$$

[15] Dies kann zu $s^e_{ij}, s^p_{ij} \in \{0\} \cup [1,\infty)$ abgeschwächt werden.

[16] Wichtig ist hier, dass die Summation bei $r = 1$ und nicht bei $r = 0$ beginnt.

Der Term $\sum_{r=1}^{m} C^{r-1}\alpha_r(t)$ ist wegen seiner stetigen Abhängigkeit von $t \in [0, t_2]$ beschränkt durch eine Konstante $\tilde{C} > 0$. Wir landen also, da auch $c_i(t) \geq 0$ ist auf $[0, t_2]$, bei

$$\frac{c_i'(t)}{c_i(t)} \geq -\tilde{C} \quad \forall t \in [0, t_2).$$

Integration liefert wegen der Monotonie des Integrals

$$\int_0^t \frac{c_i'(t)}{c_i(t)}\, dt \geq -\tilde{C}t$$

für $t \in [0, t_2]$, somit (Substitution $u := c(t)$)

$$c_i(t) \geq c_i(0)\, \exp(-\tilde{C}t), \quad t \in [0, t_2).$$

Dies ist ein Widerspruch zur Stetigkeit von c_i, da $c_i(t_2) = 0$. □

5.2 Beschränktheit von Lösungen, Existenz von globalen Lösungen

In diesem Kapitel werden zwei verschiedene Wege beschritten, wie man Beschränktheit von Lösungen des Batch-Problems bei Massenwirkungskinetik zeigen kann.

Die Übungsaufgabe 1 zeigte uns einen Weg auf, wie man unter Verwendung von Invarianten sowie der aus Kap. 5.1 hervorgehenden Nichtnegativität die Beschränktheit von Lösungen ('nach oben') zeigen kann:

Satz. Es gelten die Voraussetzungen des vorangegangenen Satzes, und es sei $\mathcal{S}^\perp \cap (\mathbb{R}^+)^I \neq \emptyset$. Dann ist die Lösung des Batch-Problems mit Massenwirkungskinetik (nicht notwendigerweise reversibel) auf jedem Existenzintervall $[0, T)$ beschränkt durch eine Konstante (unabhängig von T).
Dabei sei $\mathcal{S} := \text{Bild}(S)$ der Spaltenraum der Matrix S und $\mathcal{S}^\perp$ sein orthogonales Komplement.

Beweis. Sei $\vec{s}^\perp \in \mathcal{S}^\perp \cap (\mathbb{R}^+)^I$. Wir bilden die Linearkombination

$$\Phi(t) := \sum_{i=1}^{I} s_i^\perp\, c_i(t).$$

Dieses Φ ist eine Erhaltungsgröße:

$$\Phi'(t) = \langle s^\perp, \vec{c}\,'(t)\rangle = \langle s^\perp, S\vec{R}(\vec{c}(t))\rangle = \langle \underbrace{S^T s^\perp}_{=\vec{0}}, \vec{R}(\vec{c}(t))\rangle = 0\,,$$

denn $s^\perp \in \mathcal{S}^\perp = \text{Bild}(S)^\perp = \text{Kern}(S^T)$. Es ist also $\Phi(t) = \Phi(0) = \text{const} = \langle s^\perp, \vec{c}(0)\rangle$. Da $s_i^\perp \neq 0$, können wir die Gleichung $\sum_{i=1}^{I} s_i^\perp \, c_i(t) = \langle s^\perp, \vec{c}(0)\rangle$ nach $c_i(t)$ auflösen:

$$c_i(t) = \frac{1}{s_i^\perp} \left(\langle s^\perp, \vec{c}(0)\rangle - \sum_{j \neq i} s_j^\perp \, c_j(t) \right)$$

Da die $s_j^\perp \geq 0$ und die $c_i(j) \geq 0$, folgt

$$c_i(t) \leq \frac{1}{s_i^\perp} \, \langle s^\perp, \vec{c}(0)\rangle \, . \qquad \square$$

Folgerung. Unter den Voraussetzungen des obigen Satzes existiert die Lösung der Batch-Problems mit Massenwirkungskinetik auf ganz $[0, \infty)$; d.h. das Batch-Problem hat eine *globale* Lösung.

Der Grund dafür ist ein Satz, den man z.B. in dem Buch [Ha64], über ODE-Systeme mit lokal-lipschitzstetiger rechter Seite findet:
Gibt es eine Funktion $f : [0, \infty) \to \mathbb{R}$, so dass jede lokale Lösung des ODE-Systems auf ihrem Existenzintervall die Bedingung $|\vec{c}(t)| \leq f(t)$ erfüllt, so existiert die Lösung auf ganz $[0, \infty)$. (Hintergrund: Es gibt immer eine 'maximale' Lösung, d.h. eine Lösung, die 'bis an den Rand' des (ggf. unbeschränkten) Gebietes, z.B. $[0, \infty) \times \mathbb{R}^n$, auf dem die rechte Seite lokal L-stetig ist, reicht. Die Annahme einer Schranke $f(t)$ sorgt dafür, dass die maximale Lösung bis $t = \infty$ gehen muss.)

Physikalische (Be-)Deutung. In 'realen' chemischen Systemen ist jedes Teilchen (Molekül) aus einer echt positiven Anzahl von Atomen zusammengesetzt. Setzt man $s_i^\perp$ als Anzahl der Atome, aus denen ein Molekül des Stoffs X_i besteht (was offenbar eine strikt positive Zahl ist), so ist $\vec{s}^\perp$ orthogonal zu jeder Spalte der Matrix S (dies besagt nichts anderes als dass die Anzahl der Atome auf der linken und auf der rechten Seite einer chemischen Gleichung gleich sein müssen, Erhalt der Anzahl der Atome). Somit erfüllt der so gebildete Vektor $\vec{s}^\perp$ die Voraussetzung des obigen Satzes. Die Voraussetzung des Satzes $\mathcal{S}^\perp \cap (\mathbb{R}^+)^I \neq \emptyset$ ist also realistisch.

Andere Invarianten sind z.B. die Anzahl der Atome einer gewissen Atom-Sorte (z.B. C-Atome) oder die Ladungszahl; deren Komponenten sind jedoch i.a. nur nichtnegativ statt strikt positiv, d.h. deren Verwendung und kann nur die Beschränktheit von gewissen Komponenten des Lösungsvektors zeigen.

Beachte, dass wir hier die Zusammensetzung von Teilchen aus Atomen zum ersten mal verwenden.
Es gibt noch einen ganz anderen Zugang zur Beschränktheit von Lösungen, der auf sog. *Lyapunov-Funktionen* beruht, und der das Bestehen der Teilchen aus Atomen nicht

verwendet, dafür allerdings nur für *reversible* Systeme funktioniert. Als Motivation, diesen neuen Zugang ebenfalls zu verfolgen, mag dienen, dass er
1. ganz nützlich sein wird später für das PDE-Modell und
2. dass dieser auch bei sog. *seltsamen* (engl.: *peculiar*) reaktiven Netzwerken funktioniert. Beispiele für seltsame Netzwerke sind: 2.a.: Bei Reaktionen mit Wasser (oder mit Stoffen, die in großen Mengen vorhanden sind, und deren Menge durch Reaktionen kaum beeinflusst wird) der Stoff 'Wasser' oft eliminiert wird: Statt der chemischen Gleichung $2\mathrm{H}^+ + \mathrm{OH}^- \longleftrightarrow \mathrm{H_2O}$, zu der die Rate $R^v(\vec{c}) = k^f\, c_1^2 c_2$, $R^r(\vec{c}) = k^r\, c_3$ gehören, integriert man die (de facto konstante) Konzentration von Wasser, c_3, in die Reaktionskonstanten und bekommt hier $\tilde{R}^v(\vec{c}) = k^f\, c_1^2 c_2$, $\tilde{R}^r(\vec{c}) = \tilde{k}^r$, was wiederum der chemischen Gleichung $2\mathrm{H}^+ + \mathrm{OH}^- \longleftrightarrow$ 'nichts' entspricht; in einer solchen chemischen Gleichung gibt es keinen Atomerhalt, und es gibt kein $\vec{s}^{\perp}$ mit positiven Einträgen, das zu der Spalte $(1,1)^T$ der zugehörigen neuen stöchiometrischen Matrix orthogonal ist.
2.b.: Auch *Zu-* und *Abfluss* aus einem chemischen Reaktor kann mittels Massenwirkungskinetik für (Pseudo-)Reaktionen 'nichts' $\longrightarrow \mathrm{X_i}$ bzw. $\mathrm{X_i} \longrightarrow$ 'nichts' modelliert werden.

Die Lyapunov-Technik (für ODEs) Für ein ODE-System $\vec{y}'(t) = \vec{f}(t, \vec{y}(t))$ finde ein Funktional φ, so dass

1. $\frac{d}{dt}\varphi(\vec{y}(t)) \leq 0$, wobei $t \to \vec{y}(t)$ die Lösung der ODE ist ('φ fällt entlang von Lösungen'),

2. aus der Beschränktheit von $t \to \varphi(\vec{y}(t))$ die Beschränktheit von $t \to \vec{y}(t)$ folgt (also z.B. eine Abschätzung der Form $|\vec{y}| \leq c_1\varphi(\vec{y}) + c_2$ für φ gilt).

Die Argumentation ist dann wie folgt:
Aus 1. folgt, dass $\varphi(\vec{y}(t)) \leq \varphi(\vec{y}(0)) = \text{const}$, aus 2. folgt dann $|\vec{y}(t)| \leq c_1\varphi(\vec{y}(t)) + c_2 \leq c_1\varphi(\vec{y}(0)) + c_2$. Aus einer solchen Schranke folgt wieder die Existenz einer *globalen* Lösung.

Bemerkungen.

- Es sind Varianten/Abschwächungen der Voraussetzung 1. denkbar. So reicht es auch, ein mögliches *Anwachsen* von φ entlang von Lösungen zu 'kontrollieren', z.B. durch Abschwächung von 1. zu $\frac{d}{dt}\varphi(\vec{y}(t)) \leq g(t)$ oder zu $\frac{d}{dt}\varphi(\vec{y}(t)) \leq c\,\varphi(\vec{y}(t))$.

- Ist ein ODE-System ohne konkreten Anwendungshintergrund gegeben, so ist es oft sehr schwierig, eine geeignete Funktion φ zu finden. Bei Anwendungsproblemen ist es häufig erfolgversprechend, als φ physikalische Größen wie irgendwelche Energien als φ zu benutzen; bei reibungsbehafteten Vorgängen (Bewegung in einem Gravitationsfeld, Bewegung eines Federschwingers...) vielleicht die Summe

aus kinetischer und potenzieller Energie.
Mathematisch betrachtet zeigt die Rechnung

$$\frac{d}{dt}\varphi(\vec{y}(t)) = \langle \nabla\varphi(\vec{y}(t)), \vec{y}'(t)\rangle \overset{\text{ODE}}{=} \langle \nabla\varphi(\vec{y}(t)), f(t,\vec{y}(t))\rangle,$$

dass als Anforderung an φ zu stellen ist, dass $\langle \nabla\varphi(\vec{y}(t)), f(t,\vec{y}(t))\rangle$ nichtpositiv oder zumindest 'nicht zu groß' wird.

Im vorliegenden Fall eines reversiblen reaktiven Problems mit Massenwirkungskinetik ist ein geeignetes Funktional

$$\varphi(\vec{c}) := \sum_{i=1}^{I} (\mu_i - 1 + \ln c_i)\, c_i + e^{1-\mu_i}, \quad (\mathbb{R}_0^+)^I \longrightarrow \mathbb{R},$$

wobei der Vektor $\vec{\mu}$ eine Lösung des LGS

$$S^T\vec{\mu} = -\ln\vec{K}$$

ist, $S = S^p - S^e$, $\vec{K} \in \mathbb{R}_+^{\tilde{J}}$, $k_j = \frac{k_j^v}{k_j^r}$.
Motivation/physikalische Bedeutung von φ: Die Konstruktion von φ ist inspiriert von einer Größe (einem 'Potenzial') aus der Thermodynamik von Gemischen, der sog. *freien Gibbs-Energie*. Anschaulich kann man sich darunter eine Art von chemischer Energie vorstellen; das System versucht durch Ablaufenlassen der Reaktionen, seine chemische Energie zu minimieren ($\rightarrow$ Erwartung: φ monoton fallend entlang von Lösungen). Die additive Konstante $e^{1-\mu_i}$ verschiebt lediglich das Nullniveau von φ, was dazu führt, dass man in Punkt 2. mit $c_2 = 0$ durchkommt (man kann auch ohne additive Konstante arbeiten).
Existenz einer Lösung $\vec{\mu}$ des obigen LGS: Hierzu setzen wir voraus, dass die Spalten von S linear unabhängig sind; es ist dann $\operatorname{rang}(S) = \operatorname{rang}(S) = \tilde{J}$, somit ist das Bild von S der gesamte $\mathbb{R}^{\tilde{J}}$, d.h. das LGS ist für beliebige rechte Seite lösbar (die Lösung ist i.a. nicht eindeutig; eine Lösung ist offenbar $\vec{\mu} = -S(S^TS)^{-1}\ln\vec{K}$; die Gesamtheit der Lösungen erhält man als $\operatorname{Kern}(S^T) - S(S^TS)^{-1}\ln\vec{K}$).

Satz. Für das Batch-Problem mit reversibler Massenwirkungskinetik ist die Lösung auf jedem Existenzintervall $[0,T)$ beschränkt durch eine Konstante (unabhängig von T).

Beweis. Wir errechnen $\frac{\partial\varphi}{\partial c_i} = \mu_i + \ln c_i$, also

$$\nabla\varphi(\vec{c}) = \vec{\mu} + \ln\vec{c},$$

wobei der ln komponentenweise zu verstehen ist. Da die Lösung $t \mapsto \vec{c}(t)$ strikt positiv ist, können wir $\varphi(\vec{c}(t))$ und sogar $\nabla\varphi(\vec{c}(t))$ bilden. (Anmerkung: φ ist per stetiger

Fortsetzung auch am *Rande* des positiven Rektanden definiert, $\nabla\varphi$ jedoch nicht.) Es folgt

$$\begin{aligned}
\frac{d}{dt}\varphi(\vec{c}(t)) &= \langle \nabla\varphi(\vec{c}(t)), \vec{c}\,'(t)\rangle \\
&= \langle \vec{\mu} + \ln \vec{c}(t), S\vec{R}(\vec{c}(t))\rangle \\
&= \langle S^T(\vec{\mu} + \ln \vec{c}(t)), \vec{R}(\vec{c}(t))\rangle \\
&= \langle -\ln \vec{K} + S^T \ln \vec{c}(t), \vec{R}(\vec{c}(t))\rangle \\
&= \sum_{j=1}^{J} [-\ln K_j + (S^T \ln \vec{c}(t))_j]\, R_j(\vec{c}(t)) \\
&= \sum_{j=1}^{J} \underbrace{\left[-\ln \overbrace{K_j}^{=\frac{k_j^v}{k_j^r}} + \sum_{i=1}^{I} (s_{ij}^p - s_{ij}^e) \ln c_i(t) \right]}_{\text{(I)}} \underbrace{\left[k_j^v \prod_{i=1}^{I} c_i(t)^{s_{ij}^e} - k_j^r \prod_{i=1}^{I} c_i(t)^{s_{ij}^p} \right]}_{\text{(II)}}
\end{aligned}$$

wobei

$$\begin{aligned}
\text{(II)} \gtreqless 0 \quad &\Longleftrightarrow \quad \ln k_j^v + \sum_{i=1}^{I} s_{ij}^e c_i \gtreqless \ln k_j^r + \sum_{i=1}^{I} s_{ij}^p c_i \\
&\Longleftrightarrow \quad \text{(I)} \gtreqless 0
\end{aligned}$$

Somit ist $\varphi \circ \vec{c}$ monoton fallend, also $\varphi(\vec{c}(t)) \le \varphi(\vec{c}(0)) =: \varphi_0$.
Es bleibt noch, aus der Beschränktheit von $\varphi \circ \vec{c}$ die Beschränktheit von $\vec{c}$ zu folgern. Dazu eine kurze 'Kurvendiskussion' für φ:
Wir schreiben

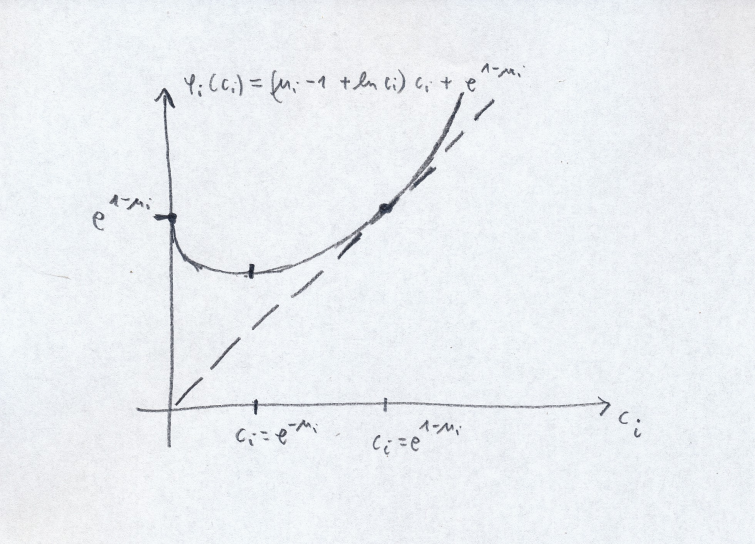
$$\varphi(\vec{c}) = \sum_{i=1}^{I} \varphi_i(c_i) \quad \text{mit } \varphi_i(x) := (\mu_i - 1 + \ln x)\, x + e^{1-\mu_i}.$$

Es ist offensichtlich $\lim\limits_{x\to 0} \varphi_i(x) = e^{1-\mu_i}$, $\lim\limits_{x\to\infty} \varphi_i(x) = \infty$. Die einzige Extremalstelle (=Minimalstelle) erhalten wir mit

$$\varphi_i'(x) = 0 \quad \Leftrightarrow \quad \mu_i + \ln x = 0 \quad \Leftrightarrow \quad x = e^{-\mu_i};$$

Der Minimalwert von φ_i ist somit $\varphi_i(e^{-\mu_i}) = e^{-\mu_i}(e-1)$.
$\varphi : (\mathbb{R}_0^+)^I \to \mathbb{R}$ hat also eine positive untere Schranke. Ferner ist $\frac{d}{dx}(\varphi_i(x) - x) = \mu_i + \ln x - 1 = 0 \Longleftrightarrow x = e^{1-\mu_i}$ und $(\varphi_i(x) - x)|_{x=\exp(1-\mu_i)} = 0$, d.h. der Graf von φ_i berührt die Identität $f(x) = x$ genau einmal (an der Stelle $x = e^{1-\mu_i}$) und verläuft sonst oberhalb. Somit

$$\varphi_i(x) \ge x \quad \forall x \in \mathbb{R}_0^+.$$



Zusammen mit der Nichtnegativität der φ_j folgt

$$\varphi(\vec{c}) \geq \sum_{i=1}^{I} c_i, \quad \text{insbesondere } \varphi(\vec{c}) \geq c_i \quad \forall i = 1, ..., I.$$

□

Der stöchiometrische Raum und Veranschaulichung der beiden Zugänge zur Beschränktheit von Lösungen.
Integration des ODE-Systems liefert

$$\vec{c}(t) - \vec{c}(0) = S \int_0^t \vec{R}(\vec{c}(\tau))\, d\tau,$$

somit liegt der Vektor $\vec{c}(t) - \vec{c}(0)$ immer im Raum $\text{Bild}(S) = \text{Kern}(S^T)^\perp$. Die Lösung kann also den affinen Unterraum

$$\vec{c}(t) \in \vec{c}(0) + \text{Bild}(S) = \vec{c}(0) + \text{Kern}(S^T)^\perp$$

nicht verlassen. Der Vektorraum $\mathcal{S} := \text{Bild}(S) = \text{Kern}(S^T)^\perp$ heißt *stöchiometrischer Raum*. Der affine Raum $\vec{x} + \mathcal{S}$ heißt *stöchiometrische Klasse* des Vektors $\vec{x} \in \mathbb{R}^I$.

Falls es einen Vektor $s^\perp \in \mathcal{S}^\perp$ gibt mit lauter echt positiven Einträgen (was, siehe oben, z.B. dann der Fall ist, wenn dem Problem ein 'Atomerhalt' innewohnt), so muss der Schnitt des Raumes $\vec{x} + \mathcal{S}$ mit dem positiven Rektanden *beschränkt* sein (s. Abb. 3, links). Da die Lösung in beiden Mengen zugleich liegen muss, folgt die Beschränktheit von Lösungen, sofern $\mathcal{S}^\perp \cap (\mathbb{R}_+)^I \neq \emptyset$, was den ersten Zugang zum Beweis der Beschränktheit von Lösungen geometrisch veranschaulicht.
Der zweite Zugang (Lyapunov-Technik) zeigt, dass die Existenz eines solchen Vektors bei *reversiblen* Systemen nicht erforderlich ist: Alle Niveaulinien von φ sind wegen $\varphi(\vec{x}) \geq |\vec{x}|_1$ beschränkt, und da $\varphi(\vec{c}(t))$ aus Monotoniegründen niemals größer als $\varphi(\vec{c}(0))$ sein kann, kann die Lösung einen beschränkten Bereich $(\vec{c}(0) + \mathcal{S}) \cap \{\vec{x} \,|\, \varphi(\vec{x}) \leq \varphi(\vec{c}(0))\}$ nicht verlassen (s. Abb. 3, rechts).

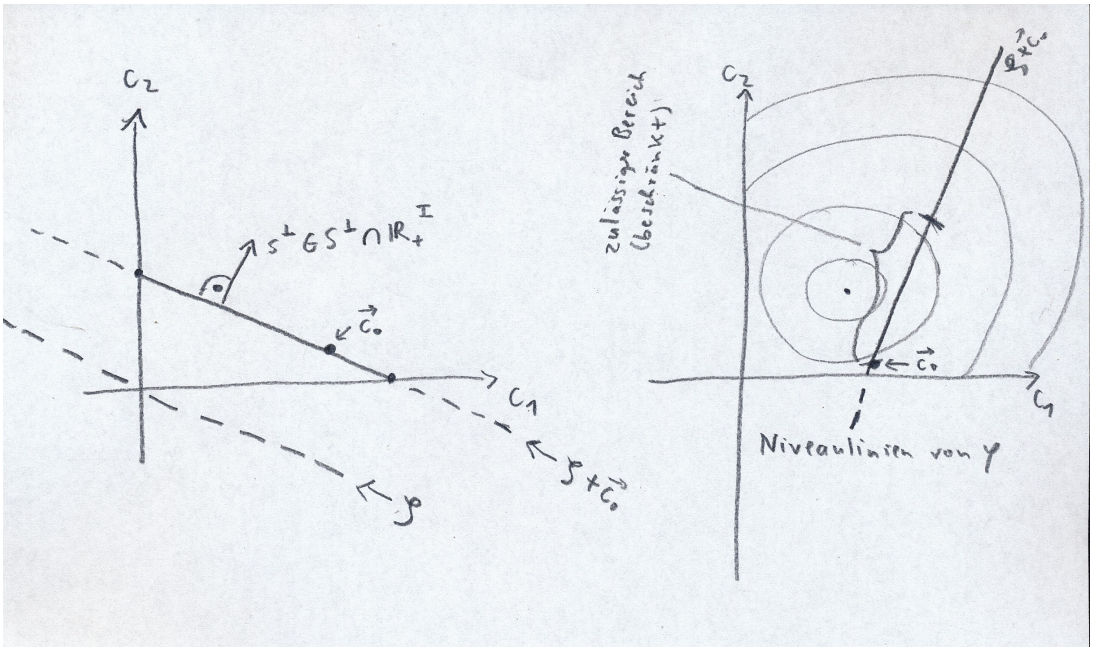


Abbildung 3: Links: Stöchiometrische Klasse im Falle, dass $\exists \vec{s}^\perp \in \mathcal{S}^\perp \cap (\mathbb{R}_+)^I \neq \emptyset$ (also z.B. im Falle eines dem reaktiven Systems innewohnenden 'Atomerhalts'). Rechts: Niveaulinien von φ bei 'beliebiger' Lage von $\vec{s}^\perp$. In beiden Fällen bekommen wir die Beschränktheit von Lösungen.

5.3 Reaktionsinvarianten

Zur Vereinfachung gehen wir in diesem Kapitel davon aus, dass die Spalten von S linear unabhängig sind; auf diese Anforderung kann jedoch auch verzichtet werden; siehe Anmerkung unten.
Wir zerlegen jeden Vektor $\vec{c} \in \mathbb{R}^I$ in eine direkte Summe, bestehend aus seiner Projektion $P_{\mathcal{S}}\vec{c}$ auf den Raum $\mathcal{S}$ sowie seiner Projektion $P_{\mathcal{S}^\perp}\vec{c}$ auf den Raum $\mathcal{S}^\perp$:

$$\vec{c} = P_{\mathcal{S}}\vec{c} \oplus P_{\mathcal{S}^\perp}\vec{c}$$

Da $\mathcal{S} = \text{Bild}(S)$ von den Spalten von S aufgespannt wird, kann man $P_{\mathcal{S}}$ als LK dieser Spalten mit Koeffizienten ξ_i schreiben; analog für $P_{\mathcal{S}^\perp}$:

$$P_{\mathcal{S}}\vec{c} = S\vec{\xi}, \quad P_{\mathcal{S}^\perp}\vec{c} = U\vec{\eta}, \quad \vec{\xi} \in \mathbb{R}^J, \vec{\eta} \in \mathbb{R}^{I-J}$$

Dabei ist U eine $(I-J) \times I$-Matrix, deren Spalten eine Basis des $I-J$-dimensionalen Raumes $\mathcal{S}^\perp$ bilden; es gelten also insbesondere die Matrixgleichungen

$$U^T S = 0, \quad S^T U = 0.$$

Wir haben also die Darstellung

$$\boxed{\vec{c} = S\vec{\xi} + U\vec{\eta}}, \qquad \vec{\xi} \in \mathbb{R}^J, \vec{\eta} \in \mathbb{R}^{I-J} \qquad (*)$$

Eine Umkehrformel für diese Transformation kann man auf zweierlei Arten finden. Entweder man kennt die Tatsache, dass sich eine Orthogonalprojektion auf einen Raum $\mathcal{S} = \text{Bild}(S)$ schreiben lässt als $P_{\mathcal{S}}\vec{c} = S(S^TS)^{-1}S^T\vec{c}$ (beachte, dass S^TS invertierbar ist aufgrund der Annahme, dass S maximalen Spaltenrang hat). Analog $P_{\mathcal{S}^\perp}\vec{c} = U(U^TU)^{-1}U^T\vec{c}$. Es folgt

$$\vec{c} = S\underbrace{(S^TS)^{-1}S^T\vec{c}}_{=\vec{\xi}} + U\underbrace{[U^TU]^{-1}U^T\vec{c}}_{=\vec{\eta}}$$

woraus man

$$\boxed{\vec{\xi} = (S^TS)^{-1}S^T\vec{c}}, \quad \boxed{\vec{\eta} = (U^TU)^{-1}U^T}$$

'ablesen' kann. Oder man wendet S^T auf die Gleichung (*) an, nutzt dann $S^TU = 0$ aus, und wendet dann die Matrix $(S^TS)^{-1}$ an, um die Formel für $\vec{\xi}$ zu bekommen; analog für $\vec{\eta}$.

Das ODE-System

$$\frac{d}{dt}\vec{c}(t) = S\vec{R}(\vec{c}(t))$$

(nicht notwendigerweise MWG, nicht notwendigerweise reversibel) kann durch Multiplikation einerseits mit $(S^TS)^{-1}S^T$, andererseits mit $[U^TU]^{-1}U^T$ auf

$$\begin{aligned}
\frac{d}{dt}\underbrace{[U^TU]^{-1}U^T\vec{c}}_{=\vec{\eta}(t)} &= [U^TU]^{-1}\underbrace{U^TS}_{=0}\vec{R}(\vec{c}(t)) \\
\frac{d}{dt}\underbrace{(S^TS)^{-1}S^T\vec{c}}_{=\vec{\xi}(t)} &= \underbrace{(S^TS)^{-1}S^TS}_{=\text{Id}}\vec{R}(\vec{c}(t))
\end{aligned}$$

äquivalent umgeformt werden. Wir erhalten somit das System

$$\begin{aligned}
\frac{d}{dt}\vec{\eta}(t) &= \vec{0} \\
\frac{d}{dt}\vec{\xi}(t) &= \vec{R}(\vec{c}(t))
\end{aligned}$$

Es ist also $\vec{\eta}(t) = \text{const} = \vec{\eta}(0) = (U^TU)^{-1}U^T\vec{c}(0) =: \vec{\eta}_0$. Wir erhalten somit das i.a. kleinere, nur noch aus J (statt I) ODEs bestehende System

$$\boxed{\tfrac{d}{dt}\vec{\xi}(t) = \vec{R}(S\vec{\xi}(t) + U\vec{\eta}_0)}.$$

Die Komponenten η_i heißen *Reaktionsinvarianten (reaction invariants)*, die ξ_i *extents of reaction*. Geometrisch bezeichnen die ξ_i Koordinaten in Richtung der stöchiometrischen Klasse, und die η_i bezeichnen Koordinaten senkrecht zur stöchiometrischen Klasse. Dass die η_i konstant sind, passt zu der vorher bereits hergeleiteten Tatsache,

dass die Lösung die stöchiometrische Klasse des Anfangswertes niemals verlässt.

Bemerkung zur Abschwächung der Voraussetzung (zur Übung): Falls die Spalten von S linear abhängig sind, dann führt man eine Matrix S^* ein, die aus einem linear unabhängigen Teilsystem von Spalten von S besteht, so dass $\text{Bild}(S^*) = \text{Bild}(S)$. Dann gibt es eine Matrix A, so dass $S = S^*A$. Die Transformation lässt auf $\vec{\xi}$-$\vec{\eta}$-Koordinaten lässt sich dann ebenfalls durchführen. (Welche Bedeutung haben die Spalten von A? Können Sie die Gleichung $S = S^*A$ nach A auflösen?)
Nutzen der Umformung dieses Kapitels ist, dass das numerische Lösen des kleineren 'reduzierten' Problems schneller geht. Dies ist insbesondere interessant, wenn es um das (zeitaufwändige) Lösen von *PDE*-Systemen geht (Übertragung von Kap. 5.3 auf PDEs: später).

5.4 Reaktionen im Gleichgewicht

Wir betrachten das ODE-Problem zunächst ohne Vorgabe von Anfangswerten. Wir suchen nach *Gleichgewichtslösungen*, also nach Lösungen mit $\vec{0} \stackrel{!}{=} \vec{c}'(t) = S\vec{R}(\vec{c}(t))$, somit nach Vektoren $\vec{c} \in (\mathbb{R}_0^+)^I$ mit

$$S\vec{R}(\vec{c}) = \vec{0},$$

also $\vec{R}(\vec{c}) \in \text{Kern}(S)$. Wir setzen nun voraus, dass die Spalten von S linear unabhängig sind. Dann können wir per Multiplikation[17] mit $(S^TS)^{-1}S^T$ die notwendige Bedingung

$$\vec{R}(\vec{c}) = \vec{0}$$

herleiten. Setzen wir zusätzlich voraus, dass das System *reversibel* ist, so erhalten wir

$$\vec{c} \text{ ist GG-Lösung} \quad \Longleftrightarrow \quad \forall\, j = 1, ..., j : R_j^v(\vec{c}) = R_j^r(\vec{c}),$$

d.h. jede einzelne Vorwärts-Rückwärts-Reaktion muss dann im Gleichgewicht sein. Im Falle des Massenwirkungsgesetztes lautet die GG-Bedingung in logarithmierter Form

$$\ln \vec{k}^v + (S^v)^T \ln \vec{c} = \ln \vec{k}^r + (S^r)^T \ln \vec{c},$$

was zu

$$\boxed{\ln \vec{K} + S^T \ln \vec{c} = 0}$$

zusammengefasst werden kann. Die Lösung ist natürlich i.a. nicht eindeutig, da es J Gleichungen für I Unbekannte gibt.
Nehmen wir nun aber noch einen konkreten Anfangswert $\vec{c}_0 \in \mathbb{R}_+^I$ an, so legt dieser die stöchiometrische Klasse fest, in der wir nach einer GG-Lösung suchen, was $I - J$ zusätzlichen Bedingungen $\eta_i = \eta_{i,0}$ entspricht. Obwohl die GG-Bedingungen nichtlinear sind, kann man unter den gemachten Annahmen zeigen, dass nun die GG-Lösung in

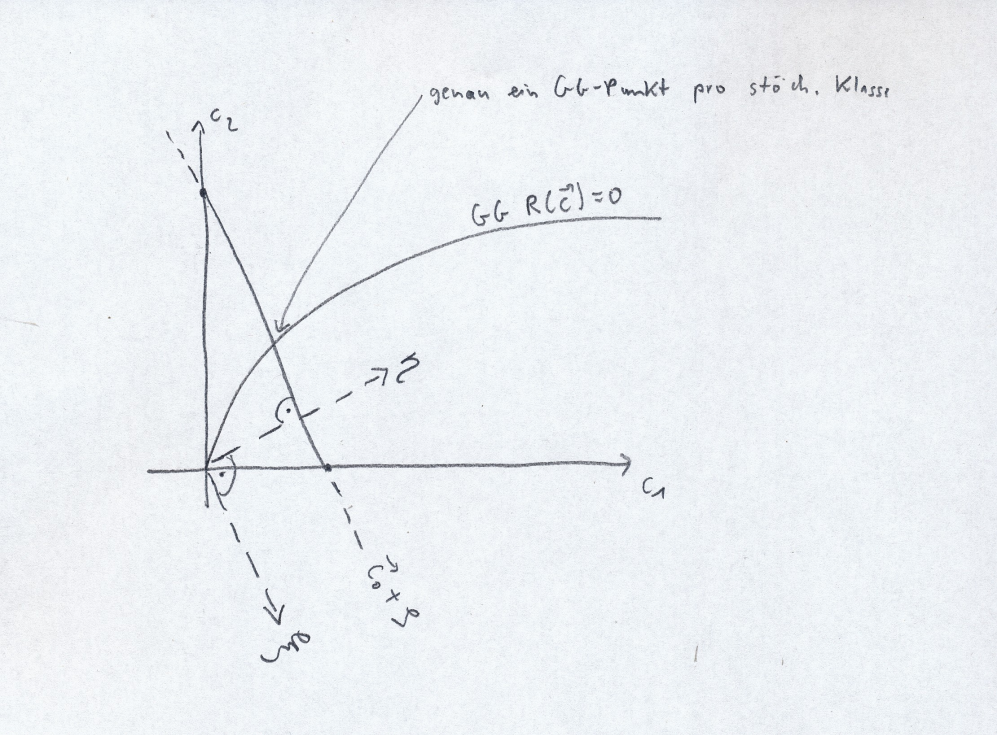


Abbildung 4: Für das reversible System $X_1 \longleftrightarrow 2X_2$, also $S = (-1, 2)^T$, und GG-Bedingung $k_1^v c_1 = k_1^r c_2^2$: Lage der stöchiometrischen Klasse(n) $(\vec{c}_0+\mathcal{S}) \cap \mathbb{R}^I_+$ und Lage der GG-Punkte; in jeder Klasse liegt genau ein GG-Punkt.

einer stöchiometrischen Klasse existiert und eindeutig bestimmt ist.

Satz. Für reversible Massenwirkungssysteme, bei denen die stöchiometrische Matrix maximalen Spaltenrang hat, gibt es in jeder nichtleeren stöchiometrischen Klasse $(\vec{c}_0+\mathcal{S}) \cap \mathbb{R}^I_+$ genau einen Gleichgewichtspunkt.

Beweis. Wir betrachten zunächst das Hilfsproblem

"Minimiere das Funktional φ aus Kap. 5.2 unter der Nebenbedingung $\vec{\eta} = (U^T U)^{-1} U^T \vec{c} \stackrel{!}{=} \vec{\eta}_0$"

(Es ist naheliegend zu vermuten, dass dieses Problem äquivalent ist zu unserem GG-Problem.) Mit dem Lagrange-Formalismus erhalten wir das äquivalente[18] System

$$\nabla\varphi(\vec{c}) = (U^T U)^{-1} U^T]^T \vec{\lambda}, \quad (U^T U)^{-1} U^T \vec{c} = \vec{\eta}_0$$

was äquivalent ist zu

$$\mu + \ln \vec{c} = U\,(U^T U)^{-1} \vec{\lambda}, \quad (U^T U)^{-1} U^T \vec{c} = \vec{\eta}_0.$$

[17] alternative Argumentation: Dann ist $\vec{R}(\vec{c}) \in \mathrm{Kern}(S) = \{\vec{0}\}$

[18] Hier geht ein, dass das Problem konvex ist.

Die erste der beiden Gleichungen wird äquivalent umgeformt durch Multiplikation einerseits mit U^T, andererseits mit S^T. Wir erhalten die drei Gleichungen

$$\underbrace{S^T(\mu+\ln\vec{c})=0}_{=\text{GG-Bedingung}}, \quad U^T(\mu+\ln\vec{c})=\vec{\lambda}, \quad \underbrace{(U^TU)^{-1}U^T\,\vec{c}=\vec{\eta}_0}_{\Leftrightarrow\, \vec{c}\in\vec{c}_0+\mathcal{S}}.$$

Die mittlere können wir fallen lassen, da sie lediglich das (uninteressante) $\vec{\lambda}$ definiert. Wir sehen: Das obige restringierte Minimierungsproblem ist äquivalent zur Suche nach GG-Punkten in der stöchiometrischen Klasse.
Für das restringierte Minimierungsproblem wiederum kann man Existenz und Eindeutigkeit der Lösung relativ leicht zeigen: Die "zulässige Menge" (also die Menge, die durch die Nebenbedingung beschrieben wird) ist konvex. Die Zielfunktion φ ist strikt konvex, denn die Hesse-Matrix

$$H\varphi(\vec{c})=\operatorname{diag}(\frac{1}{c_1},...,\frac{1}{c_I})$$

ist positiv definit. Somit hat das restringierte Minimierungsproblem höchstens eine Lösung. Um die Existenz einer Lösung zu zeigen, ist es hinreichend, dass es eine nichtleere, kompakte Niveaumenge

$$M_l := \{\vec{c}\in(\mathbb{R}_0^+)^I \mid \varphi(\vec{c})\leq l\}$$

gibt. Dass es eine solche gibt, folgt aus der Abschätzung $\varphi(\vec{c}) \geq |\vec{c}|$ aus Kap. 5.2, da diese Abschätzung $M_l \subseteq K_l(\vec{0})$ beinhaltet, somit die Beschränktheit liefert, und da $\vec{c}$ auf dem nichtnegativen abgeschlossenen Rektanden stetig ist, was die Abgeschlossenheit von M_l liefert. □

6 Feinberg'sche Netzwerktheorie

6.1 Einführung

Das Hauptziel der Feinberg'schen Netzwerktheorie ist es, Aussagen über **Existenz und Eindeutigkeit von Gleichgewichtszuständen** des Batch-Problems zu liefern. Wir haben bereits in Kap. 5.4 gesehen, dass unter Annahme der Reversibilität Existenz und Eindeutigkeit eines GG-Zustandes in jeder nichtleeren stöchiometrischen Klasse gezeigt werden kann. Feinberg gelang es, die Voraussetzungen abzuschwächen; eine wesentliche Rolle spielt dabei der Begriff der *schwachen Reversibilität*. Interessanterweise wird in der Feinberg'schen Netzwerktheorie zu jedem reaktiven System ein mathematischer Graf aufgestellt, und Kriterien für Existenz/Eindeutigkeit von GG-Zuständen werden mit Hilfe grafentheoretischer Begriffe formuliert.

Aufstellen des Grafen für ein reaktives System anhand eines Beispiels:
Die chemischen Gleichungen seien

$$\begin{array}{rcll} X_1 + X_2 & \rightleftarrows & X_3 & \\ X_3 & \to & X_4 + X_5 & \\ X_6 & \rightleftarrows & X_4 + X_5 & \\ 2X_2 & \to & X_2 + X_7 & \\ X_8 & \to & 2X_2 & \\ X_8 & \rightleftarrows & X_2 + X_7 & \qquad I = 8 \text{ Spezies} \end{array}$$

Das, was auf einer Seite eines chemischen Reaktionspfeiles steht, bezeichnen wir als *Komplex*, z.B. $X_2 + X_7$. Ein Komplex hat *keine* Konzentration. Komplexe werden nur als mathematisches Hilfsmittel eingeführt; sie entsprechen keiner physikalischen Substanz.

Darstellung als gerichteter Graf:

$$\begin{array}{rcl} \text{Komplexe} & \widehat{=} & \text{Knoten} \\ \text{Reaktionen} & \widehat{=} & \text{Kanten} \end{array}$$

im Beispiel:

$$\begin{array}{rcl} X_1 + X_2 & \rightleftarrows & X_3 \to X_4 + X_5 \rightleftarrows X_6 \\ 2X_2 & \to & X_2 + X_7 \\ \nwarrow & & \swarrow\nearrow \\ & X_8 & \end{array}$$

Auch wenn in den Reaktionen ein Komplex mehrfach vorkommt, so wird er im Grafen mit *einem* Knoten repräsentiert. Hier: $n = 7$ Knoten/Komplexe und $J = 9$ Kanten/Reaktionen (dabei "$\rightleftarrows$" als zwei Kanten gezählt).

Wir können den Grafen als „Gebilde“ im $\mathbb{R}^I$ ($I = 8$) auffassen:

Knoten/Komplexe können als Elemente des $(\mathbb{R}_0^+)^I$ aufgefasst werden; z.B. $X_1 + X_2 \widehat{=} \vec{e}_1 + \vec{e}_2 = (1, 1, 0, 0, 0, 0, 0, 0)^T$, wobei $\vec{e}_1, \ldots, \vec{e}_I$ die Standardbasis-Vektoren des $\mathbb{R}^I$ sind.
Also: Knotenmenge/Komplexmenge $\mathcal{C} \subseteq (\mathbb{R}_0^+)^I$
Beachte: Die Komponenten des Vektors sind gerade die *stöchiometrischen Koeffizienten* s_{ij}^e, s_{ij}^p aus den vorherigen Kapiteln!

Die Menge der Kanten/Reaktionen kann einerseits als *Relation* $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ auf der Menge $\mathcal{C}$ aufgefasst werden. Daneben haben die Kanten/Reaktionen auch eine Interpretation als Richtungsvektoren=Differenz der Endpunkte, somit als Elemente des $\mathbb{R}^I$ aufgefasst werden: z.B. die Reaktion „$2X_2 \to X_2 + X_7$“$\widehat{=} \vec{e}_2 + \vec{e}_7 - 2\vec{e}_2 = \vec{e}_7 - \vec{e}_2 = (0, -1, 0, 0, 0, 0, 1, 0)^T \in \mathbb{R}^I$

Beachte: Die Komponenten des Vektors sind gerade die stöchiometrischen Koeffizienten $s_{ij} = s_{ij}^p - s_{ij}^e$ aus den vorherigen Kapiteln. Da jeder Knoten/Komplex sowohl als Edukt wie auch als Produkt auftreten kann, ist die Notation $(\vec{s}_i, \vec{s}_j) \in \mathcal{R}$ für eine Reaktion, die von Knoten/Komplex $\vec{s}_i$ zum Knoten/Komplex $\vec{s}_j$ abläuft, sinnvoller; die zugehörige Ratenfunktion wird mit $R_{(\vec{s}_i,\vec{s}_j)}$ oder $R_{\vec{s}_i \to \vec{s}_j}$ bezeichnet.

Unser Batch-Modell, das in der alten Notation

$$\vec{c}\,'(t) = (S^p - S^e)\,\vec{R}(\vec{c}(t)) = \sum_{j=1}^{J} (\vec{s}_j^{\,p} - \vec{s}_j^{\,e})\,R_j(\vec{c}(t))$$

lautet, wird in der Feinberg'schen Notation zu

$$\vec{c}\,'(t) = \sum_{(\vec{s}_i,\vec{s}_j)\in\mathcal{R}} (\vec{s}_j - \vec{s}_i)\,R_{\vec{s}_i \to \vec{s}_j}(\vec{c}(t))$$

oder

$$= \sum_{(\vec{s},\vec{s}\,')\in\mathcal{R}} (\vec{s}\,' - \vec{s})\,R_{\vec{s}\to\vec{s}\,'}(\vec{c}(t))$$

Def. (Chemisches Reaktionsnetzwerk, Reaktives System) Ein chemisches Reaktionsnetzwerk (RNW) ist ein Tripel $(\mathcal{M}_\mathcal{S}, \mathcal{C}, \mathcal{R})$, wobei
$\mathcal{M}_\mathcal{S} = \{1, 2 \ldots, I\}, I \in \mathbb{N}$, die *Menge der Spezies* ist,
$\mathcal{C} \subset \bar{\mathbb{R}}_+^N$ mit $|\mathcal{C}| =: n \in \mathbb{N}$ die (endliche) *Menge der Komplexe* ist,
und wobei $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$, die Menge der Reaktionen, eine *Relation* auf $\mathcal{C}$ ist mit den Eigenschaften $(\vec{s}, \vec{s}) \notin \mathcal{R}\ \forall\ \vec{s} \in \mathcal{C}$; (i.a. ist die Relation nicht symmetrisch, d.h. der Graf ist *gerichtet*).
Ist darüber hinaus jeder Rate/Kante $(\vec{s}, \vec{s}\,') \in \mathcal{R}$ eine stetig differenzierbare Ratenfunktion $R_{\vec{s}\to\vec{s}\,'} : (\mathbb{R}_0^+)^I \longrightarrow \mathbb{R}_0^+$, die zudem die Minimalanforderung aus Kap. 4.2, $R(\vec{c}) > 0 \quad \Leftrightarrow \text{supp}\ \vec{s}^{\,e} \subseteq \text{supp}\ \vec{c}$, erfüllt, so spricht Feinberg von einem *reaktiven System*.

6.2 Schwache Reversibilität, Zusammenhangskomponenten, Rang, Defekt

Die Definition der Reversibilität lautet in der Feinberg'schen Notation:

Ein RNW heißt *reversibel*, falls

$$\forall \vec{s}, \vec{s}\,' \in \mathcal{C} :\ (\vec{s}, \vec{s}\,') \in \mathcal{R}\ \Rightarrow (\vec{s}, \vec{s}\,') \in \mathcal{R}.$$

(dies entspricht der Symmetrie der Relation $\mathcal{R}$.)

Def. (schwach reversibel). Ein RNW heißt *schwach reversibel*, falls es für alle $\vec{s}, \vec{s}' \in \mathcal{C}$, für die es einen gerichteten Pfad im Grafen von $\vec{s}$ nach $\vec{s}'$ gibt, es einen gerichteten Pfad von $\vec{s}'$ nach $\vec{s}$ gibt:

$$\forall \vec{s}, \vec{s}' \in \mathcal{C}: \quad \big(\exists : m \in \mathbb{N}_0, \vec{s}_1, ..., \vec{s}_m \in \mathcal{C}: \ (\vec{s}, \vec{s}_1), (\vec{s}_1, \vec{s}_2), ..., (\vec{s}_{m-1}, \vec{s}_m), (\vec{s}_m, \vec{s}') \in \mathcal{R}$$
$$\Longrightarrow \exists : m' \in \mathbb{N}_0, \vec{s}_1{}', ..., \vec{s}'_{m'} \in \mathcal{C}: \ (\vec{s}', \vec{s}_1{}'), (\vec{s}_1{}', \vec{s}_2{}'), ..., (\vec{s}'_{m'-1}, \vec{s}'_{m'}), (\vec{s}'_{m'}, \vec{s}) \in \mathcal{R} \big)$$

Aus der Reversibilität folgt offensichtlich die schwache Reversibilität.

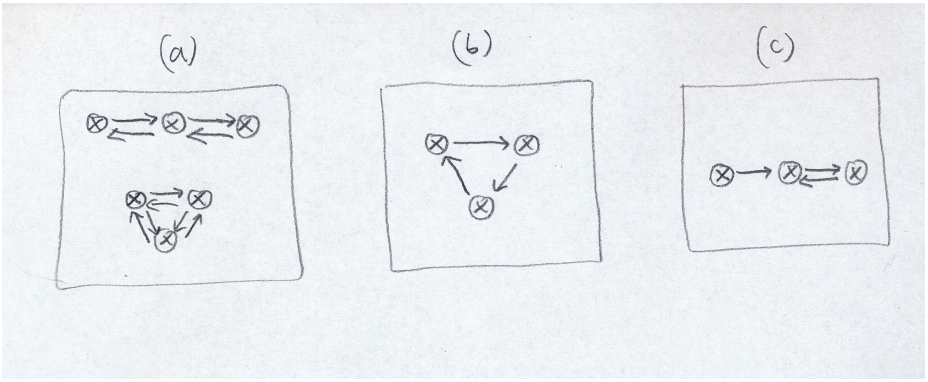


Abbildung 5: Netzwerk (a) ist reversibel, Netzwerk (b) ist schwach reversibel, aber nicht reversibel, Netzwerk (c) ist nicht schwach reversibel, aber zusammenhängend.

Zu dem gerichteten Grafen kann man offensichtlich durch 'Weglassen der Pfeilspitzen' einen *ungerichteten Grafen* $(\mathcal{M}_\mathcal{S}, \mathcal{C}, \tilde{\mathcal{R}})$ machen, der also die gleiche Knotenmenge hat sowie die Kantenmenge

$$\tilde{\mathcal{R}} \subset \mathcal{C} \times \mathcal{C}, \quad (\vec{s}_1, \vec{s}_2) \in \tilde{\mathcal{R}} :\Longleftrightarrow (\vec{s}_1, \vec{s}_2) \in \mathcal{R} \ \vee \ (\vec{s}_2, \vec{s}_1) \in \mathcal{R};$$

die Relation $\tilde{\mathcal{R}}$ ist also symmetrisch.

Def. (Zusammenhangskomponente) Wir definieren eine weitere Relation, $\sim$, auf $\mathcal{C}$:

$$\begin{aligned} \vec{s} \sim \vec{s}' \quad &:\Longleftrightarrow \quad \text{Es gibt im ungerichteten Grafen einen Pfad}^{19} \text{ von } \vec{s} \text{ nach } \vec{s}': \\ &\Longleftrightarrow \quad \exists : m \in \mathbb{N}_0, \vec{s}_1, ..., \vec{s}_m \in \mathcal{C}: \ (\vec{s}, \vec{s}_1), (\vec{s}_1, \vec{s}_2), ..., (\vec{s}_{m-1}, \vec{s}_m), (\vec{s}_m, \vec{s}') \in \tilde{\mathcal{R}} \end{aligned}$$

'$\sim$' ist offensichtlich eine Äquivalenzrelation. Die Äquivalenzklassen von $\mathcal{C}$ unter '$\sim$' heißen *Zusammenhangskomponenten* (ZHKs, engl: linkage classes) des RNW. Wir bezeichnen die Anzahl von ZHKs mit l.
Beispiel: Das RNW

$$s_1 \to s_2 \leftarrow s_3$$
$$s_4 \leftrightarrows s_5$$

[19] Unter 'Pfad' wollen wir hier und im Folgenden verstehen 'Pfad der Länge ≥ 0', d.h. jeder Knoten soll bzgl. '$\sim$' zu sich selbst in Relation stehen.

hat $l=2$ ZHKs.

Def. (Rang). Der *Rang* $s \in \mathbb{N}_0$ eines RNW ist die Dimension des Vektorraums, der von den Reaktionen aufgespannt wird, also des stöchiometrischen Raums:

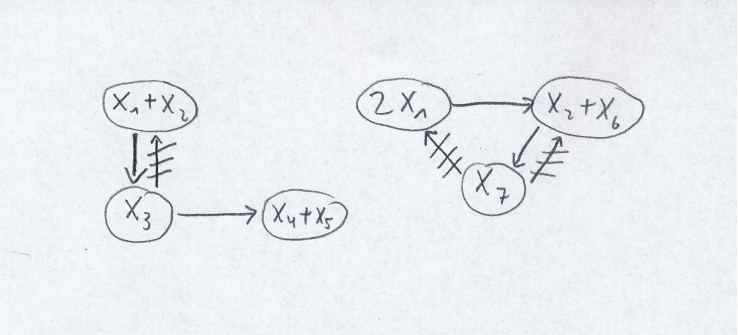
$$s := \dim(\mathrm{span}\{\vec{s}-\vec{s}' \,|\, (\vec{s},\vec{s}') \in \mathcal{R}\}) = \dim(\mathrm{Bild}(S^p - S^e)) = \dim(\mathrm{Bild}(S)) = \dim(\mathcal{S})$$

Da $S \in \mathbb{R}^{I\times J}$ ist also $0 \leq s \leq \min\{I, J\}$.

Offensichtlich ist der Rang eines RNW *invariant* gegenüber

- Umkehrung der Richtung einer Kante ($\widehat{=}$ Multiplikation eines aufspannenden Vektors mit (-1)),
- Einfügen/Weglassen einer Kante, sofern sich dabei die ZHKs nicht verändern ($\widehat{=}$ Wegstreichen/Hinzufügen von linear abhängigen Vektoren im EZS von $\mathcal{S}$)

Beispiel. Zum RNW



gehört die stöchiometrische Matrix

$$S = \begin{pmatrix} -1 & 1 & 0 & -2 & 0 & 0 & 2 \\ -1 & 1 & 0 & 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 \end{pmatrix}$$

Die ersten beiden Spalten sind linear abhängig, ebenso die letzten vier. Es können hier drei Spalten/Reaktionen gestrichen werden, ohne dass sich die ZHKs/der Rang des RNWs ändert. Streichen wir die zweite sowie die beiden letzten Spalten aus S, so entspricht dies den gestrichenen Reaktionen im Grafen, die in der Skizze zu sehen sind. Die verbleibenden vier Spalten/Reaktionen sind linear unabhängig; es ist hier also $s=4$.

Der Rang eines RNW hängt nur insofern von den Kanten ab, als diese die ZHKs definieren; man kann Grafen immer 'maximal ausdünnen' ohne den Rang zu verändern.

In einem 'maximal ausgedünnten' Grafen ist in jeder ZHK die Anzahl der Kanten gleich der Anzahl der Knoten $|\mathcal{C}_i|$ minus eins. Eine obere Schranke für den Rang ist somit:

$$s \leq \text{Anzahl Kanten im maximal ausgedünnten Grafen} = \sum_{i=1}^{l}(|\mathcal{C}_i|-1) = \underbrace{|\mathcal{C}|}_{=n} - l,$$

also[20]

$$s \leq n - l.$$

Def. (Defekt). Die Zahl

$$\delta := n - l - s \ \ (\geq 0)$$

heißt *Defekt* eines RNW.

Im obigen Beispiel: $n=6$ Komplexe, $l=2$ ZHKs, $s=4 \ \Rightarrow \ \delta=6-2-4=0$

6.3 Das Defekt-Null-Theorem

Der zentrale Satz der Feinberg'schen Netzwerktheorie ist das Defekt-Null-Theorem:

Satz (Defekt-Null-Theorem). Für jedes RNW [21] mit Defekt $\delta=0$ gilt:

(a) Wenn das RNW *nicht schwach reversibel* ist, dann hat das System keine positive stationäre Lösung.

(b) Ist das RNW *schwach reversibel*, und wird Massenwirkungskinetik angenommen, dann gibt es in jeder stöchiometrischen Klasse, die mit $\mathbb{R}^I_+$ einen nichtleeren Schnitt hat, *genau eine* positive stationäre Lösung, und diese ist asymptotisch stabil ('bezüglich der stöchiometrischen Klasse'[22]).

Beweisskizze[23] Im Satz geht es um Existenz/Eindeutigkeit von Gleichgewichten, genauer gesagt, um *Speziesgleichgewichteé*, also um Vektoren $\vec{c}$ mit

$$0 \stackrel{!}{=} \frac{d\vec{c}}{dt} = \sum_{(\vec{s},\vec{s}')\in\mathcal{R}} (\vec{s}'-\vec{s})\, R_{\vec{s}\to\vec{s}'}(\vec{c}).$$

[20] Auch ohne die obige Argumentation über das Ausdünnen des Grafen sollte klar sein: Der Raum, der von Vektoren augespannt wird, die $|\mathcal{C}_i|$ viele Punkte miteinander verbinden, kann höchstens $|\mathcal{C}_i|-1$-dimensional sein.

[21] nicht notwendigerweise mit Massenwirkungskinetik, sondern nur die 'Minimalanforderungen' an Raten sollen erfüllt sein

[22] Was Stabilität der GG-Lösung *bezüglich der stöchiometrischen Klasse* genau bedeutet, wird in einer Übungsaufgabe verdeutlicht

[23] Mehr Details aus dem Beweis: Siehe mein Vorlesungsskript 'Reaktive Netzwerke'.

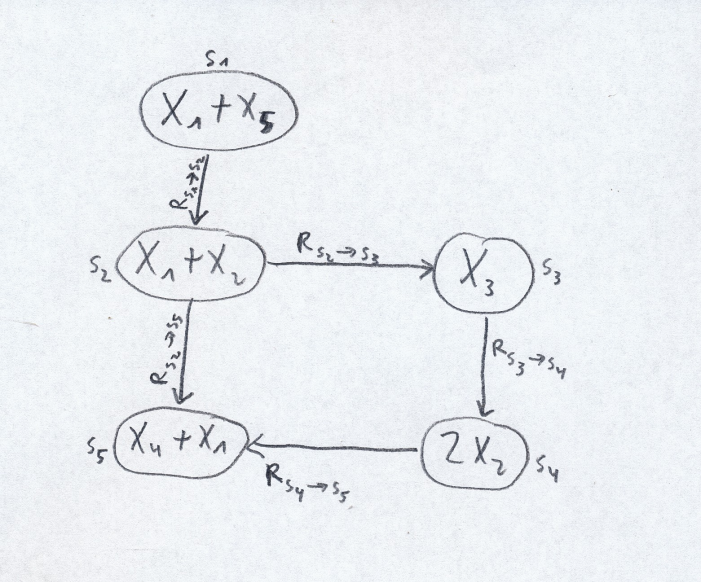


Abbildung 6:

Daneben gibt es noch den Begriff des *Komplex-* oder *Knoten-Gleichgewichts*. Das sind Vektoren $\vec{c}$, für die gilt: Für jeden Knoten $\vec{s} \in \mathcal{C}$ gilt, dass sich die Raten, die zum Knoten hin- und wegführen, aufheben, also[24]

$$\forall \vec{s} \in \mathcal{C} : \; r_{\vec{s}}(\vec{c}) := \underbrace{\sum_{\substack{\vec{s}', \\ (\vec{s}', \vec{s}) \in \mathcal{R}}} R_{\vec{s}' \to \vec{s}}(\vec{c})}_{\text{hin zu } \vec{s}} - \underbrace{\sum_{\substack{\vec{s}', \\ (\vec{s}, \vec{s}') \in \mathcal{R}}} R_{\vec{s} \to \vec{s}'}(\vec{c})}_{\text{weg von } \vec{s}} = 0$$

Anschaulich leuchtet leicht ein (und es lässt sich auch leicht zeigen), da sich jedes $\frac{d}{dt} c_i(t)$ als Linearkombination von 'Knoten-Raten' (s.o.) schreiben lässt, dass also jedes Knotengleichgewicht ein Speziesgleichgewicht ist. Die Umkehrung ist jedoch i.a. falsch! Dazu ein Beispiel (siehe Fig. 6): Es ist dort $\frac{d}{dt} c_2 = r_{\vec{s}_2}(\vec{c}) + 2 r_{\vec{s}_4}(\vec{c})$ (da X_2 in diesen beiden Komplexen vorkommt), jedoch kann $r_{\vec{s}_2}(\vec{c}) + 2 r_{\vec{s}_4}(\vec{c})$ null werden ohne dass $r_{\vec{s}_2}(\vec{c})$ und $r_{\vec{s}_4}(\vec{c})$ beide null sind (Komplex-Raten können negativ sein). Feinberg zeigt jedoch: Falls $\delta = 0$, dann gilt auch die Rückrichtung, d.h. $\vec{c}$ ist genau dann Spezies-GG, wenn es Komplex-GG ist. (Als Hilfsmittel für diese Zwischenbehauptung verwendet er Lineare Algebra; er konstruiert eine Matrix A, die von einem Parameter $\vec{\rho} \in \mathbb{R}^J$ abhängt, derart, dass $\vec{c}$ genau dann Knoten-GG ist, wenn $(1, ..., 1)^T \in \mathrm{Kern}(A_{\vec{\rho}})$, und $\vec{c}$ genau dann Spezies-GG ist, wenn $(1, ..., 1)^T \in \mathrm{Kern}(S A_{\vec{\alpha}})$, wobei $\vec{\rho} := \vec{R}(\vec{c})$; siehe Skript 'Reaktive Netzwerke' S. 36-38. [25])

[24] Beachte, dass Knoten/Komplexe keine phys. Konzentrationen haben; die Knoten-Raten sind also reine theoretische Hilfsmittel, deren Betrachtung aber aufgrund des mathematischen Konzepts 'Graf' recht naheliegend ist: Die Kanten werden als 'Rohrleitungen' betrachtet, die Knoten als 'Depots'.

[25] Die lineare Abbildung $A_{\vec{\rho}} : \mathbb{R}^n \to \mathbb{R}^n$ lautet: $\vec{x} \longmapsto \sum_{\substack{\vec{s}', \\ (\vec{s}, \vec{s}') \in \mathcal{R}}} \rho_{\vec{s}'} x_{\vec{s}} (\vec{e}_{\vec{s}'} - \vec{e}_{\vec{s}})$, dabei Komponentenschreibweise: $\vec{x} = (x_{\vec{s}})_{\vec{s} \in \mathcal{C}}$, $\vec{\rho} = (\rho_{\vec{s}})_{\vec{s} \in \mathcal{C}}$

Anschließend reicht es also, Existenz und Eindeutigkeit von *Komplex-Gleichgewichten* zu untersuchen. Der Beweis ist sehr aufwändig (Skript 'Reaktive Netzwerke' S. 38-44); der Beweis der Existenz und Eindeutigkeit dieses Gleichgewichts für (b) erfolgt in drei Schritten:
(1.) Zeige, dass (nicht notwendigerweise in jeder Klasse) in $\mathbb{R}^I_+$ ein Komplex-GG $\vec{c}_*$ gibt (Dieses ist, nach Vorüberlegung, auch Spezies-GG).
(2.) Zeige die Gleichheit der Mengen $\{\vec{c} \in \mathbb{R}^I_+ \mid \vec{c} \text{ ist GG}\}$ und $\{\vec{c} \in \mathbb{R}^I_+ \mid \ln \vec{c} - \ln \vec{c}_* \in \mathcal{S}^\perp\} =: E_{\vec{c}_*}$
(3.) Zeige: $E_{\vec{c}_*} \cap (\vec{c}_0 + \mathcal{S})$ hat genau einen Punkt.
Bemerkung zu Punkt (2.): Im Fall *echter Reversibilität* ist diese Mengengleichheit sofort klar: Sei $\vec{c}_*$ ein GG-Punkt, also $\ln \vec{K} + S^T \ln \vec{c}_* = \vec{0}$. Es ist dann $\vec{c}$ genau dann ein GG-Punkt, wenn $\ln \vec{K} + S^T \ln \vec{c} = \vec{0}$, also genau dann, wenn $S^T(\ln \vec{c} - \ln \vec{c}_*) = \vec{0}$, was als $\ln \vec{c} - \ln \vec{c}_* \in \text{Kern}(S^T) = \mathcal{S}^\perp$ geschrieben werden kann. (Im Fall von nur schwacher Reversibilität muss anders argumentiert werden, da es ein solches $\vec{K}$ dann i.a. gar nicht gibt.)
Bemerkung zu Punkt (3.): Die Existenz und Eindeutigkeit eines Elementes von $E_* \cap (\vec{c}_0 + \mathcal{S})$ wird, wie schon im Fall der echten Reversibilität, mit Hilfe eines Funktionals φ gezeigt, indem man zeigt, dass sich das gesuchte Element als Lösung eines Minimierungsproblems für φ unter der Nebenbedingung $\vec{c} \in \vec{c}_0 + \mathcal{S}$ ist. In Ermangelung eines "$\vec{K}$" gibt es jedoch auch kein "$\vec{\mu}$", welches in Kap. 5.4 (siehe auch Kap. 5.2) zur Definition des dortigen φ verwendet wurde; jedoch kann man hier stattdessen das Funktional $\varphi(\vec{c}) := \sum_{i=1}^I (-\ln c_i^* - 1 + \ln c_i)\, c_i$ nehmen. □

6.4 Weitere grafentheoretische Begriffe und das Defekt-Eins-Theorem

Wir definieren eine neue Relation auf dem gerichteten Grafen:

$\vec{s}_1 \equiv \vec{s}_2 \iff$ Es gibt einen gerichteten Pfad von $\vec{s}_1$ nach $\vec{s}_2$ und einen von $\vec{s}_2$ nach $\vec{s}_1$

Trivial ist:
(1) Wenn $\vec{s}_1 \equiv \vec{s}_2$, dann sind $\vec{s}_1$ und $\vec{s}_2$ in der gleichen ZHK.
(2) '$\equiv$' ist Äquivalenzrelation.

Die bezüglich '$\equiv$' gebildeten Äquivalenzklassen heißen *starke Zusammenhangskomponenten* (engl: strong linkage classes). Wegen (1) ist jede ZHK die Vereinigung von ein oder mehreren ganzen, disjunkten starken ZHKs.
eine starke ZHK heißt *terminal*, falls kein Pfad aus ihr herausführt. Im Beispiel der Skizze mit $l = 2$ gibt es 4 starke ZHKs, von denen $t = 3$ terminal sind. Man kann leicht allgemein zeigen, dass jede ZHK mindestens eine starke terminale ZHK enthalten muss.

Da jede ZHK als eigenständiges RNW aufgefasst werden kann, kann man jeder ZHK $Z_1, ..., Z_l$ einen Defekt δ_k, $k = 1, ..., l$, zuordnen:

$$\delta_j := n_j - s_j - 1$$

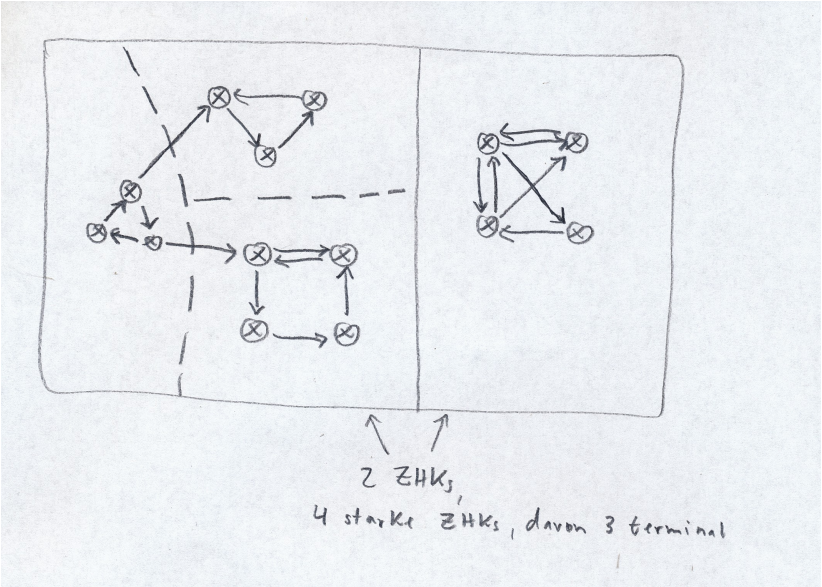


Abbildung 7: RNW mit 2 ZHKs, 4 starken ZHKs, davon 3 terminal.

Dabei ist n_j die Anzahl der Knoten der ZHK; s_j ist die Dimension des Raumes, der von Reaktionen aufgespannt wird, die von Knoten der ZHK zu Knoten derselben ZHK gehen, und die 1 ist die Anzahl der ZHKs, aus denen eine ZHK besteht. Addition liefert

$$\sum_{k=j}^{l} \delta_j = n - \sum_{j=1}^{l} s_j - l \leq n - s - l = \delta, \quad \text{sowie } \delta_j \leq \delta \forall\, j.$$

Nun ein weiteres wichtiges Theorem. Anders als der Name vermuten lässt, beschäftigt es sich nicht mit Netzwerken, die $\delta = 1$ haben:

Satz (Defekt-Eins-Theorem). (Feinberg, 1987). Wir betrachten ein RNW mit Massenwirkungskinetik und $\delta \in \mathbb{N}_0$ beliebig. Es seien $\mathcal{C}_1, ..., \mathcal{C}_l$ die ZHKs mit den Defekten $\delta_1, ..., \delta_l \in \mathbb{N}_0$. Es gelte

(V1) $\delta_j \leq 1 \;\forall\, j = 1, ..., l$

(V2) $\sum_{j=1}^{l} \delta_j = \delta$

(V3) Jede ZHK enthält genau eine starke terminale ZHK.

Dann gilt:

(a) Jede stöchiometrische Klasse enthält *höchstens einen* positiven GG-Zustand.

(b) Falls das System einen positiven GG-Zustand hat (in irgendeiner stöchiometrischen Klasse), dann enthält *jede* stöchiometrische Klasse $\vec{c}_0 + \mathcal{S}$ (die mit $\mathbb{R}^I_+$ nichtleeren Schnitt hat) *genau einen* positiven GG-Zustand.

(c) Falls das RNW zusätzlich (V3') *schwach reversibel* ist, so hat das System einen positiven GG-Zustand, es tritt also (b) in Kraft.

Bemerkungen:

- (V3') ist eine Verschärfung der Bedingung (V3), denn

  (V3') $\Leftrightarrow$ schwach reversibel $\Leftrightarrow$ jede ZHK *ist gleich einer* terminalen starken ZHK $\Rightarrow$ (V3)

- die Voraussetzungen des Defekt-1-Theorems sind Abschwächungen der Voraussetzung des Defekt-0-Theorems:

  $$\delta = 0 \quad \begin{array}{c} \Longrightarrow \\ \not\Longleftarrow \end{array} \quad \text{(V1) \& (V2)}$$

  (In der Tat zeigt Feinberg, dass (V1)–(V3) hinreichend sind für die Äquivalenz von Spezies-GG und Komplex-GG; die stärkere Forderung $\delta = 0$ aus dem Defekt-Null-Theorem ist dazu nicht notwendig.)

- Die beiden Theoreme sagen nur etwas aus über Existenz/Eindeutigkeit von strikt positiven GG-Lösungen; darüber gibt es Überlegungen von Feinberg auch über die Existenz von GG-Lösungen, die sich *am Rande* des positiven Rektanden befinden.

Hier könnte man ein Kapitel einfügen über GG-Punkte am Rande des positiven Rektanden.

7 Das PDE-Modell

In diesem Kapitel wollen wir die Ideen aus Kap. 5 auf das PDE-Modell übertragen. Wir treffen im Verlauf des Kapitels folgende Annahmen:

- Alle Reaktionen nach Massenwirkungsgesetz.
- Alle Reaktionen reversibel.
- Dispersion-Diffusion ist speziesunabhängig: $L = \text{diag}(L_1, ..., L_I)$ mit $L_1 = ... = L_I$
- Nur zur Vereinfachung der Darstellung: Der Diffusions-Dispersionskoeffizient skalar und konstant, auch die Porosität sei konstant
- homogene Neumann-Randbedingungen (typische Ausström-Randbedingung) (Ergebnis kann auf sog. (inhomogene) Fluss-Randbedingungen übertragen werden $\rightarrow$ typische Randbedingung für den Einströmrand))

Wir starten mit einem Nachweis der *Eindeutigkeit* von Lösungen. Dies geschieht mittels Energie-Methoden und Gronwall-Lemma. Die Existenz von Lösungen ist schwieriger zu zeigen. Zur Vorbereitung des Existenz-Resultats zeigen wir zunächst die Nichtnegativität von Lösungen. Der anschließende Beweis der Existenz einer (globalen!) Lösung des PDE-Problems (in einem noch zu bestimmenden Funktionenraum) verwendet einen sog. *Fixpunktsatz*. Die wesentliche Voraussetzung zur Anwendung eines Fixpunktsatzes ist, dass jede potenzielle Lösung eine sog. *a-priori-Schranke* erfüllt. Das Prinzip der a-priori-Schranken/Fixpunktsätze lautet in etwa wie folgt: Wenn man nachweisen kann, dass jede Lösung der PDE (Anfangs-Randwert-Problem) eine Schranke erfüllt, die nur von den Daten abhängt, so liefert der Fixpunktsatz die Existenz einer Lösung.
Zum Nachweis der Existenz einer a-priori-Schranke verwenden wir das Lyapunov-Funktional aus Kap. 5.2 erneut.
Unter 'Existenz einer *globalen* Lösung' verstehen wir hier, dass für *beliebiges* $T>0$ eine Lösung auf dem Zeitintervall $[0,T]$ gefunden werden kann.

Als Funktionenraum, in dem wir nach Lösungen suchen, wählen wir (siehe [WYW] S. 13) den 'anisotropen' Hölder-Raum[26] $C^{2+\alpha,1+\frac{\alpha}{2}}(Q_T)^I$,

$$C^{2+\alpha,1+\frac{\alpha}{2}}(Q_T) := \{u \in C^{2,1}(\overline{Q}_T) \,|\, \sup_{(t,x),(s,y)\in Q_T} \frac{|u(t,x)-u(s,y)|}{(|t-s|^2+|x-y|)^{\frac{\alpha}{2}}} < \infty\}, \quad 0<\alpha<1.$$

Dabei bezeichnet $C^{2,1}(Q_T)$ Funktionen, die zweimal nach x und einmal nach t differenzierbar sind.[27] Es sind auch andere Räume mit geringerer Regularität verwendbar[28] (die dann auch geringere Anforderungen an die Daten stellen), wie der Sobolev-Raum

$$W_p^{2,1}(Q_T)^I := \{u : Q_T \longrightarrow \mathbb{R}^I \,|\, \|u_i\|_{W_p^{2,1}(Q_T)} < \infty\},$$
$$\|u_i\|_{W_p^{2,1}(Q_T)} := \left[\|u\|^p_{L^p(Q_T)} + \|\partial_t u\|^p_{L^p(Q_T)} + \sum_i \|\partial_{x_i} u\|^p_{L^p(Q_T)} + \sum_{i,j} \|\partial_{x_i}\partial_{x_j} u\|^p_{L^p(Q_T)}\right]^{\frac{1}{p}}$$
$$Q_T := (0,T]\times\Omega.$$

wobei $p \geq n+1$ gefordert wird, damit $W_p^{2,1}(Q_T) \subset W_p^1(Q_T)$ kompakt eingebettet ist in $C(\overline{Q}_T)$; für Gebiete M im m-dimensionalen ist der Raum $W_p^1(M)$ kompakt in $C(M)$ eingebettet für $p>m$ (beachte, dass $Q_T \subset \mathbb{R}^{n+1}$, d.h. $m=n+1$).

Die Wahl des Raums wird erst in Kap. 7.3-7.4 relevant; in Kap. 7.1-7.2 reicht es, dass 'alle vorkommenden Terme' existieren.

[26] vgl. [MiSi04]

[27] Klassischerweise bezeichnet man als Hölder-Raum $C^{m+\alpha}(\overline{\Omega})$ für $m \in \mathbb{N}_0$ und $\alpha \in [0,1)$ den Raum $C^{m+\alpha}(\overline{\Omega}) := \{u \in C^m(\overline{\Omega}) \,|\, \|u\|_{m+\alpha} < \infty\}$ mit der Norm $\|u\|_{m+\alpha} := \sum_{\substack{|\beta|\leq m \\ \beta\in(\mathbb{N}_0)^n}} \|D^\beta u\|_{L^\infty(\overline{\Omega})} + \sup_{\substack{x\neq y \\ x,y\in\overline{\Omega}}} \frac{|u(x)-u(y)|}{|x-y|^\alpha}$, $D^\beta := \partial^{\beta_1}_{x_1}...\partial^{\beta_n}_{x_n}$, $|\beta| := \sum_{i=1}^n \beta_i$.

[28] siehe [Habil]

7.1 Eindeutigkeit von Lösungen

Wir fangen an mit der Eindeutigkeit von Lösungen, da diese viel einfacher zu zeigen ist als die Existenz. Wir verwenden eine Form der sog. *Energie-Methode*. Bei dieser testet man die PDE mit der Lösung, um so eine 'Energie-Abschätzung' der Lösung zu bekommen. Um Eindeutigkeit zu zeigen, verwenden wir im folgenden die Differenz zweier Lösungen.

Satz. Seien $u, v \in W_p^{2,1}(Q_T)^I$, mit $p > n+1$, $T > 0$ Lösungen des PDE-Systems $\theta\partial_t u - d\Delta u + q \cdot \nabla u = SR(u)$ mit $\theta = \text{const} > 0$, $d = \text{const} > 0$, $q \in C^0(Q_T)$, die Reaktionen seien gemäß dem Massenwirkungsgesetz und reversibel. Es sei ein Anfangswert $u_{t=0} = u_0 \in L^2(\Omega)$ gegeben sowie Dirichlet- oder Neumann-Randwerte $u|_{[0,T]\times\partial\Omega} = g$ oder $\frac{\partial u}{\partial \nu}\big|_{[0,T]\times\partial\Omega} = g$. Dann ist $u = v$.

Beweis. Sei $w := u - v$. Dann erfüllt w die PDE

$$\theta\partial_t w - d\Delta w + q \cdot \nabla w = SR(u) - SR(v), \quad w|_{t=0} = 0 \quad \& \text{ hom. RW.}$$

Teste die i-te PDE mit w_i:

$$\frac{\theta}{2}\frac{d}{dt}\int_\Omega |w_i|^2\,dx + d\int_\Omega |\nabla w_i|^2\,dx - d\underbrace{\int_{\partial\Omega} w_i \frac{\partial w_i}{\partial \nu}\,do}_{=0} + \int_\Omega q\cdot\nabla w_i\, w_i\,dx = \int_\Omega [[SR(u)]_i - [SR(v)]_i] w_i\,dx$$

Wir möchten nun die rechte Seite abschätzen durch eine Konstante mal die L^2-Norm von w zum Quadrat, wobei die Konstante von u und v abhängen darf. Der Einfachheit halber führen wir dies, etwas unmathematisch, anhand eines Beispiel-Terms vor; man sieht, dass dies allgemein für MW-Raten durchführbar ist: Für die Einzelrate $u \mapsto k u_1^3 u_2^2$ expandieren wir den Term

$$\begin{aligned}(k u_1^3 u_2^2 - k v_1^3 v_2^2)\, w_i &= k[(u_1^3 u_2^2 - u_1^3 u_2 v_2) + (u_1^3 u_2 v_2 - u_1^3 v_2^2) + (u_1^3 v_2^2 - u_1^2 v_1 v_2^2) \\ &\quad + (u_1^2 v_1 v_2^2 - u_1 v_1^2 v_2^2) + (u_1 v_1^2 v_2^2 - v_1^3 v_2^2)]\, w_i \qquad (7.1)\\ &= k\,[u_1^3 u_2 w_2 + u_1^3 v_2 w_2 + u_1^2 v_2^2 w_1 + u_1 v_1 v_2^2 w_1 + v_1^2 v_2^2 w_1]\, w_i;\end{aligned}$$

bei den eingefügten Termen wurden sukzessive die v-Potenzen um eins erhöht und die u-Potenzen um eins verringert. Allgemein kann man so die rechte Seite schreiben als $\int_\Omega \sum_{j=1}^{I} w_i w_j f_{ij}(t,x)\,dx$, wobei sich die f_{ij} aus den Komponenten von u und v zusammensetzen. Da die $u_i, v_j \in W_p^{2,1}(Q_T) \subset C^\infty(\overline{Q}_T)$ liegen, können wir dies abschätzen unter Verwendung einer Konstante c, die von der $L^\infty(Q_T)$-Norm der f_{ij} (also der u_i, v_i) abhängt:

$$\frac{\theta}{2}\frac{d}{dt}\int_\Omega |w_i|^2\,dx + d\int_\Omega |\nabla w_i|^2\,dx \le \|q\|_{L^\infty(Q_T)} \int_\Omega \sum_{i=1}^{n} |\nabla w_i|\,|w_i|\,dx + c\,\|w_i\|_{L^2(\Omega)}\|w\|_{L^2(\Omega)^I}$$

Der 'gemischte' Term auf der rechten Seite wird mit der Ungleichung $ab \leq \epsilon a^2 + \frac{1}{4\epsilon} b^2$ (die unmittelbar aus $(\sqrt{\epsilon}a - \frac{1}{2\sqrt{\epsilon}}b)^2 \geq 0$ folgt) abgeschätzt:

$$\frac{\theta}{2}\frac{d}{dt}\int_\Omega |w_i|^2\,dx + d\int_\Omega |\nabla w_i|^2\,dx \leq \|q\|_{L^\infty(Q_T)}\,\epsilon \int_\Omega |\nabla w_i|^2\,dx + \frac{\|q\|_{L^\infty(Q_T)^n}}{4\epsilon}\int_\Omega |w_i|^2\,dx + c\,\|w_i\|_{L^2(\Omega)}\|w\|_{L^2(\Omega)^I}$$

Summation über i liefert

$$\frac{\theta}{2}\frac{d}{dt}\|w\|^2_{L^2(\Omega)^I} + (d - \epsilon\,\|q\|_{L^\infty(Q_T)^n})\sum_{i=1}^{I}\|\nabla w_i\|^2_{L^2(\Omega)^n} \leq c\,\|w\|^2_{L^2(\Omega)^I}$$

Man wählt nun $\epsilon > 0$ hinreichend klein, so dass der Koeffizient vor dem 'Energie-Term' positiv (oder nicht-negativ) ist (diese Technik heißt *Absorption*) und bekommt mit $h(t) := \|w(t,\cdot)\|^2_{L^2(\Omega)^n}$ die Differentialungleichung

$$h'(t) \leq \frac{2c}{\theta}\,h(t)$$

mit Anfangswert $h(0) = 0$, da $w|_{t=0} = 0$.
Mit dem *Gronwall-Lemma* folgt

$$h(t) \leq 0 \;\forall\, t \in [0,T],$$

somit $h(t) \equiv 0$, somit $w \equiv 0$, somit $u \equiv v$. □

7.2 Nichtnegativität von Lösungen

Auch wenn wir immer noch nichts über die Existenz von Lösungen wissen, können wir die Nichtnegativität von (potenziell existierenden) Lösungen zeigen. Wir verwenden wieder eine Form der Energie-Methode; jedoch testen wir die PDE mit dem negativen Anteil der Lösung.
Wir zerlegen (auch vektorwertige Funktionen, dann komponentenweise) $u = u^+ - u^-$ mit $u^+ := \max\{0,u\}$, $u^- := \max\{0,-u\} = -\min\{u,0\}$; es sind also $u^+, u^- \geq 0$. Es sei $\Omega_i^-(t) := \{x \in \Omega \mid u_i(t,x) < 0\}$. Für stetiges $u_i : Q_T \longrightarrow \mathbb{R}$ ist dies eine wohldefinierte, offene Menge. Im folgenden Satz betrachten wir ein leicht modifiziertes Problem mit rechter Seite $SR(u^+)$ statt $SR(u)$:

Satz. Sei $u \in W_p^{2,1}(Q_T)^I$ eine Lösung des PDE-Problems $\theta\partial_t u + Lu = SR(u^+)$ mit $u|_{t=0} \geq 0$ und homogenen Neumann-Randwerten, auf einem Existenzintervall $[0,T]$. Es sei q in L^∞ beschränkt. Dann ist $u \geq 0$ auf ganz Q_T.

Beweis. Sei u eine Lösung des modifizierten Problems auf einem Existenzintervall $[0,T]$. Wir testen die i-te Gleichung des System mit u_i^- und bekommen

$$\theta \int_\Omega u_i^- \partial_t u_i \, dx - d_i \int_\Omega u_i^- \, \Delta u_i \, dx + \int_\Omega u_i^- \, \nabla u_i \cdot q \, dx = \sum_{j=1}^J (s_{ij}^p - s_{ij}^e) \left[R_j^v(u^+) - R_j^r(u^+) \right] u_i^-$$

$$= \int_\Omega \sum_{j=1}^J \overbrace{\Big[\underbrace{s_{ij}^p R_j^v(u^+)}_{= s_{ij}^p k_j^v \prod_{k=1}^I (u_k^+)^{s_{kj}^e}} + \underbrace{s_{ij}^e R_j^r(u^+)}_{= s_{ij}^e k_j^r \prod_{k=1}^I (u_k^+)^{s_{kj}^p}} \Big] u_i^-}^{=:(\mathrm{I})} - \sum_{j=1}^J \overbrace{\Big[\underbrace{s_{ij}^e R_j^v(u^+)}_{= s_{ij}^e k_j^v \prod_{k=1}^I (u_k^+)^{s_{kj}^e}} + \underbrace{s_{ij}^p R_j^r(u^+)}_{= s_{ij}^p k_j^r \prod_{k=1}^I (u_k^+)^{s_{kj}^p}} \Big] u_i^-}^{=:(\mathrm{II})} \, dx$$

Alle auftretenden Faktoren und Exponenten sind nichtnegativ. Im vorderen Term von (II) ist immer dann, wenn der Vorfaktor s_{ij}^e nicht verschwindet, im Produkt ein Faktor $(u_i^+)^{s_{ij}^e}$ mit nichtverschwindendem Exponenten vorhanden, d.h. wegen der Multiplikation mit u_i^- ist das Produkt null; analog für den hinteren Term von (II). Es ist also (II)$=0$. Der Term (I) verschwindet dagegen i.a. nicht; es ist i.a. (I)≥ 0. Auf der linken Seite nutzen wir aus, dass $u_i^- = -u_i$ auf $\Omega_i^-(t)$ und $=0$ sonst, um nach Multiplikation mit (-1)

$$\theta \int_\Omega u_i^- \partial_t u_i^- \, dx - d_i \int_{\Omega_i^-(t)} u_i \, \Delta u_i \, dx + \int_{\Omega_i^-(t)} u_i \, \nabla u_i \cdot q \, dx = -(I) \leq 0$$

zu bekommen. Nun geht es weiter wie im Beweis der Eindeutigkeit in Kap. 7.1 und wir erhalten

$$\frac{\theta}{2} \frac{d}{dt} \int_\Omega |u_i^-(t,x)|^2 \, dx \leq \text{const} \int_\Omega |u_i^-(t,x)|^2 \, dx \, .$$

Daraus folgt, zusammen mit der Nichtnegativität des Anfangswertes, also $\int_\Omega |u_i^-(0,x)|^2 \, dx = 0$, und mit dem Gronwall-Lemma, dass $\int_\Omega |u_i^-(t,x)|^2 \, dx \leq 0$, somit $u_i^- \equiv 0$, somit $u_i \geq 0$. □

Jede Lösung des modifizierten Problems (mit nichtnegativen Anfangswerten) ist also nichtnegativ, und ist somit auch Lösung des ursprünglichen Problems: $\mathbb{L}_{mod} \subseteq \mathbb{L}$. In Kap. 7.1 haben wir gezeigt, dass $\mathbb{L}$ höchstens ein Element enthält. Wir müssen also nur noch zeigen, dass $\mathbb{L}_{mod}$ mindestens ein Element enthält, um $\mathbb{L}_{mod} = \mathbb{L}$ und somit die Existenz und Eindeutigkeit der Lösung des ursprünglichen Problems zu zeigen, d.h. beim Existenzbeweis können wir uns auf das modifizierte Problem konzentrieren.
Die Existenz einer Lösung des modifizierten Problems wird in Kap. 7.4 gezeigt; Kap. 7.3 ist als Vorbereitung nötig.

7.3 A priori-Schranken

Als Vorbereitung auf den Beweis der Existenz einer Lösung des modifizierten Problems brauchen wir eine sog. a priori-Schranke. Wir verwenden dazu das

Lyapunov-Funktional aus Kap. 5.2 hier erneut. Als ersten Schritt zeigen wir:

Lemma. Sei $\varphi : (\mathbb{R}_0^+)^I \longrightarrow \mathbb{R}$ wie in Kap. 5.2. Es sei $g : Q_T \longrightarrow \mathbb{R}$ definiert durch $g := \varphi \circ u$, wobei u eine Lösung des modifizierten Problems sei auf einem Existenzintervall $[0, T]$. Dann[29] gilt

$$\theta \partial_t g + Lg \leq 0$$

wobei

$$Lu := (-d_i \Delta u_i + q \cdot \nabla u_i)_{i=1,\ldots,I}.$$

Beweis.[30] Mit μ aus Kap. 5.2 und der Kettenregel ist

$$\begin{aligned}
\partial_t g &= ((\nabla\varphi) \circ u) \cdot \partial_t u = (\mu + \ln u) \cdot \partial_t u \\
\partial_{x_i} g &= (\mu + \ln u) \cdot \partial_{x_i} u = \sum_{k=1}^{I} (\mu_k + \ln u_k)\, \partial_{x_i} u_k \\
\partial_{x_i}^2 g &= \sum_{k=1}^{I} \underbrace{\frac{1}{u_k} (\partial_{x_i} u_k)^2}_{\geq 0} + \sum_{k=1}^{I} (\mu_k + \ln u_k)\, \partial_{x_i}^2 u_k \geq (\mu + \ln u_k) \cdot \partial_{x_i}^2 u \\
&\Longrightarrow -\Delta g \leq -(\mu + \ln u) \cdot \Delta u
\end{aligned}$$

Es folgt

$$\begin{aligned}
\theta \partial_t g + Lg \leq (\mu + \ln u) \cdot (\theta \partial_t u + Lu) &\overset{\text{(PDE)}}{=} (\mu + \ln u) \cdot SR(u) = S^T (\mu + \ln u) \cdot R(u) \\
&= (-\ln K + S^T \ln u) \cdot R(u) \leq 0
\end{aligned}$$

wobei die Nichtpositivität am Ende wie in Kap. 7.2 folgt. □

Beachte, dass im Beweis die Spezies-Unabhängigkeit des Diffusions-Dispersionskoeffizienten verwendet wurde.

Wir verwenden nun:

Satz ((parabolisches, schwaches) Maximum-Prinzip). Eine Funktion $g \in C^2(Q_T) \cap C(\overline{Q}_T)$ erfülle

$$\partial_t g - \sum_{i,j} a_{ij}(t,x) \partial_{x_i} \partial_{x_j} g + \sum_i b_i(t,x)\, \partial_{x_i} g \leq 0$$

[29] Damit ∇g wohldefiniert ist, braucht man die strikte Positivität von Lösungen u, wohingegen wir in Kap. 7.2 lediglich die Nichtnegativität von Lösungen gezeigt haben. Als Ausweg kann man $g_\delta := \varphi \circ u_\delta$, wobei $u_\delta(t,x) := u(t,x) + \delta$, für $\delta > 0$ betrachten; es ist ∇g_δ wohldefiniert. Der Beweis des Lemma zeigt dann, dass $\theta \partial_t g_\delta + L g_\delta \leq f(\delta)$ ist, wobei f eine Funktion ist mit $f(\delta) \overset{(\delta \to 0)}{\longrightarrow} 0$; diese Schranke ist für die nachfolgenden Überlegungen ebenfalls hinreichend.

[30] Der Beweis folgt in etwa der Argumentation in [MiSi04].

mit $\sum_{i,j} a_{ij}(t,x)\xi_i\xi_j \geq 0\ \forall\, \xi$ ('Elliptizität'). Dann ist

$$\max_{Q_T} g = \max_{\partial_p Q_T} g,$$

wobei $\partial_p Q_T := \overline{Q}_T \backslash Q_T$ der sog. parabolische Rand von Q_T ist; $Q_T = (0,T]\times\Omega$. (D.h. das Maximum von g wird für $t=0$ oder für $x\in\partial\Omega$ angenommen.)[31]

Beweis: Siehe [Evans], Kap. 7.1.4.

Anwendung des Maximum-Prinzips auf g liefert, sofern wir 'geeignete' Randbedingungen anehmen (s.u.), dass g sein Maximum bei $t=0$ annimmt, somit g (und damit auch u!) durch die Anfangsdaten $u_0 = u|_{t=0}$ abgeschätzt werden kann:

Satz. Unter den Voraussetzungen des obigen Lemmas ist u auf Q_T beschränkt durch eine Schranke, die nur von den Daten abhängt:

$$0 \leq u(t,x) \leq c \ \ \forall\ (t,x) \in \overline{Q}_T$$

Beweis. Wir bilden $\tilde{g}(t,x) := g(t,x) - \epsilon t/\theta$. Aus obigem Lemma folgt sofort, dass $\theta\tilde{g} + L\tilde{g} \leq -\epsilon$. Anwendung des Maximum-Prinzips auf die Funktion $\tilde{g}$ liefert, dass (a) $\tilde{g}$ sein Maximum bei einem $(t_0,x_0)\in\{0\}\times\overline{\Omega}$ oder (b) bei einem $(t_0,x_0)\in(0,T]\times\partial\Omega$ annimmt. Angenommen der Fall (b) tritt ein.
Da $\frac{\partial\tilde{g}}{\partial\nu} = \frac{\partial g}{\partial\nu} = (\mu + \ln u)\,\frac{\partial u}{\partial\nu}$ und da für u homogene Neumann-Randbedingungen angenommen wurden, ist $\frac{\partial\tilde{g}}{\partial\nu} = 0$. Da (t_0,x_0) Maximalstelle sein soll, folgt $\frac{\partial\tilde{g}}{\partial t}(t_0,x_0) = 0$ und $\frac{\partial\tilde{g}}{\partial\tau}(t_0,x_0) = 0$ (τ sei eine beliebige Tangentialrichtung an $\partial\Omega$ im Punkt (t_0,x_0)), sowie $-\Delta\tilde{g}(t,x) \geq 0$ in einer Ω-Umgebung[32] von (t_0,x_0). Es folgt also $\theta\tilde{g} + L\tilde{g} \geq 0$ in einer Ω-Umgebung von (t_0,x_0). Widerspruch. Es tritt also Fall (a) ein.
Es ist somit $\max_{(t,x)\in\overline{Q}_T} \tilde{g}(t,x) \leq \max_{x\in\overline{\Omega}} \tilde{g}(0,x) = \max_{x\in\overline{\Omega}} g(0,x) = \max_{x\in\overline{\Omega}} \varphi(u_0(x))$. Ferner ist, wie in Kap. 5.2 festgestellt, $u(t,x) \leq \varphi(u(t,x)) = g(t,x) \leq \tilde{g}(t,x) + \epsilon T/\theta$. Es folgt die Beschränktheit von u durch die Schranke $\frac{\epsilon T}{\theta} + \max_{x\in\Omega} \varphi(u_0(x))$. Da dies für alle $\epsilon > 0$ funktioniert, kann mit $\epsilon \to 0$ der ϵ-Term sogar fallengelassen werden.[33] □

Alternative Strategien/Räume. In schwächeren Räumen, in denen das Maximum-Prinzip nicht anwendbar ist, kann die Funktion[34]

$$\psi_r : W \to \mathbb{R}_+, \quad \psi(u) := \int_\Omega \varphi(u(\cdot,x))^r\, dx, \quad r\in\mathbb{N},$$

[31] Für die Koeffizienten der PDE wird keine Regularität vorausgesetzt. Essentiell ist aber, dass g 'glatt' ist; "$g\in H^2(Q_T)$" würde *nicht* reichen!

[32] An der Stelle (t_0,x_0) selbst ist Δg nicht definiert

[33] im Fall, dass wir wie in Fußnote 29 vorgehen, nehmen wir $\tilde{g}_\delta(t,x) := g_\delta(t,x) - \epsilon t/\theta$ anstelle von $\tilde{g}$, und wir wählen ϵ,δ so, dass $f(\delta) < \epsilon$.

[34] Auch hier in der Definition von ψ_r muss u strenggenommen durch $u_\delta := u + \delta$ ersetzt werden, wenn nicht a priori klar ist, dass u echt positiv ist.

betrachtet werden, man kann zeigen, dass $t \longmapsto (\psi_r \circ u)(t)$, wobei u eine Lösung ist, monoton fallend ist (bei inhomogenen Fluss-Randbedingungen: Beschränktes Wachstum hat). Für $r \neq 1$ ist der Beweis zwar ähnlich wie in obigem Lemma, aber mühsamer, da beim Ableiten zusätzliche Terme auftreten (siehe [Habil]). Als Lohn der Mühe bekommt man eine Schranke für die $L^\infty([0,T], L^r(\Omega))$-Norm (insbesondere also für die $L^r(Q_T)$-Norm) der Lösung u. Eine solche Schranke ist viel 'wertvoller' als eine Schranke in $L^\infty([0,T], L^1(\Omega))$; siehe Verwendung dieser Schranke am Ende des folgenden Kapitels.

7.4 Existenz von globalen Lösungen

Eine Möglichkeit, Existenz von Lösungen von nichtlinearen PDEs zu zeigen, ist es, die PDE als Fixpunktproblem zu schreiben in einem geeigneten Funktionenraum, und dann einen Fixpunktsatz anzuwenden, der die Existenz eines Fixpunktes und somit einer Lösung der PDE liefert. Die typischen Voraussetzungen zur Anwendung eines Fixpunktsatzes sind (neben ggf. einigen technischen Anforderungen) die Kompaktheit des Fixpunktoperators oder dessen Definitonsbereichs und das Vorliegen einer a priori-Schranke für Fixpunkte. (Bei ODEs folgte aus dem Vorliegen einer a priori-Schranke viel einfacher durch Ausnutzen der Existenz von 'maximalen' Lösungen die Existenz einer globalen Lösung. Deren Existenz beruht allerdings ebenfalls auf einem Fixpunktprinzip, FP-Satz von Banach.)

Ein Fixpunktsatz, der recht wenige technische Anforderungen stellt, ist der Fixpunktsatz von Schaefer ([Schae55], siehe auch [Evans] S. 504):

Satz (Fixpunktsatz von Schaefer). Sei X ein reeller Banach-Raum und sei $Z : X \longrightarrow X$ kompakt. Ferner sei die Menge

$$M := \{x \in X \mid \exists\, \lambda \in [0,1] :\ x = \lambda Z(x)\}$$

beschränkt. Dann hat Z (mindestens) einen Fixpunkt.

Beweis: Siehe Anhang; wird auf den Fixpunktsatz von Schauder zurückgeführt, siehe Anhang. □

Dabei gilt:

Def. (kompakte Abbildung). Eine (i.a. nichtlineare) Abbildung $Z : X \longrightarrow Y$ zwischen zwei Banach-Räumen X, Y heißt *kompakt*, wenn sie stetig ist und für jede beschränkte Menge $M \subset X$ gilt dass $\overline{Z(M)}$ kompakt ist.

Umwandlung des PDE-Problems in ein FP-Problem. Das gegebene PDE-System

$$\partial_t u + Lu = SR(u^+)$$

mit Anfangsbedingung $u|_{t=0} = u_0 \geq 0$ und Randbedingungen kann als Fixpunktproblem für die nichtlineare(!) Abbildung

$$Z : X \longrightarrow X, \quad u \longmapsto v = Z(u),$$

wobei v Lösung des (linearen!) Problems

$$\partial_t v + Lv = SR(u^+),$$

mit den entsprechenden Anfangs- und Randbedingungen, geschrieben werden; X ist ein noch zu wählender Lösungsraum. Offensichtlich ist ein $x \in X$ genau dann ein Fixpunkt von Z, wenn u das obige PDE-Problem löst.
Um den FP-Satz von Schaefer anwenden zu können, müssen wir also (a) eine Schranke für die Menge M finden und (b) – durch Wahl eines geeigneten Funktionenraums X – dafür Sorge tragen, dass Z kompakt ist.

Zu (a): Beschränktheit von M: Sei $u \in X$ und $v = Z(u)$ Lösung der obigen linearen PDE; sei $w := \lambda v$. Die Menge M ist also diejenige Menge, für die $u = w$ ist. Multiplikation der linearen PDE mit λ ergibt, dass $\partial_t w + Lw = \lambda SR(u^+)$ ist; ferner $w|_{t=0} = \lambda u_0$, und auch in der Randbedingung tritt, falls sie inhomogen ist, ein solcher Faktor λ auf. Die Menge M ist also charakterisiert durch

$$M = \{x \in X \mid \exists\, \lambda \in [0,1] : \ \partial_t u + Lu = \lambda\, SR(u^+),\ u|_{t=0} = \lambda u_0, \text{ und Randbedingung}\}.$$

Da λ aus einer beschränkten Menge zu wählen ist, übertragen sich alle Abschätzungen aus den früheren Kapiteln (dort $\lambda = 1$) auch auf den Fall $\lambda \in [0,1]$. Nach Kapitel 7.3 wissen wir also, dass die Menge M beschränkt ist unter den dort gemachten Annahmen.

Zu (b): Kompaktheit von $Z : X \longrightarrow X$: Um Wohldefiniertheit und Kompaktheit von Z zu bekommen, nutzen wir die 'regularisierende Wirkung' des Lösens einer linearen parabolischen PDE

$$\partial_t v + Lv = f$$

aus.
Wir können z.B. den Raum $X := C^{2+\alpha, 1+\frac{\alpha}{2}}(\overline{Q}_T)$ wählen, mit $0 < \alpha < 1$. Sei nun $\alpha < \beta < 1$. Sei $u \in X$. Trivialerweise folgt, dass $u \in C^{\beta, \frac{\beta}{2}}(\overline{Q}_T)$. Daraus folgt, dass die gekappte Funktion $u^+ \in C^{\beta, \frac{\beta}{2}}(\overline{Q}_T)$. Damit sind auch alle Potenzen und polynomiellen Ausdrücke von u^+, somit auch $SR(u^+) \in C^{\beta, \frac{\beta}{2}}(\overline{Q}_T)$. Nach einem Satz aus der Theorie *linearer* parabolischer Differentialgleichungen ist bei rechter Seite $f = SR(u^+) \in C^{\beta, \frac{\beta}{2}}(\overline{Q}_T)$ die Lösung $v \in C^{2+\beta, 1+\frac{\beta}{2}}(\overline{Q}_T)$ ([WYW] Kap. 8.3.1). Dieser Raum wiederum ist kompakt eingebettet in X. (Dass die Hölder-Räume kompakt ineinander eingebettet sind, kann man unter Verwendung des Satzes von Arzela-Ascoli zeigen.)
Nehmen wir eine beschränkte Folge (u_n) in X, so zeigt obige Kette, dass die zugehörige Folge von Lösungen (v_n) eine in X konvergente Teilfolge hat; Z ist also kompakt.

Die Wohldefiniertheit und Kompaktheit von Z ergibt sich auch für die Wahl $X := W_p^{2,1}(Q_T)$ für $p > n+1$: Dieser Raum ist, wie am Anfang von Kap. 7 bereits festgestellt, kompakt eingebettet in $C^0(\overline{Q}_T)$; $u \in X \subset C^0(\overline{Q}_T)$ hat trivialerweise $f = SR(u^+) \in C^0(\overline{Q}_T)$ zur Folge, somit $f \in L^p(Q_T)$. Nach einem Satz über lineare parabolische Differentialgleichungen (siehe [WYW] Kap. 9.2.3 oder [Lady68]) ist dann die Lösung $v \in W_p^{2,1}(Q_T) = X$. Nehmen wir wieder eine beschränkte Folge (u_n) in X, so zeigt obige Kette, dass die zugehörige Folge von Lösungen (v_n) eine in X konvergente Teilfolge hat; Z ist also kompakt.
Bei dieser Wahl des Raumes X ist zu beachten, dass die a priori-Abschätzung für ψ_r (siehe Ende von Kap. 7.3) zunächst nur eine Abschätzung der Lösung v in der $L^\infty(0,T;L^r(Q_T))$ liefert, nicht in der Norm von $X = W_p^{2,1}(Q_T)$. Jedoch folgt, indem man $r = r(p)$ hinreichend groß wählt, aus $u \in L^\infty(0,T;L^r(Q_t))$, dass $f = SR(u^+) \in L^p(Q_T)$, und die Theorie linearer parabolischer Differentialgleichungen liefert damit, dass die Lösung $v \in W_p^{2,1}(Q_T) = X$ (siehe [Habil]).

7.5 Reaktiver Transport mit Gleichgewichtsreaktionen, Herleitung eines Modells

Wir wollen nun das Problem mit Reaktionen im Gleichgewicht, das wir in Kap. 5.4 für das Batch-Problem betrachtet hatten, auf das PDE-Problem übertragen. Ausgangspunkt ist das 'kinetische' PDE-System

$$\partial_t u + Lu = SR(u) \tag{7.2}$$

für die I Konzentrationen und die Annahme, dass alle J Reaktionen reversibel sind; es gilt also $R_j(u) = R_j^v(u) - R_j^r(u)$. Es sei $I < J$.

Ein erster Versuch. Für lange Zeiträume, d.h. im Limes $t \to \infty$, kann man vermuten (und auch zeigen), dass sich ein Konzentrationsvektor u derart einstellt, dass sich Vorwärts- und Rückwärtsrate jeweils kompensieren. Sind die Reaktionen sehr schnell, kann man erwarten, dass sich so ein Zustand sehr schnell einstellt. In einer Idealisierung (Reaktionsgeschwindigkeiten extremst schnell) wollen wir annehmen, dass sich ein solcher Zustand 'instantan' einstellt, dass also zu allen Zeiten und an allen Orten u die Gleichungen $R_j^v(u) = R_j^r(u)$, $j = 1, ..., J$, d.h.

$$R(u) = 0\,, \tag{7.3}$$

erfüllt wird. Im Falle des Massenwirkungsgesetzes und echter Positivität von Lösungen kann dies als

$$-\ln K + S^T \ln u = 0$$

geschrieben werden. Wir haben somit J Gleichungen für die I Unbekannten u_i, was sicherlich kein sinnvolles Modell ergibt. Wie können wir weitere Gleichungen

bekommen? Ein naives Einsetzen der Gleichung $R(u) = 0$ in das PDE-System (7.2) würde zusätzlich I homogene PDEs $\partial_t u_i + Lu_i = 0$ ergeben; jedoch hätten wir dann insgesamt $I+J$ Gleichungen für I Unbekannte zu erfüllen, was fragwürdig erscheint. Wie lautet das korrekte GG-System?

Heuristische Herleitung des GG-Problems. Um, anders als oben, die 'richtige' Anzahl von Gleichungen für unsere Unbekannten zu bekommen, gehen wir wie folgt vor: Wir eliminieren zunächst einmal die schnellen Reaktionen aus möglichst vielen der PDEs, d.h. wir konzentrieren die Raten auf möglichst wenige PDEs. Um dies zu erreichen, übernehmen wir aus Kap. 5.4 die Koordinatentransformation $u = S\xi + U\eta$, $\xi = (S^T S)^{-1} S^T u$, $\eta = (U^T U)^{-1} U^T u$. Diese liefert, analog zu Kap. 5.4, das zu obigem kinetischen PDE-System äquivalente System[35]

$$\begin{aligned} \partial_t \eta + L\eta &= 0 \\ \partial_t \xi + L\xi &= R(S\xi + U\eta). \end{aligned}$$

Beachte, dass wir hier ausgenutzt haben, dass der Transportoperator L und die stöchiometrischen Matrizen kommutieren, was nur im Fall von spezies-unabhängigem Diffusions-Dispersionskoeffizienten der Fall ist.[36] Die obigen η-PDEs beschreiben die Tatsache, dass für kinetische Raten die Lösung die stöchiometrische Klasse $u_0 + \mathcal{S}$ nie verlässt. Es ist plausibel zu fordern, dass dies auch im Grenzfall von 'unendlich schnellen' Reaktionen so sein soll. Dies sind $I - J$ Gleichungen. Zusammen mit den J GG-Bedingungen (7.3) ergibt dies I Gleichungen für die I Unbekannten u_i:

$$\begin{aligned} \partial_t \eta + L\eta &= 0 \\ R(S\xi + U\eta) &= 0\,, \end{aligned}$$

bzw. gleichbedeutend:

$$\partial_t \eta + L\eta = 0 \tag{7.4}$$

$$S^T \ln(S\xi + U\eta) - \ln K = 0\,, \tag{7.5}$$

Man beachte, dass die Gleichungen *entkoppeln*, d.h. man kann *zuerst* die (skalaren, linearen) PDEs für die η_i lösen und *anschließend*, an jedem Punkt des Rechengebietes, die nichtlinearen algebraischen Gleichungen für die ξ_i. Die numerische Berechnung einer Lösung ist somit deutlich weniger zeitaufwändig als die des nichtlinearen, voll gekoppelten *kinetischen* PDE-Problems – wir haben ja nur $I - J$ viele PDEs anstelle von I vielen, und diese sind obendrein linear statt nichtlinear.

[35] Diese Transformation ist nur im Fall, dass die Spalten von S linear unabhängig sind, wohldefiniert; im anderen Fall ist jedoch, wie in Kap. 5.4 erwähnt, eine Modifikation der Vorgehensweise möglich, die vor dem R eine Matrix A erscheinen lässt.

[36] Diese Annahme braucht man eigentlich immer dann, wenn man Linearkombinationen der PDEs bilden und neue Variablen einführen will; das Anwenden von Matrizen auf das PDE-System entspricht in der Tat dem Bilden von Linearkombinationen der PDEs.

Die Existenz und Eindeutigkeit einer Lösung kann analog zum ODE-Fall gezeigt werden: Die Lösung der η-Gleichungen existiert und ist eindeutig, unter Verwendung sinnvoller Annahmen und Anfangs-/Randbedingungen[37] (lineare parabolische Theorie). Das Berechnen eines $\xi(t,x)$ kann an jedem einzelnen Punkt des Gebietes Q_T wieder als (endlichdimensionales!) Optimierungsproblem (im $\mathbb{R}^J$, wir brauchen keinen Funktionenraum!) umgeschrieben werden, und unter Ausnutzung der Eigenschaften der Zielfunktion φ (u.a. strikte Konvexität) kann Existenz und Eindeutigkeit von $\xi(t,x)$ für jedes $(t,x) \in Q_T$ gezeigt werden. Alternativ kann man auch den Satz über implizite Funktionen heranziehen, der zumindest die Existenz einer lokalen Auflösungsfunktion $\xi = \xi(\eta)$ liefert, da $\frac{\partial}{\partial \xi}[S^T \ln(S\xi + U\eta) - \ln K] = S^T \mathrm{diag}(u_1^{-1}, ..., u_I^{-1})S$, und diese Matrix ist symmetrisch positiv definit, also invertierbar.
Zur Deutung des Modells: Die Annahme, dass alle Reaktionen im lokalen (oder: dynamischen) GG sind, bedeutet nicht, dass diese Reaktionen nicht mehr ablaufen. Es bedeutet auch nicht, dass sich zu jedem Zeitpunkt an jedem Ort die Vorwärts- und die Rückwärts-Reaktionsrate genau ausgleichen (r ist nicht 0). Die GG-Bedingung $R(u) = 0$ beschreibt eine Mannigfaltigkeit, auf der die Lösung liegen muss, und wenn z.B. durch Transportvorgänge die Lösung diese Mannigfaltigkeit zu verlassen droht, dann laufen instantan Reaktionen dahingehend ab, dass die Lösung auf der Mannigfaltigkeit bleibt. Falls man neben den Konzentrationen u_i auch die sich dann ergebenen Reaktionsraten berechnen möchte, so kann man diese per

$$r = \partial_t \xi + L\xi \tag{7.6}$$

a posteriori berechnen.
Übrigens kann man die Transformation $u \mapsto (\eta, \xi)$ nun für das System (7.4)-(7.6) nun auch wieder rückabwickeln; man erhält

$$\begin{aligned} \partial_t u + Lu &= Sr \\ S^T \ln u - \ln K &= 0, \end{aligned}$$

In dieser Formulierung sind die I Konzentrationen u_i und die J Reaktionsraten r_j die Unbekannten, die man als Lösung eines Systems, bestehend aus I linearen PDEs und J nichtlinearen algebraischen (GG-)Gleichungen, bekommt. Diese Formulierung besteht also aus mehr Gleichungen und Unbekannten als (7.4)-(7.5).

Verallgemeinerung: Gemischtes GG-Kinetik-Problem. Man kann auch reaktive Systeme betrachten, in denen einige Rekationen 'schnell' und andere 'langsam' sind, d.h. nur für einen Teil $j = 1, ..., J_{GG}$ der Reaktionen nimmt man lokales Gleichgewicht an, und die übrigen $J_{kin} = J - J_{GG}$ Reaktionen $j = J_{kin} + 1, ..., J$ beschreibt man weiterhin kinetisch. In diesem Fall kann man das Ausgangsproblem zunächst als

$$\partial_t u + Lu = S_{GG} R_{GG}(u) + S_{kin} R_{kin}(u)$$

[37] Wir müssen nur für η Anfangs- und Randbedingungen stellen. Falls wir Anfangs-/Randbedingungen auch für ξ haben - oder anders ausgedrückt, falls wir Anfangs-/Randbedingungen für u haben – dann sollten diese konsistent zu den Gleichgewichtsbedingungen sein.

aufgeschreiben, wobei wir zerlegt haben $S = (S_{GG}|S_{kin})$ und $R_{kin}^T = (R_{GG}^T|R_{kin}^T)$. Wir wollen nun nur die schnellen Reaktionen aus möglichst vielen PDEs herauswerfen. Dazu kann nun die Koordinatentransformation mit S ersetzt durch S_{GG} durchgeführt werden, also $u = S_{GG}\xi + U\eta$, $\xi = (S_{GG}^T S_{GG})^{-1} S_{GG}^T\, u$, $\eta = (U^T U)^{-1} U^T\, u$, wobei nun $U \in \mathbb{R}^{I\times(I-J_{GG})}$ eine Matrix ist, deren Spalten $S_{GG}^{\perp}$ aufspannen; es ist nun $\xi \in \mathbb{R}^{J_{GG}}$, $\eta \in \mathbb{R}^{I-J_{GG}}$. Man erhält

$$\begin{aligned} \partial_t\eta + L\eta &= (U^T U)^{-1} U^T\, S_{kin} R_{kin}(S_{GG}\xi + U\eta) \\ \partial_t\xi + L\xi &= R_{GG}(S_{GG}\xi + U\eta) + (S_{GG}^T S_{GG})^{-1} S_{GG}^T\, S_{kin} R_{kin}(S_{GG}\xi + U\eta) \end{aligned}$$

und nach Annahme eines Gleichgewichts, also Ersetzung der ξ-PDEs durch die GG-Bedingungen,

$$\begin{aligned} \partial_t\eta + L\eta &= (U^T U)^{-1} U^T\, S_{kin} R_{kin}(S_{GG}\xi + U\eta) \\ R(S_{GG}\xi + U\eta) &= 0 \\ r_{GG} &= \partial_t\xi + L\xi - (S_{GG}^T S_{GG})^{-1} S_{GG}^T\, S_{kin} R_{kin}(S_{GG}\xi + U\eta) \end{aligned}$$

Anfangs- und Randbedingungen sind wieder nur für η zu fordern. Im Unterschied zum *reinen* GG-Problem sind beim gemischten GG-Kinetik-Problem die Gleichungen der $I - J_{GG}$-vielen η- und der J_{GG}-vielen ξ-Variablen nicht mehr entkoppelt. Nur die r_{GG}-Gleichungen sind weiterhin entkoppelt und können fallengelassen oder a posteriori berechnet werden. Einige Techniken, wie man dafür sorgen kann, dass immerhin *einige* der η-Gleichungen auch im gemischten Problem noch entkoppeln, findet man u.a. in [Kr07]. Der Existenzbeweis, der oben für das *reine* GG-Problem geführt wurde, lässt sich offenbar nicht auf das gemischte Problem übertragen, da die Nebenbedingung (=die η-PDEs) nun durch das Vorkommen von R_{kin} i.a. nichtlinear sein dürften, die zulässige Menge eines entsprechenden Optimierungsproblems also nicht mehr konvex ist. Das obige ξ-η-r_{GG}-Problem kann man per Rücktransformation wieder in eine leichter lesbare Form (aber numerisch aufwändiger zu lösende Form, da nun r_{GG} nicht mehr entkoppelt ist) bringen:

$$\begin{aligned} \partial_t u + Lu &= S_{GG} r_{GG} + S_{kin} R_{kin}(u) \\ R_{GG}(u) &= 0 \end{aligned} \tag{7.7}$$

In dieser Formulierung hat man $I + J_{GG}$ Gleichungen und Unbekannte.

8 Reaktionen mit immobilen Spezies (Mineralienausfällung und -auflösung), Komplementaritätsprobleme

Bisher haben wir nur die Reaktionen von im Fluid gelösten Spezies untereinander betrachtet; diese Vorgänge sind nicht an ein poröses Medium gebunden, sondern sie

können allgemein in Fluiden vorkommen (chemische Reaktoren, Verbrennung von Gasen,...) Im folgenden betrachten wir auch Reaktionen zwischen *mobilen* (d.h. im Fluid gelösten) und *immobilen* in der Bodenmatrix vorhandenen oder an der Bodenmatrix haftenden Spezies. Die mobilen Spezies werden durch PDEs, die immobilen durch ODEs beschrieben, wobei all diese Gleichungen i.a. gekoppelt sind. Reaktionen, die zwischen mobilen und immobilen Spezies ablaufen, werden als *heterogen* bezeichnet.

Es gibt, was die Struktur der entstehenden Gleichungen angeht, prinzipiell zwei Klassen von heterogenen Reaktionen:

- Sorptionsreaktionen
- Mineralienreaktionen

8.1 Sorptionsreaktionen

Ein Modell der Sorption auf Mikro-Skala: Die Oberfläche des Feststoffskeletts besteht aus einem oder mehreren Mineralien (z.B. FeOH). Die Mineralien-Teilchen, die sich an der Oberfläche der Bodenmatrix befinden, können mit Ionen des Fluids reagieren (z.B. mit H^+, Ca^{2+}, SO_4^{2-} zu $FeOH_2^+, FeO^-, FeOCa^+, FeSO_4^-$). Das FeOH-Oberflächen-Teilchen wird als *freier Sorptionsplatz* (*uncomplexed surface site*) bezeichnet, das Reaktionsprodukt (ebenfalls immobil) als *besetzter Sorptionsplatz* (*complexed surface site*). Die Reaktionen können oft mit dem MWG (näherungsweise) beschrieben werden, sowohl kinetisch als auch im lokalen Gleichgewicht.

Ein detaillierteres Modell: Double Layer Model (siehe [Be96] Kap. 8)
Annahme: Die Oberfläche kann eine Netto-Ladung bekommen; Oberflächenladungsdichte

$$\sigma = \frac{F n_w}{A} \sum_i z_i m_i \, [C/m^2],$$

wobei $F = 96.48$ Coulomb/mol die Faraday-Konstante ist, n_w die Wassermenge in kg pro Volumen, $z_i \in \mathbb{Z}$ die Ladung der Sorptionsplätze, m_i die Molarität der Sorptionsplätze (in Mol pro kg Wasser) un A die Groöße der Oberfläche pro Volumen ist.
Annahme: Es bildet sich im Fluid eine Schicht (Layer) mit Ionen (aus z.B. H^+, Ca^{2+}, SO_4^{2-}) aus, die diese Oberflächenladung kompensiert ($\rightarrow$ double layer). Die advektive Geschwindigkeit in Rand-Nähe (Poiseuille-Strömungsprofil) ist sehr gering, d.h. der Transport in der Randschicht wird völlig von Diffusion (thermische Bewegung) und elektrostatischen Kräften dominiert. Die Beschreibung dieser elektrostatischen Kräfte sowie einige Näherungen führen darauf, dass die 'effektive' Rate, mit der die Reaktion abläuft (die bestimmt wird von der Geschwindigkeit, mit der Ionen zur Oberfläche vordringen), beschrieben werden kann durch einen Reaktionsratenparameter

$$k = k_0 \exp\left(\frac{\Psi F}{RT}\right),$$

wobei T die Temperatur in Kelvin, R die universelle Gaskonstante in Joule/(Kelvin mal Mol)=Volt mal Coulomb pro Kelvin mal Mol und Ψ das elektrostatische Potenzial der Oberfläche ist. Das Argument der Exponentialfunktion ergibt sich aus der Beziehung zwischen Oberflächenladungsdichte und dem Potential:

$$\sigma = \sqrt{8 \cdot 10^3 RT\epsilon\epsilon_0 I} \, \sinh\left(\frac{\Psi F}{2RT}\right)$$

sowie der obigen Formel, die die Oberflächenladungsdichte in Beziehung zu den Molaritäten (also gewissermaßen: Konzentrationen) z_i setzt. Es ist dabei I die Ionenstärke im Fluid und $\epsilon_0 = 8.85 \cdot 10^{-12}$ und $\epsilon = 78.5$ bei 25 Grad Celsius.

Allgemeines (makroskopisches) Mehrkomponentenmodell. In einem allgemeinen Mehrkomponentenproblem haben wir zwei Phänomene, die sowohl die Analysis als auch die Numerik erschweren:

- sowohl mobile als auch immobile Spezies
- sowohl kinetische als auch GG-Reaktionen

Ein Modell, dass diese beiden Schwierigkeiten beinhaltet, hat die Struktur

$$\begin{aligned} \begin{pmatrix} \partial_t \vec{c} + L\vec{c} \\ \partial_t \vec{s} \end{pmatrix} &= S_{GG}\, \vec{r}_{GG} + S_{kin}\, \vec{R}_{kin}(\vec{c}, \vec{s}) \\ R_{GG}(\vec{c}, \vec{s}) &= 0 \end{aligned}$$

Dabei sind $\vec{c}$ und $\vec{s}$ die Vektoren der mobilen bzw. der immobilen Spezies-Konzentrationen, und $\vec{R} = (\vec{R}_{GG}^T, \vec{R}_{kin}^T)^T$ ist ein Vektor von gegebenen Ratenfunktionen, z.B. nach dem MWG. Teilt man die stöchiometrische Matrix weiter auf gemäß

$$S = (S_{GG}|S_{kin}) = \left(\frac{S^{mob}}{S^{immo}}\right) = \left(\begin{array}{c|c} S_{GG}^{mob} & S_{kin}^{mob} \\ \hline S_{GG}^{immo} & S_{kin}^{immo} \end{array}\right),$$

so kann man dies auch schreiben als

$$\begin{aligned} \partial_t \vec{c} + L\vec{c} &= S_{GG}^{mob}\, \vec{r}_{GG} + S_{kin}^{mob}\, \vec{R}_{kin}(\vec{c}, \vec{s}) \\ \partial_t \vec{s} &= S_{GG}^{immo}\, \vec{r}_{GG} + S_{kin}^{immo}\, \vec{R}_{kin}(\vec{c}, \vec{s}) \\ R_{GG}(\vec{c}, \vec{s}) &= 0. \end{aligned}$$

Die GG-Bedingung (s. Kap. 5.4) lautet im Fall des Massenwirkungsgesetzes

$$(S_{GG}^{mob})^T \ln \vec{c} + (S_{GG}^{immo})^T \ln \vec{s} = \ln \vec{K}.$$

Drei-Spezies-Sorptionsmodell. Häufig wird statt des allgemeinen Mehrspezies-Modells ein Drei-Spezies-Modell verwendet mit einer Sorptionsreaktion:

$$\alpha\, \mathrm{C} + \beta\, \mathrm{S} \rightleftarrows \mathrm{CS}$$

Dabei ist C ein mobiler Stoff (Konzentration: c), S bezeichnet die freien Sorptionsplätze ('Konzentration': $\tilde{s}$), und CS bezeichnet die belegten Sorptionsplätze ('Konzentration': s). Meist ist $\beta = 1$. Wir können entscheiden, ob wir die Reaktions als kinetisch oder als GG-Reaktion beschreiben wollen:

Kinetisches Modell: Unter Annahme des MWG ist die Reaktionsrate

$$R(c, s, \tilde{s}) = k_v\, c^\alpha\, \tilde{s}^\beta - k_r\, s,$$

und der Massenerhalt wird beschrieben durch

$$\begin{array}{rcl} \partial_t c + Lc & = & -\alpha R(c, s, \tilde{s}) \\ \partial_t \tilde{s} & = & -\beta R(c, s, \tilde{s}) \\ \partial_t s & = & R(c, s, \tilde{s}) \end{array}$$

Vereinfachung kann man erreichen, indem man weitere Annahmen trifft, z.B. dass $\tilde{s}$ näherungsweise konstant ist[38] oder nur wenig Einfluss auf die Rate hat. Dann ist

$$R(c, s, \tilde{s}) \approx \tilde{k}_v\, c^\alpha\ - k_r\, s = k_r \left(\frac{k_v}{k_r}\, c^\alpha - s\right) =: K\,(\varphi(c) - s).$$

Eine Funktion φ in diesem Modell heißt *Isotherme*; die Funktion $\varphi(c) = \frac{k_v}{k_r} c^\alpha$ heißt *Freundlich-Isotherme*. Das *Isothermen-Modell der Sorption (kinetisch)* lautet

$$\begin{array}{rcl} \partial_t(c - \alpha s) + Lc & = & 0 \\ \partial_t s & = & K\,(\varphi(c) - s) \end{array}$$

Gleichgewichtsmodell: Wenn man wieder annimmt, dass $\tilde{s}$ näherungsweise konstant ist oder dass die GG-Lage kaum von $\tilde{s}$ abhängt, kommt man von

$$\begin{array}{rcl} \partial_t c + Lc & = & -\alpha r \\ \partial_t s & = & r \\ R(c, s) & = & 0 \end{array}$$

auf

$$\begin{array}{rcl} \partial_t(c + \alpha s) + Lc & = & 0 \\ R(c, s) & = & 0. \end{array}$$

Geht man ferner davon aus, dass die Gleichung $R(c, s) = 0$ eine Auflösungsfunktion $s = \psi(c)$ hat, so lautet das *Isothermen-Modell der Sorption (im GG)*:

$$\partial_t(c + \alpha\psi(c)) + Lc = 0.$$

Die Funktion ψ heißt *Gleichgewichtsisotherme*. Es gibt verschiedene Modelle für ψ. Die zum kinetischen Freundlich-Modell passende GG-Isotherme ist offensichtlich $\psi(c) = \text{const} \cdot c^\alpha$.

Ohne die Annahme $\tilde{s} = \text{const}$ bekommt man, wenn $\alpha = \beta = 1$ offensichtlich $\tilde{s} + s = \text{const} =: K_S$ (was auch anschaulich völlig klar ist); die GG-Bedingung nach MWG lautet $s = K_{GG} c \tilde{s}$. Kombination ergibt $s = K_{GG} c (K_S - s)$, was sich zu

$$s = K_S\, \frac{K_{GG}\, c}{1 + K_{GG}\, c} =: \psi(c)$$

auflösen lässt. Obiges ψ heißt *Langmuir-Isotherme*. Lässt man allgemeines α zu, so erhält man die *verallgemeinerte Langmuir-Isotherme*

$$\psi(c) = K_S\, \frac{K_{GG}\, c^\alpha}{1 + K_{GG}\, c^\alpha}.$$

[38] Diese Annahme ist gerechtfertigt, wenn $s \ll s + \tilde{s}$; beachte auch: $\beta s + \tilde{s} = \text{const}$

8.2 Reaktionen mit Mineralien

In einem Drei-Spezies-Modell lautet eine Mineralienreaktion ("precipitation-dissolution reaction"), bestehend aus zwei Teilreaktionen

$$\begin{aligned} \alpha\,\mathrm{A} + \beta\,\mathrm{B} &\longrightarrow \mathrm{C} \quad \text{(Ausfällung, precipitation)},\\ \alpha\,\mathrm{A} + \beta\,\mathrm{B} &\longleftarrow \mathrm{C} \quad \text{(Auflösung, in Lösung gehen, dissolution)}. \end{aligned}$$

Dabei ist C ein immobiler Bestandteil der Bodenmatrix, und A, B sind mobile Stoffe (i.a. Ionen). Es gibt auch kompliziertere Mineralien-Reaktionen:

$$\sum_i \alpha_i\,\mathrm{A_i} \rightleftarrows \mathrm{C} + \sum_i \beta_i\,\mathrm{B_i},$$

wobei C wieder das Mineral ist und $\mathrm{A_i}, \mathrm{B_i}$ mobile Spezies. Natürlich gibt es oft *mehrere* solche Reaktionen in einem System. Man nimmt jedoch immer an, dass es zu jedem Mineral nur *eine* Reaktion gibt; der Block der Mineralien-Reaktionen lautet also, wenn man alle mobilen Spezies $\mathrm{X_i}$ auf eine Seite stellt (d.h. vorzeichenbehaftete $s_{ij} \in \mathbb{R}$ zulässt)

$$\sum_i s_{ij}\,\mathrm{X_i} \rightleftarrows \mathrm{C_j}, \quad j = \ldots$$

was in der stöchiometrischen Matrix einem Block $\begin{pmatrix} S_{min} \\ \hline -\mathrm{Id} \end{pmatrix}$ ausmacht.

Die wesentliche Besonderheit bei Reaktionen mit Mineralien ist, dass die Aktivität von Mineralien als konstant (oBdA $= 1$) angenommen wird, d.h. die Reaktionsrate der Auflösungsreaktion ist unabhängig von der Mineralienkonzentration.[39] Das in Aktivitäten formulierte MWG für das Drei-Spezies-Modell

$$R(c_1, c_2, c_3) = k_v \underbrace{a_1(\vec{c})}_{\approx c_1}{}^{\alpha} \underbrace{a_2(\vec{c})}_{\approx c_2}{}^{\beta} - k_r \underbrace{a_3(\vec{c})}_{\approx 1}$$

wird so zu

$$R(c_1, c_2) = \underbrace{k_v\, c_1^{\alpha} c_2^{\beta}}_{=:R_{prec}(c_1,c_2)} - \underbrace{k_r}_{=:R_{diss}}.$$

Die zugehörige GG-Bedingung ist

$$c_1^{\alpha} c_2^{\beta} = \frac{k_r}{k_v} =: K_{GG}\,.$$

Die Konstante K_{GG} heißt *Löslichkeitsprodukt* des Minerals C (der Begriff könnte aus dem Chemieunterricht der Schule bekannt sein). Die Unabhängigkeit der Rate von der

[39] Als Begründung kann man anführen, dass man annimmt, dass sich die Oberflächengröße des Minerals nur wenig verändert, wenn sich die Menge des Minerals verändert. Es gibt auch Modelle, in denen man versucht, die Größe der Mineraloberfläche (als Funktion der Mineralmenge) mit zu modellieren und die Rate von der Oberflächengröße abhängen zu lassen.

Mineralienkonzentration hat eine gewichtige Auswirkung: Für Mineralienkonzentration $c_3 \longrightarrow 0$ geht i.a. *nicht* der mit negativem Vorzeichen behaftete Quellterm der c_3-Differentialgleichung gegen null, d.h. der Quellterm für c_3 kann negativ sein, selbst wenn $c_3 = 0$; die Nichtnegativität von c_3 wäre so nicht gesichert (siehe Abb. 8). Das Modell muss also modifiziert werden. Für $c_3 = 0$ muss $0 \leq R_{diss} \leq R_{prec}$ gefordert werden

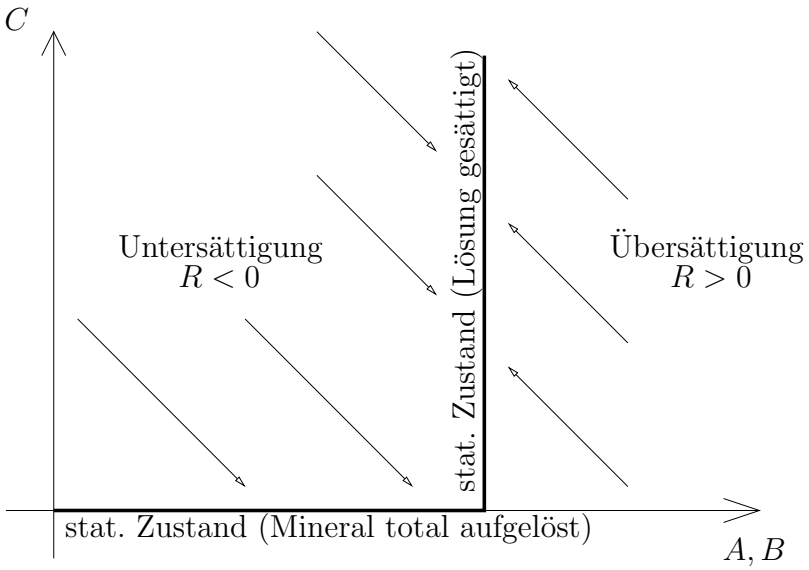


Abbildung 8: Die Pfeile beschreiben mögliche Veränderungen des Konzentrationsvektor (räumlicher Transport u. andere Reaktionen ignoriert). Die fette L-förmige Linie bezeichnet die stationären Zustände. Auf der vertikalen fetten Linie ist die Lösung gesättigt, auf der horizontalen ist das Mineral vollständig aufgelöst.

(statt: $R_{prec} = K_{GG}$). Eine mögliche Beschreibung lautet daher (s. Publikationen von Knabner & vanDuijn):

$$\begin{aligned} \partial_t c_1 + L c_1 &= -\alpha\, r \\ \partial_t c_2 + L c_2 &= -\beta\, r \\ \partial_t c_3 &= r \\ \text{wobei } \; r &\in k_v c_1^\alpha c_2^\beta - k_r\, H(c_3), \end{aligned}$$

und wobei H die mengenwertige Heaviside'funktion'

$$H(x) = \begin{cases} \{1\}, & x > 0 \\ [0,1], & x = 0 \\ \{0\}, & x < 0 \end{cases}$$

ist. Die Lösung hat i.a. recht geringe Regularität (siehe Übung); $t \mapsto c_3(t,x)$ kann unstetig sein (springen)! (Bei Knabner & vanDuijn wird an Sprungstellen $\partial_t c_3$ im Sinne von $\lim_{\epsilon \to 0, \epsilon > 0} \partial_t c(t+\epsilon)$ verstanden.) In der Dissertation von J. Hoffmann (Uni Erlangen,

2009) wird das Modell

$$\begin{aligned}
&\partial_t c_1 + Lc_1 = -\alpha\, r \\
&\partial_t c_2 + Lc_2 = -\beta\, r \\
&(c_3 = 0 \ \wedge\ \partial_t c_3 - k_v c_1^\alpha c_2^\beta + k_r \geq 0) \quad \vee \quad (c_3 \geq 0 \ \wedge\ \partial_t c_3 - k_v c_1^\alpha c_2^\beta + k_r = 0)
\end{aligned}$$

verwendet und Äquivalenz zur obigen Darstellung gezeigt.
Eine Bedingung der obigen Form, also

$$(f_i(\vec{x}) \geq 0 \ \wedge x_i = 0) \quad \vee \quad (f_(\vec{x}) = 0 \ \wedge x_i \geq 0)$$

heißt *Komplementaritätsbedingung* (KB). Eine solche kann immer äquivalent umgeschrieben werden zu

$$f_i(\vec{x}) \cdot x_i = 0 \ \wedge\ f_i(\vec{x}) \geq 0 \ \wedge\ x_i \geq 0.$$

Ein Problem, das (eine) Komplementaritätsbedingung(en) enthält, heißt *Komplementaritätsproblem* (KP). Eine KB heißt linear, falls das/die f_i linear sind. KPs können umgeschrieben werden in sog. *Variationsungleichungen*; es gibt eine ganze Theorie zu KPs/VUs (siehe Buch *Kinderlehrer, Variational Inequalities*) sowie zu numerischen Lösungsverfahren für diese Probleme. Diese Lösungsverfahren wurden im Rahmen der *Optimierung* entwickelt (denn Optimierungsprobleme mit Ungleichungs-Nebenbedingungen kann man über ihre 'KKT'-Bedingung als KPs schreiben).

Zum zugehörigen GG-Problem: Es gibt (siehe auch Skizze) zwei Fälle von GG:[40]
1. $R_{prec} = R_{diss}$ (und $c_3 \geq 0$), entspricht: Lösung ist gesättigt
2. $c_3 = 0$ und $R_{prec} \leq R_{diss}$, entspricht: Mineral ist total aufgelöst

Zusammenfassen lässt sich das als KB

$$(c_3 = 0 \ \wedge\ c_1^\alpha c_2^\beta \leq K_{GG}) \quad \vee \quad (c_3 \geq 0 \ \wedge\ c_1^\alpha c_2^\beta = K_{GG})$$

bzw.

$$(K_{GG} - c_1^\alpha c_2^\beta) \cdot c_3 = 0 \ \wedge\ c_3 \geq 0 \ \wedge\ K_{GG} - c_1^\alpha c_2^\beta \geq 0,$$

was zusammen mit den GG-Bedingungen

$$\begin{aligned}
\partial_t c_1 + Lc_1 &= -\alpha\, r \\
\partial_t c_2 + Lc_2 &= -\beta\, r \\
\partial_t c_3 &= r
\end{aligned}$$

oder kurz

$$\begin{aligned}
\partial_t (c_1 + \alpha c_3) + Lc_1 &= 0 \\
\partial_t (c_2 + \beta c_3) + Lc_2 &= 0
\end{aligned}$$

[40] Sie können dies selbst beobachten, wenn Sie Salz in ein Glass Wasser schütten: Nehmen Sie wenig Salz (so wenig, dass das Produkt aus Natrium- und Chlor-Ionen kleiner als das Löslichkeitsprodukt ist), dann löst sich das Salz komplett auf. Andernfalls löst sich nur so viel, wie sich aus dem Löslichkeitsprodukt ergibt; die Lösung ist dann gesättigt, un der Rest bleibt ungelöst als Feststoff auf dem Boden des Glases liegen.

zu lösen ist. Alternativ können die PDEs auch unter Einführung einer Reaktionsinvariante $\eta := \beta c_1 - \alpha c_2$ umgeschrieben werden zu

$$\begin{array}{lll} \partial_t \eta + L\eta & & = 0 \\ \partial_t(\underbrace{c_1}_{=\frac{1}{\beta}(\eta+\alpha c_2)} + \alpha c_3) + L \underbrace{c_1}_{=\frac{1}{\beta}(\eta+\alpha c_2)} & & = 0 \end{array}$$

was zusammen mit der KB drei Gleichungen für die drei Unbekannten η, c_2, c_3 darstellt. Der Vorteil dieser letzten Darstellung ist, dass die Gleichung für η entkoppelt ist. Zur Entkopplung ist ausgenutzt, dass η als Linearkombination von nur mobilen Spezies gebildet wurde.

Numerisches Lösen. Numerisches Lösen von nichtlinearen PDEs erfordert i.a eine Diskretisierung in Raum und Zeit sowie die Anwendung des Newton-Verfahrens (oder eines ähnlichen Verfahrens) zur Zurückführung auf lineare Gleichungssysteme. Doch wie behandelt man Komplementaritätsbedingungen, also Ungleichungen? Zum numerischen Lösen von KPs gibt es diverse Verfahren, die im Bereich der Optimierung entwickelt wurden. Eine recht einfache ist folgende: Man wählt eine Funktion $\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ mit der Eigenschaft

$$\varphi(a,b) = 0 \iff ab = 0 \ \wedge \ a \geq 0 \ \wedge \ b \geq 0.$$

Beispiele für Funktionen mit dieser Eigenschaft sind $\varphi(a,b) = \min\{a,b\}$ und $\varphi(a,b) = a + b - \sqrt{a^2+b^2}$. Man kann dann die KB (die ja *Ungleichungen* enthält) äquivalent umschreiben in eine *Gleichung(!)*

$$\varphi(f_i(\vec{x}), x_i) = 0;$$

die Ungleichungen sind 'verschwunden'. Der Preis, der dafür zu zahlen ist: Die obigen Funktionen φ sind nicht sehr glatt, sie erfüllen nicht die Anforderungen, die man klassischerweise fürs Newton-Verfahren (d.h. Funktion differenzierbar, Ableitung Lipschitzstetig) braucht. Jedoch: Man kann die Theorie des Newton-Verfahrens erweitern auf Funktionen recht geringer Regularität, auf sog. *strongly semismooth functions* (die Definition dieser Funktionenklasse ist recht technisch); das Newton-Verfahren für solche Funktionen ist, wie das klassische Newton-Verfahren, lokal quadratisch konvergent; man nennt es *semismooth Newton Method*.
Weitere Verfahren für nichtlineare KPs sind u.a. sog. *Innere-Punkte-Verfahren* und *Aktive-Mengen-Strategien* (siehe Bücher über Numerische Methoden für restringierte Optimierungsprobleme).

Allgemeines Mehrspezies-Modell mit Mineralienreaktionen.
Modellumfang:

- mehrere Mineralien(-reaktionen), jedoch alle im Gleichgewicht,

- daneben weitere GG-Reaktionen (unterteilt in Sorptionsreaktionen und Reaktionen innerhalb der mobilen Phase),
- ferner kinetische Reaktionen (sowohl Sorptionsreaktionen als auch Reaktionen innerhalb der mobilen Phase, jedoch keine Mineralien-Reaktionen)

Stöchiometrische Matrix:

$$S = \left(\begin{array}{c|c|c||c} S^1_{mob} & S^1_{sorp} & S^1_{min} & S^1_{kin} \\ \hline\hline 0 & S^2_{sorp} & 0 & S^2_{kin} \\ \hline 0 & 0 & -\mathrm{Id} & 0 \end{array}\right)$$

Die ersten drei Spalten enthalten GG-Reaktionen, in der vierten Spalte kinetische Reaktionen. Die erste Zeile gehört zu mobilen Spezies, die beiden anderen Zeilen zu immobilen Spezies, und zwar zuerst die Sorptionsplätze, danach die Mineralien. Der Vektor der (unbekannten) GG-Reationsraten ist dementsprechend unterteilt in $r_{GG} = (r^T_{mob}, r^T_{sorp}, r^T_{min})^T$. Das System für die unbekannten $c_{mob}, c_{sorp}, c_{min}, r_{mob}, r_{sorp}, r_{min}$ lautet (ablesbar an der Blockstruktur von S)

$$\begin{array}{lcll} \partial_t c_{mob} + L c_{mob} & = & S^1_{mob}\, r_{mob} + S^1_{sorp}\, r_{sorp} + S^1_{min}\, r_{min} & + S^1_{kin} R_{kin}(c_{mob}, c_{sorp}) \\ \partial_t c_{sorp} & = & S^2_{sorp}\, r_{sorp} & + S^2_{kin} R_{kin}(c_{mob}, c_{sorp}) \\ \partial_t c_{min} & = & -r_{min} & \end{array}$$

mit den GG-Bedingungen, die im Fall von Massenwirkungskinetik ("$S^T_{GG} \ln \vec{c} = \ln \vec{K}$") lauten

$$\begin{array}{ll} (S^1_{mob})^T \ln c_{mob} - \ln K_{mob} = 0 & \text{(GG der 'mobilen' Reaktionen)} \\ (S^1_{sorp})^T \ln c_{mob} + (S^2_{sorp})^T \ln c_{sorp} - \ln K_{sorp} = 0 & \text{(GG der Sorptionsreaktionen)} \\ \varphi(\ln K_{min} - S^1_{min} \ln c_{mob}, c_{min}) = 0 & \text{(GG der Mineralienreaktionen)} \end{array}$$

wobei die Komplementaritätsfunktion φ komponentenweise anzuwenden ist.

In den letzten 1980er bis 2010er Jahren sind viele verschiedene Umwandlungen dieses Systems bzw. meist von Modellen von etwas geringerem Umfang publiziert worden. Ein Verfahren, um dieses System in ein System reduzierter Größe umzuwandeln (Abkoppeln von Gleichungen, Elimination der r_{GG},...), welches sich dann schneller numerisch lösen lässt, findet man in [Habil], Varianten davon in der Dissertation von J. Hoffmann, Erlangen, 2009.[41] **** Hier könnte man noch ein Kapitel zu dem Thema einfügen ****

9 Anhang: Fixpunktsätze

In Sec. 7.4 haben wir einen Fixpunktsatz benötigt, um für das PDE-Modell die Existenz einer Lösung zu zeigen. Hier im Anhang wird in einer Kette von (Fixpunkt-)sätzen der

[41] Für dieses Verfahren gab es im Zusammenhang mit dem Lösen eines internationalen großen Benchmark-Problems für reaktiven Transport einen Preis für besondere Effizienz des Verfahrens. [CKK10, MoMaS10]

dort verwendete Fixpunktsatz von Schaefer hergeleitet. Beachte, dass die hier vorkommenden FP-Sätze immer nur die Existenz und nie die Eindeutigkeit eines Fixpunktes liefern (anders als der FP-Satz von Banach).

9.1 Fixpunktsatz im Endlichdimensionalen

Satz (Brouwer)[42] Sei B die abgeschlossene Einheitskugel im $\mathbb{R}^n$ bzgl. der Euklidischen Norm und sei $Z : B \longrightarrow B$ in $C^\infty(B)$. Dann hat Z (mindestens) einen Fixpunkt, also ein $x \in B$ mit $Z(x) = x$.

Bemerkung: Im Fall $n = 1$ ist die Aussage des Satzes anschaulich völlig klar. Sie ist eine elementare Folgerung aus dem Zwischenwertsatz. Der nun folgende Beweis lässt sich für alle $n \in \mathbb{N}$ führen; wir werden uns jedoch bei einem Beweisschritt auf den Fall $n = 2$ beschränken um die Darstellung weniger technisch zu machen.

Beweis. Angenommen, es gibt keinen Fixpunkt, also $x \neq Z(x)$ für alle $x \in B$. Für jedes $x \in B$ betrachten wir die Gerade $g_x(\alpha) := x + \alpha \underbrace{(x - Z(x))}_{\neq 0}$. g_x hat genau zwei Schnittpunkte mit ∂B (dazu mache man sich klar, dass eine tangentiale Lage von g_x nicht möglich ist). Es existieren also zwei verschiedene $\alpha_1, \alpha_2 \in \mathbb{R}$ mit $1 = \|x + \alpha(x - Z(x))\|^2$. Diese Bedingung kann als quadratische Gleichung geschrieben werden mit Lösung

$$\alpha_{1,2} = \alpha_{1,2}(x) = \frac{\langle x, Z(x) - x\rangle}{\|x - Z(x)\|^2} \pm \sqrt{\frac{|\langle x, Z(x)\rangle - x|^2 + (1 - \|x\|^2)\|x - Z(x)\|^2}{\|x - Z(x)\|^4}}.$$

Eine der Lösungen ist positiv, die andere negativ (ist auch anschaulich klar). Sei $\alpha_1(x)$ die positive Lösung. Sei $r(x) := g_x(\alpha_1) \in \partial B$ "der zu x gehörende Randpunkt". Die Abbildung

$$x \longmapsto r(x), \quad B \longrightarrow \partial B$$

ist glatt, da die Diskriminante strikt positiv ist und Z glatt ist. Betrachte nun die Abbildung

$$F : [0,1] \times B \longrightarrow B, \quad F(t,x) := \underbrace{x + t\alpha_1(x)(x - Z(x))}_{\in B}.$$

Es ist $F(0,x) = x$, also $F(0,\cdot) = \mathrm{id} : B \to B$.
Es ist $F(1,x) = r(x)$, also $F(1,\cdot) = r : B \to \partial B$.
Wir berechnen nun das Volumen $V(t) := \mathrm{Vol}(F(t,B))$ der Menge $F(t,B)$, $t \in [0,1]$:

[42] siehe [GilbTrud] S. 236

Sei $J_xF(t,x)$ die bezüglich des Arguments x gebildete Jacobi-Matrix von F. Es ist

$$\begin{aligned} V(t) &= \int\limits_B \det J_xF(t,x)\,dx, \\ V(0) &= \int\limits_B \det J_x(\mathrm{id})\,dx = \int\limits_B 1\,dx = \mathrm{Vol}(B), \\ V(1) &= \mathrm{Vol}(\partial B) = 0. \end{aligned}$$

Um einen Widerspruch dazu zu finden, wollen wir zeigen, dass $\frac{d}{dt}V(t) \equiv 0$. Dazu starten wir mit

$$\frac{d}{dt}V(t) = \int\limits_B \frac{\partial}{\partial t} \det\left[\frac{\partial}{\partial x_1}F(t,x), ..., \frac{\partial}{\partial x_n}F(t,x)\right] dx\,.$$

Zur Vereinfachung der Darstellung nehmen wir hier $n=2$ an; es ist
$\frac{\partial}{\partial t} \det\left[\frac{\partial}{\partial x_1}F(t,x_1,x_2), \frac{\partial}{\partial x_2}F(t,x_1,x_2)\right] = \frac{\partial}{\partial x_1} \det\left[\frac{\partial}{\partial t}F(t,x_1,x_2), \frac{\partial}{\partial x_2}F(t,x_1,x_2)\right] - \frac{\partial}{\partial x_2} \det\left[\frac{\partial}{\partial x_1}F(t,x_1,x_2), \frac{\partial}{\partial t}F(t,x_1,x_2)\right]$,
was elementar nachgerechnet werden kann durch Entwicklung der Determinanten.[43]
Nun ist

$$\begin{aligned} \int\limits_B \frac{\partial}{\partial x_1} \det\left[\frac{\partial}{\partial t}F, \frac{\partial}{\partial x_2}F\right] dx &= \int\limits_{-1}^{1} \left(\int\limits_{x_1=-\sqrt{1-x_2^2}}^{x_1=+\sqrt{1-x_2^2}} \frac{\partial}{\partial x_1} \det\left[\frac{\partial}{\partial t}F, \frac{\partial}{\partial x_2}F\right] dx_1 \right) dx_2 \\ &= \int\limits_{-1}^{1} \left(\det\left[\frac{\partial}{\partial t}F(t,x_1,x_2), \frac{\partial}{\partial x_2}F(t,x_1,x_2)\right] \Big|_{x_1=-\sqrt{1-x_2^2}}^{x_1=+\sqrt{1-x_2^2}} \right) dx_2. \end{aligned}$$

Die beiden dort vorkommenden Punkte (x_1,x_2) liegen auf dem Rand von B, dort aber ist $x=r(x)$, also $\alpha_1(x)=0$, also $F(\cdot,x)=x$ für diese Punkte x, also $\frac{\partial}{\partial t}F(t,x)=0$ für $x=(\pm\sqrt{1-x_2^2},x_2)$. So folgt, dass $\int\limits_B \frac{\partial}{\partial x_1} \det\left[\frac{\partial}{\partial t}F, \frac{\partial}{\partial x_2}F\right] dx = 0$. Analog zeigt man, dass $\int\limits_B \frac{\partial}{\partial x_2} \det\left[\frac{\partial}{\partial x_1}F, \frac{\partial}{\partial t}F\right] dx = 0$. Es folgt $\frac{d}{dt}V(t) \equiv 0$, was ein Widerspruch zu $V(1) \neq V(0)$ ist. □

Abschwächung der Regularitätsvoraussetzung:

Satz (Fixpunktsatz von Brouwer) (eigentlich: Bohl 1904)
Sei B wie oben, sei $Z : B \longrightarrow B$ stetig. Dann hat Z einen Fixpunkt.

Beweis. Sei $Z = (Z_1, ..., Z_n)$, also $Z_i : B \longrightarrow [-1,1]$. Nach dem Weierstraß'schen Approximationssatz gibt es Folge von Polynomen $Z_i^k : B \longrightarrow \mathbb{R}$ mit $|Z_i^k(x) - Z_i(x)| \leq$

[43] Eine Verallgemeinerung dieser Gleichung auf $n \in \mathbb{N}$ findet man im Buch [GilbTrud], Lemma 10.12, S. 235.

$\epsilon\,\forall\, x\in B\,\forall\, k\geq K(\epsilon)$.[44] Wegen der Äquivalenz von Normen des $\mathbb{R}^n$ folgt $\|Z^k(x)-Z(x)\|\leq \sqrt{n}\,\epsilon$. Setze $\tilde{Z}^k(x) := (1-\sqrt{n}\,\epsilon)\,Z^k(x)$. Es ist dann $\tilde{Z}^k : B \longrightarrow \mathbb{R}^n$ eine C^∞-Funktion, und das Bild ist für $\epsilon\leq \frac{1}{\sqrt{n}}$ in B, denn:

$$\begin{aligned}\|\tilde{Z}^k(x)\| &= \underbrace{(1-\sqrt{n}\,\epsilon)}_{\geq 0}\,\|Z^k(x)\| \leq (1-\sqrt{n}\,\epsilon)\,\big(\underbrace{\|Z(x)\|}_{\leq 1}+\underbrace{\|Z^k(x)-Z(x)\|}_{\leq\sqrt{n}\epsilon}\big)\\ &\leq (1-\sqrt{n}\,\epsilon)\,(1+\sqrt{n}\,\epsilon) = 1-n\epsilon^2 \leq 1\end{aligned}$$

Die C^∞-Version des FP-Satzes kann also auf $\tilde{Z}^k$ angewendet werden; es sei x_k Fixpunkt von $\tilde{Z}^k$. Nach dem Satz von Bolzano-Weierstraß hat die Folge (x_k) eine Teilfolge, die wieder mit (x_k) bezeichnet wird, die gegen ein $x\in B$ konvergiert (da B abgeschlossen und beschränkt). Es folgt

$$\|Z(x)-x\| \leq \underbrace{\|Z(x)-Z(x_k)\|}_{=:\text{(I)}}+\underbrace{\|Z(x_k)-\tilde{Z}^k(x_k)\|}_{=:\text{(II)}}+\underbrace{\|\tilde{Z}^k(x_k)-x_k\|}_{=0}+\underbrace{\|x_k-x\|}_{\to 0}$$

Term (I) konvergiert gegen null für $k\to\infty$, da Z stetig ist und $x_k\to x$. Term (II) konvergiert gegen null, da $\tilde{Z}_k$ gleichmäßig gegen Z konvergiert. Es folgt $Z(x)=x$. □

Satz (Weitere Verallgemeinerung). Statt einer Kugel B kann man in obigen Sätzen jede Menge $\tilde{B}\subset\mathbb{R}^n$ nehmen, die homöomorph zu B ist (d.h. es existiert eine bijektive, stetige Abbildung von B nach $\tilde{B}$, deren Umkehrabbildung auch stetig ist).[45]

Beweis. Sei $H : B \longrightarrow \tilde{B}$ ein Homöomorphismus und sei $Z : \tilde{B}\longrightarrow\tilde{B}$ stetig. Dann ist $\tilde{Z} := H^{-1}\circ Z\circ H : B\longrightarrow B$ stetig, hat also nach obigem Satz einen Fixpunkt $x\in B$. Es folgt $Z\circ H(x)=H(x)$, d.h. Z hat den Fixpunkt $H(x)\in\tilde{B}$. □

9.2 Fixpunktsätze in Banach-Räumen

Es geht nun darum, obiges Existenzresultat für Fixpunkte auf (i.a. unendlichdimensionale) Banach-Räume zu übertragen.

Satz (Fixpunktsatz von Schauder). (Schauder 1930, siehe [Evans] S. 502)
Sei X ein reeller Banach-Raum. Sei $K\subset X$ kompakt und konvex. Sei $Z : K\longrightarrow K$ stetig. Dann hat Z einen Fixpunkt.

[44] Die direkte Anwendung der obigen C^∞-Version des FP-Satzes auf diese C^∞-Funktionen Z^k scheitert daran, dass sie i.a. nicht nach B abbilden.

[45] Anschaulich und etwas vereinfachend heißt das, dass man abgeschlossene Mengen, die keine Löcher haben, nehmen kann, und deren Rand sogar Ecken/Knicke haben kann mit Winkeln, die strikt zwischen 0 und 2π liegen.

Beweis. Sei $\epsilon > 0$. Die Menge aller offenen Kugeln $B(x, \epsilon)$, wobei $x \in K$, bildet trivialerweise eine offene Überdeckung von K. Da K kompakt, gibt es eine endliche Überdeckung $B(x_i, \epsilon)$, $i = 1, ..., m = m(\epsilon)$ von K:

$$K \subseteq \bigcup_{i=1}^{m} B(x_i, \epsilon)$$

Sei K_ϵ definiert als die konvexe Hülle der Menge $\{x_1, ..., x_m\}$:

$$K_\epsilon := \{x = \sum_{i=1}^{m} \lambda_i x_i \mid \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1\}$$

Es ist $K_\epsilon \subseteq K$, da $x_i \in K$ und da K konvex.
$K_\epsilon \subset X$ ist homöomorph[46] zu einem Polyeder (:=Schnitt von Halbräumen) im $\mathbb{R}^n$. Nach Kap. 9.1 hat somit jede stetige Abbildung von K_ϵ nach K_ϵ einen Fixpunkt.
Wir definieren nun zunächst $f_\epsilon : K \longrightarrow K_\epsilon$ durch

$$f_\epsilon(x) := \frac{\sum_{i=1}^{m} \operatorname{dist}(x, K \setminus B(x_i, \epsilon))\, x_i}{\sum_{i=1}^{m} \operatorname{dist}(x, K \setminus B(x_i, \epsilon))} \in K_\epsilon \quad \forall\, x \in K\,.$$

Die Abbildung ist wohldefiniert, da im Nenner für jedes x niemals alle Summanden verschwinden können. Die Abbildung ist stetig, da 'dist' stetig ist. Es gilt

$$\|f_\epsilon(x) - x\| \leq \frac{\sum_{i=1} \operatorname{dist}(x, K \setminus B(x_i, \epsilon))\, \|x_i - x\|}{\sum_{i=1} \operatorname{dist}(x, K \setminus B(x_i, \epsilon))} \leq \epsilon;$$

dabei gilt die letzte Ungleichung, weil für jedes x und i entweder $x \in B(x_i, \epsilon)$, somit $\|x_i - x\| \leq \epsilon$ ist, oder aber $\operatorname{dist}(x, K \setminus B(x_i, \epsilon)) = 0$ ist.
Nun definieren wir noch

$$Z_\epsilon : K_\epsilon \longrightarrow K_\epsilon, \quad Z_\epsilon(x) := \underbrace{f_\epsilon(\overbrace{Z(\underbrace{x}_{\in K_\epsilon})}^{\in K})}_{\in K_\epsilon}\,.$$

Z_ϵ ist als Komposition von stetigen Funktionen stetig; nach obigen Überlegungen über K_ϵ hat Z_ϵ somit einen Fixpunkt $x_\epsilon \in K_\epsilon$. Sei (ϵ_j) eine positive Nullfolge. Da K kompakt,

[46] Falls man voraussetzt, dass die x_i linear unabhängig in X sind (was man o.B.d.A. tun kann), dann kann man als Homöomorphismus die Abbildung $H : \mathbb{R}^m \to X$, die den Vektor λ auf das zugehörige x schickt (siehe Def. von K_ϵ), nehmen.

gibt es eine Teilfolge, wieder mit (ϵ_j) bezeichnet, so dass (x_{ϵ_j}) konvergent ist gegen ein $x \in K$. Wir zeigen, dass dieses x Fixpunkt von Z ist:

$$\|x - Z(x)\| \leq \underbrace{\|x - x_{\epsilon_j}\|}_{\to 0} + \underbrace{\|x_{\epsilon_j} - Z_{\epsilon_j}(x_{\epsilon_j})\|}_{=0} + \overbrace{\| \underbrace{Z_{\epsilon_j}(x_{\epsilon_j})}_{= f_{\epsilon_j}(Z(x_{\epsilon_j}))} - Z(x_{\epsilon_j})\|}^{\leq \epsilon_j} + \underbrace{\|Z(x_{\epsilon_j}) - Z(x)\|}_{\to 0 \text{ da } Z \text{ stetig}} \overset{(j \to \infty)}{\longrightarrow} 0$$

$$\Longrightarrow x = Z(x) \qquad \square$$

Def. (kompakte Abbildung) Seien X, Y Banach-Räume. Eine (ggf. nichtlineare) Abbildung $Z : X \longrightarrow Y$ heißt kompakt, wenn Z stetig ist und für jede beschränkte Menge $M \subset X$ gilt, dass $\overline{Z(M)}$ kompakt ist.

Satz. $Z : X \longrightarrow Y$ ist genau dann kompakt, wenn gilt: Jede beschränkte Folge (x_n) in X hat eine Teilfolge (x_{n_k}), so dass $(Z(x_{n_k}))$ konvergent ist in Y. (Kurz: Unter Z werden aus beschränkten Folgen konvergente Folgen.)

Aus dem Fixpunktsatz von Schauder kann man den Fixpunktsatz von Schaefer folgern:[47]

Satz (Fixpunktsatz von Schaefer, [Schae55]) Sei $Z : X \longrightarrow X$ kompakt, und die Menge

$$M := \{x \in X \mid \exists\, \lambda \in [0,1] : x = \lambda Z(x)\}$$

sei beschränkt. Dann hat Z einen Fixpunkt.

Motivation/Anwendung: Der Satz von Schaefer erfordert, anders als der von Schauder, nicht, eine passende kompakte konvexe Menge zu identifizieren. Stattdessen ist die Kompaktheit eines Operators zu zeigen, was im Zusammenhang mit PDE-Problemen oft leicht geschehen kann unter Rückgriff auf bekannte Resultate über die kompakte Einbettung von Funktionenräumen in andere Funktionenräume.

Beweis. Sei c eine Schranke der Menge M, jedoch nicht die kleinste Schranke. Setze

$$\tilde{Z} : X \longrightarrow X, \quad \tilde{Z}(x) := \begin{cases} Z(x), & \text{falls } \|Z(x)\| \leq c \\ \frac{c\, Z(x)}{\|Z(x)\|}, & \text{falls } \|Z(x)\| > c \end{cases}$$

("Abschneiden von Z"). Es folgt $\|\tilde{Z}(x)\| \leq c \,\forall\, x \in X$, also insbesondere

$$\tilde{Z}(\overline{B(0,c)}) \subseteq \overline{B(0,c)} \quad \text{und} \quad \tilde{Z}(\tilde{Z}(\overline{B(0,c)})) \subseteq \tilde{Z}(\overline{B(0,c)}). \qquad (*)$$

[47] Der Fixpunktsatz von Schaefer kann als Spezialfall des Fixpunktsatzes von Leray-Schauder aufgefasst werden. Der letztere ist erheblich bekannter als der erstere, obwohl der erstere i.a. angenehmer in der Anwendung ist.

Wir betrachten nun die Einschränkung von $\tilde{Z}$ auf den Definitionsbereich $\tilde{Z}(\overline{B(0,c)})$: Es ist, siehe (*),

$$\tilde{Z} : \tilde{Z}(\overline{B(0,c)}) \longrightarrow \tilde{Z}(\overline{B(0,c)}).$$

Sei nun K der Abschluss der konvexen Hülle von $\tilde{Z}(\overline{B(0,c)})$. Es ist dann

$$\tilde{Z}(\overline{B(0,c)}) \subseteq K \subseteq \overline{B(0,c)},$$

wobei die erste Inklusion trivial ist, und die zweite gilt, da $\overline{B(0,c)}$ bereits konvexe abgeschlossene Menge ist, die (nach (*)) $\tilde{Z}(\overline{B(0,c)})$ enthält, und K *kleinste* konvexe abgeschlossene Menge ist, die $\tilde{Z}(\overline{B(0,c)})$ enthält. somit bekommen wir, wenn wir $\tilde{Z}$ eingeschränkt auf K betrachten,

$$\tilde{Z} : K \longrightarrow \tilde{Z}(K) \subseteq \tilde{Z}(\overline{B(0,c)}) \subseteq K.$$

Da Z eine kompakte Abbildung ist, ist auch $\tilde{Z}$ eine kompakte Abbildung. Da außerdem $\overline{B(0,c)}$ beschränkte Menge ist, ist $\overline{\tilde{Z}(\overline{B(0,c)})}$ (nach Def. komp. Abb.) kompakte Menge. Daraus kann man mittels der Definition von K als abgeschlossene konvexe Hülle folgern, dass K kompakt ist. Wir können auf dieses $\tilde{Z} : K \to K$ den Fixpunktsatz von Schauder anwenden, denn K ist konvex und kompakt, und $\tilde{Z} : K \to K$ ist stetig. Es folgt die Existenz eines Fixpunktes $x \in K$: $\tilde{Z}(x) = x$.
Es ist nun auch $Z(x) = x$. Denn angenommen, dies ist falsch. Dann muss nach Definition von $\tilde{Z}$ gelten dass $\|Z(x)\| > c$ (denn andernfalls wäre $Z(x) = \tilde{Z}(x) = x$). Es ist also $x = \tilde{Z}(x) = \frac{c\,Z(x)}{\|Z(x)\|}$. Daraus folgt einerseits $\|x\| = c$, aber andererseits $x = \lambda Z(x)$ mit $\lambda := \frac{c}{\|Z(x)\|} \in [0,1]$, also $x \in M$, also $\|x\| < c$ nach Definition der Schranke c. Widerspruch. □

Literatur

[Ada03] R. Adams, J. Fournier, *Sobolev spaces*, Elsevier Science, Oxford, (2nd ed.), 2003.

[Be96] C. Bethke, *Geochemical reaction modeling*, Oxford University Press, 1996.

[MoMaS10] J. Carrayrou, J. Hoffmann, P. Knabner, S. Kräutle, C. de Dieuleveult, J. Erhel, J. Van der Lee, V. Lagneau, K.U. Mayer, K.T.B. MacQuarrie, *Comparison of numerical methods for simulating strongly nonlinear and heterogeneous reactive transport problems-the MoMaS benchmark case*, Comp. Geosci., 14 (2010), 483–502.

[CKK10] J. Carrayrou, M. Kern, P. Knabner, *Reactive transport benchmark of MoMaS*, Comput. Geosci., 14 (2010), 385–392.

[Evans] L.C. Evans, *Partial differential equations*, American Mathematical Society, Providence, 1998.

[GilbTrud] D. Gilbarg, N. Trudinger, *Elliptic partial differential equations of second order*, Springer, 1977.

[Ha64] P. Hartman, *Ordinary differential equations*, John Wiley & Sons, 1964.

[Kr21] S. Kräutle, J. Hodai, P. Knabner, *Robust simulation of mineral precipitation-dissolution problems with variable mineral surface area*, Journal of Engineering Mathematics 129, doi:10.1007/s10665-021-10132-4, 2021.

[Kr07] S. Kräutle, P. Knabner, *A reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions*, Water Resour. Res., 43, W03429, doi:10.1029/2005WR004465, 2007.

[Habil] S. Kräutle, *General multi-species reactive transport problems in porous media: Efficient numerical approaches and existence of global solutions*, Habilitation thesis, University of Erlangen-Nuremberg, Germany, 2008.

[Lady68] O.A. Ladyženskaja, V.A. Solonnikov, N.N. Uralceva, *Linear and quasi-linear equations of parabolic type*, American Mathematical Society, 1968.

[Logan] J. Logan, *Transport modeling in hydrogeochemical systems*, Springer, 2001.

[MiSi04] M. Mincheva, D. Siegel, *Stability of mass action reaction-diffusion systems*, Nonlinear Analysis, 56 (2004), 1105–1131.

[Schae55] H. Schaefer, *Über die Methode der a priori-Schranken*, Math. Annalen, 129 (1955), 415–416.

[WYW] Z. Wu, J. Yin, C. Wang, *Elliptic and parabolic equations*, World Scientific Publishing, 2006.