

Stochastische Simulation

Michael Ziller

Zusammenfassung: Zur quantitativen Beschreibung der Prozesse in biochemischen Reaktionsnetzwerken werden topologische Modelle mit mathematischen kombiniert.

In den meisten Fällen werden zur mathematischen Charakterisierung eines Netzwerkmodells Systeme gewöhnlicher Differentialgleichungen verwendet. Die einzelnen Differentialgleichungen beschreiben kanonischerweise die zeitlichen Konzentrationsveränderungen der beteiligten Stoffe.

In dieser Arbeit wird eine stochastische Beschreibung eines Reaktionsnetzwerks als mächtigere Modellierungsstrategie vorgestellt. Nach einer Einführung in den mathematischen Formalismus wird der Gillespie-Algorithmus zur einfachen numerischen Simulation der abgeleiteten Gleichungen beschrieben.

1 Einführung

Der Fortschritt der experimentellen Techniken zur *in-vivo*-Untersuchung biochemischer Reaktionsnetzwerke hat in den letzten Jahren größere Datenmengen zur zeitlichen Entwicklung der beteiligten Stoffe verfügbar gemacht. Diese Daten gestatten mittlerweile eine relativ realistische Simulation kleinerer bis mittlerer metabolischer Netzwerke.

Das Ziel einer Simulation ist es, ein möglichst einfaches und biologisch plausibles mathematisches Modell des betrachteten Netzwerks zu generieren. Dieses Modell sollte in der Lage sein, die zur Verfügung stehenden Daten möglichst gut zu reproduzieren.

Der erste Schritt zur mathematischen Beschreibung besteht darin, eine Zuordnung zwischen den veränderlichen chemischen Entitäten des Systems und Zustandsvariablen herzustellen. Sämtliche zeitlich invarianten Größen gehen als Parameter in das Beschreibungsmodell ein.

Die mathematische Modellierung der zeitlichen Evolution der Zustandsvariablen ist durch zwei grundsätzlich verschiedene Methoden möglich:

1. Systeme gewöhnlicher Differentialgleichungen
2. Stochastische Differentialgleichungen

2 Gewöhnliche Differentialgleichungen

Der traditionelle Ansatz der chemischen Kinetik betrachtet die Stoffkonzentration der beteiligten Substanzen als Zustandsvariable und beschreibt deren zeitlichen Verlauf durch

ein System gewöhnlicher Differentialgleichungen:

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{f}(\nu(\mathbf{S}, \mathbf{p}, t), \mathbf{p}, t) \quad (1)$$

Die vektorwertige Funktion \mathbf{f} hängt direkt von den Reaktionsraten ν , den Systemparametern \mathbf{p} und ggf. von der Zeit ab. Die genaue mathematische Struktur des Gleichungssystems ist von dem Modell des zugrundeliegenden Reaktionsmechanismus abhängig. Ein häufig verwendetes Modell ist etwa das Massenwirkungsgesetz.

Komplexere Netzwerke, in denen mehr als eine Substanz an einer Reaktion beteiligt ist, führen so auf nichtlineare Systeme gewöhnlicher Differentialgleichungen, die sich meistens nicht analytisch lösen lassen.

Diese Formulierung der Substanzkonzentrationen als reelle, stetige Funktionen der Zeit basiert auf der Annahme eines deterministischen Prozesses. Obwohl diese Hypothese in vielen Fällen gerechtfertigt ist, gibt es Situationen, in denen auf mikroskopischer Ebene auftretende zufällige Effekte eine ausschlaggebende Rolle spielen.

Ein Beispiel ist ein nichtlineares System in der Nähe von Instabilitätspunkten. Die Aussagen des deterministischen Modells werden hier unzuverlässig [1].

3 Stochastische Differentialgleichungen

In diesen Fällen ist es sinnvoll, das Reaktionssystem mit stochastischen Methoden zu beschreiben. Dieser weitaus mächtigere Ansatz zur Modellierung allgemeiner physikalischer Systeme beinhaltet die deterministische Beschreibung als Spezialfall. In konkreten Anwendungen ist die stochastische Methode jedoch erheblich aufwändiger zu realisieren als die deterministische.

Die stochastische Formulierung der chemischen Kinetik betrachtet die Zustandsgrößen des Systems, z. B. die Stoffkonzentrationen, als zeitabhängige Zufallsvariablen. Reaktionen werden im Gegensatz zum deterministischen Modell nicht mithilfe von Geschwindigkeitskonstanten beschrieben, sondern durch eine (zeitlich veränderliche) Reaktionswahrscheinlichkeit (pro Zeitintervall).

Im Folgenden wird zunächst auf die mathematischen Grundlagen dieser Perspektive eingegangen.¹

3.1 Stochastische Prozesse

Ein stochastischer Prozess wird in dem hier betrachteten Kontext als eine zeitabhängige Zufallsvariable $Z(t)$ betrachtet.² Im Fall diskreter Zeit handelt es sich also um eine Folge von Zufallsvariablen: $\{Z(t_i), i = 1, \dots\}$. Die zugehörigen Wahrscheinlichkeitsdichten jeder

¹Die Ausführungen orientieren sich an [2, S. 155-175]

²Für eine mathematisch exakte Definition siehe [3]

Instanz der Zufallsvariable können dabei verschieden sein. Weiterhin sind die verschiedenen Instanzen im Allgemeinen nicht unabhängig voneinander.

Die Festlegung der folgenden Größen definiert einen stochastischen Prozess:

1. die Wahrscheinlichkeitsdichte $\rho_1(z, t)$, die die Wahrscheinlichkeit von Zustand z zum Zeitpunkt t angibt,
2. die bedingte Wahrscheinlichkeit $\rho_2(z_2, t_2|z_1, t_1)$, dass zum Zeitpunkt t_2 der Zustand z_2 vorliegt unter der Bedingung, dass zum Zeitpunkt $t_1 \leq t_2$ der Zustand z_1 realisiert war.
3. Allgemeiner:

$$\rho_n(z_n, t_n|z_{n-1}, t_{n-1}, \dots, z_1, t_1) \quad (2)$$

Es ist zu beachten, dass die Indizes ausschließlich zur übersichtlichen Darstellung diskret gewählt sind. Die Zustände z und die Zeit t können auch kontinuierliche Variablen sein. Gilt für einen speziellen stochastischen Prozess die Markov-Eigenschaft

$$\rho_n(z_n, t_n|z_{n-1}, t_{n-1}, \dots, z_1, t_1) = \rho_2(z_n, t_n|z_{n-1}, t_{n-1}), \quad (3)$$

so spricht man von einem gedächtnislosen oder Markovprozess. Hier ist die Zufallsvariable Z_n zum Zeitpunkt t_n unabhängig von der früheren Geschichte Z_{n-2}, \dots, Z_1 und hängt ausschließlich von der Realisierung Z_{n-1} ab.

Als Konsequenz der Markov-Eigenschaft gilt:

$$\rho_3(z_3, t_3; z_2, t_2; z_1, t_1) = \rho_3(z_3, t_3|z_2, t_2; z_1, t_1)\rho_2(z_2, t_2|z_1, t_1) \quad (4)$$

$$= \rho_2(z_3, t_3|z_2, t_2)\rho_2(z_2, t_2|z_1, t_1)\rho_1(z_1, t_1) \quad (5)$$

also:

$$\rho_2(z_3, t_3; z_1, t_1) = \int \rho_2(z_3, t_3; z_2, t_2; z_1, t_1) dz_2 \quad (6)$$

$$= \int \rho_2(z_3, t_3|z_2, t_2)\rho_2(z_2, t_2|z_1, t_1)\rho_1(z_1, t_1) dz_2 \quad (7)$$

woraus direkt die Chapman-Kolmogorov Gleichung folgt:

$$\rho_2(z_3, t_3|z_1, t_1) = \int \rho_2(z_3, t_3|z_2, t_2)\rho_2(z_2, t_2|z_1, t_1) dz_2 \quad (8)$$

Für $\rho_1(z_3, t_3)$ gilt mit einer analogen Argumentation:

$$\rho_1(z_3, t_3) = \int \rho_2(z_3, t_3|z_2, t_2)\rho_1(z_2, t_2) dz_2 \quad (9)$$

Es ist also möglich, alle Wahrscheinlichkeitsdichten ρ_3, \dots, ρ_n durch eine initiale Übergangswahrscheinlichkeit $\rho_2(z, t|z_0, t_0) = \rho_1(z, t)$ auszudrücken. Hängt diese bedingte Wahrscheinlichkeit ausschließlich von der zeitlichen Differenz $t - t'$ ab, so spricht man von einem zeitlich homogenen Markovprozess 1. Ordnung.

3.2 Die Master-Equation

Zur Untersuchung des zeitlichen Verlaufs der Wahrscheinlichkeitsdichte betrachtet man die Taylor-Entwicklung der Übergangswahrscheinlichkeit nach τ :

$$\rho_2(z, t + \tau | z', t) = (1 - a(z', t)\tau)\delta(z - z') + \tau\omega(z, z', t) + O(\tau^2) \quad (10)$$

³ Mit der Nebenbedingung wegen Normierung:

$$a(z', t) = \int \omega(z, z', t) dz \quad (11)$$

Die Gleichung (10) besagt anschaulich, dass die Zufallsvariable $Z(t)$ für jeden ausreichend kleinen Zeitschritt τ den aktuellen Zustand mit einer gewissen Wahrscheinlichkeit beibehält oder genau mit der Gegenwahrscheinlichkeit in einen anderen Zustand übergeht. Mit der Chapman-Kolmogorov Gleichung (8) folgt:

$$\begin{aligned} \rho_2(z, t + \tau | z', t') &= \int \rho_2(z, t + \tau | z'', t)\rho_2(z'', t | z', t') dz'' & (12) \\ &= (1 - a(z, t)\tau)\rho_2(z, t | z', t') \\ &+ \tau \int w(z, z'', t)\rho_2(z'', t | z', t') dz'' + O(\tau^2) \end{aligned} \quad (13)$$

Mit $\tau \rightarrow 0$ und Berücksichtigung von (11) folgt schließlich die Master-Equation:

$$\begin{aligned} \frac{\partial \rho_2(z, t | z', t')}{\partial t} &= \int w(z, z'', t)\rho_2(z'', t | z', t') dz'' \\ &- \int w(z'', z, t)\rho_2(z, t | z', t') dz'' \end{aligned} \quad (14)$$

Multiplikation von (14) mit $\rho_1(z', t')$ und Integration über z' liefert für $\rho_1(z, t)$:

$$\frac{\partial \rho_1(z, t)}{\partial t} = \int w(z, z', t)\rho_1(z', t) dz' - \int w(z', z, t)\rho_1(z, t) dz' \quad (15)$$

Diese Gleichungen beschreiben die zeitliche Evolution der Zustandswahrscheinlichkeitsdichtefunktion. Im Fall diskreter Zustände läßt sich die Master-Equation für Zustand n schreiben als:

$$\frac{d\rho_n(t)}{dt} = \sum_m \left(w_{nm}(t)\rho_m(t) - w_{mn}(t)\rho_n(t) \right) \quad (16)$$

Bei einem zeitlich homogenen Prozesses hängen die $w(z, z', t)$ nicht von der Zeit ab. Konkret beschreibt diese Gleichung den netto *Wahrscheinlichkeitsfluss* aus dem bzw. in den Zustand n .

³ $\delta(z - z')$ bezeichnet hier die Delta-Distribution.

3.3 Stochastische Beschreibung eines deterministischen Prozesses

Deterministische Prozesse, die durch gewöhnliche Differentialgleichungen beschrieben werden, lassen sich im Rahmen der stochastischen Theorie als Spezialfall ansehen. Betrachtet man die gewöhnliche Differentialgleichung:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \quad (17)$$

mit $x : \mathbb{R} \rightarrow \mathbb{R}^n$, dann lässt sich für diesen Prozess die Übergangswahrscheinlichkeit angeben als:

$$\rho_2(\mathbf{z}, t | \mathbf{x}_0, t_0) = \delta(\mathbf{z} - \mathbf{x}(t)) \quad (18)$$

mit $\mathbf{x}(t)$ der Lösung von Gleichung (17) zu den Anfangsbedingungen $\mathbf{x}(t_0) = \mathbf{x}_0$. Offensichtlich handelt es sich bei ρ_2 um eine Wahrscheinlichkeitsdichte. Für das Kurzzeitverhalten der Gleichung folgt:

$$\rho_2(\mathbf{z}, t + \tau | \mathbf{z}', t) = \delta\left(\mathbf{z} - (\mathbf{z}' + \mathbf{f}(\mathbf{z}')\tau)\right) \quad (19)$$

$$= \delta(\mathbf{z} - \mathbf{z}') - \mathbf{f}(\mathbf{z}') \cdot \frac{\partial}{\partial \mathbf{z}} \delta(\mathbf{z} - \mathbf{z}')\tau + O(\tau^2). \quad (20)$$

Mit

$$w(\mathbf{z}, \mathbf{z}') = -\mathbf{f}(\mathbf{z}') \frac{\partial}{\partial \mathbf{z}} \delta(\mathbf{z} - \mathbf{z}') \quad (21)$$

gilt also:

$$a(\mathbf{z}') = \int d^n z w(\mathbf{z}, \mathbf{z}') = - \int d^n z \mathbf{f}(\mathbf{z}') \frac{\partial}{\partial \mathbf{z}} \delta(\mathbf{z} - \mathbf{z}') = 0. \quad (22)$$

Aus Gleichung (15) ergibt sich dann für die Master-Equation:

$$\frac{\partial \rho_1(\mathbf{z}, t)}{\partial t} = \int d^n z' \left(-\mathbf{f}(\mathbf{z}') \frac{\partial}{\partial \mathbf{z}} \delta(\mathbf{z} - \mathbf{z}') \right) \rho_1(\mathbf{z}', t) \quad (23)$$

$$= -\frac{\partial}{\partial \mathbf{z}} \cdot (\mathbf{f}(\mathbf{z}) \rho_1(\mathbf{z}, t)) \quad (24)$$

Diese als *Liouville-Gleichung* bekannte Differentialgleichung beschreibt die zeitliche Entwicklung einer Wahrscheinlichkeitsdichte, die auf einem deterministischen Prozess basiert. Bezeichnet $\rho(p, q)$ eine Dichte im Orts- und Impulsraum, so lässt sich die *Liouville-Gleichung* mit der Hamilton-Funktion $H(p, q)$ schreiben als:

$$\frac{\partial}{\partial t} \rho(p, q; t) = \{H, \rho\} \quad (25)$$

$\{H, \rho\}$ bezeichnet hier die Poisson-Klammer.

4 Stochastische Simulation von chemischen Reaktionsnetzwerken.

In diesem Abschnitt werden zunächst die physikalischen Grundannahmen dargelegt, die eine stochastische Formulierung der chemischen Reaktionskinetik gestatten. Basierend auf diesen Voraussetzungen wird im Anschluss daran die mathematische Beschreibung im Rahmen der Theorie der stochastischen Prozesse diskutiert.

Der Abschnitt schließt mit einer Diskussion des Gillespie-Algorithmus.

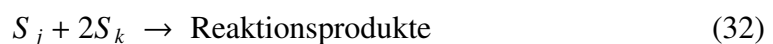
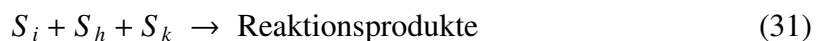
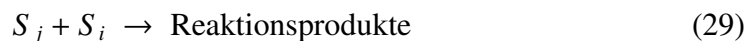
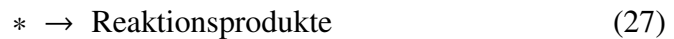
Das gesamte Kapitel orientiert sich an der Originalarbeit von Gillespie [1].

Das allgemeine Problem, dessen Lösung in den folgenden Abschnitten diskutiert wird, lässt sich wie folgt formulieren:

Ein beliebiges Volumen V enthalte N chemisch aktive Substanzen S_i ($i = 1, \dots, N$) und eine beliebige Anzahl inaktiver Moleküle. Es sei

$$X_i \equiv \text{Momentane Teilchenzahl der Substanz } S_i \quad (26)$$

Diese Substanzen können an M verschiedenen Reaktionen teilnehmen R_μ ($\mu = 1, \dots, M$). Die R_μ bezeichnet man als Reaktionskanäle. Zur Vereinfachung seien nur die nachfolgenden Reaktionstypen zugelassen (die Indizes sind als paarweise verschieden zu betrachten):



Diese Reaktionen werden in Analogie zur deterministischen Formulierung durch den Reaktionsparameter c_μ charakterisiert.

Die fundamentale Hypothese, die der stochastischen Formulierung der chemischen Kinetik zugrunde liegt, ist die Annahme, dass der Reaktionsparameter c_μ definiert werden kann als:

Fundamentale Hypothese 1

$c_\mu \delta t \equiv$ mittlere Wahrscheinlichkeit zur ersten Ordnung in δt , dass eine spezifische Kombination von R_μ Molekülen im nächsten Zeitintervall δt gemäß R_μ reagieren wird.

4.1 Physikalische Grundlagen

Als Ausgangspunkt der physikalischen Analyse einer chemischen Reaktion bietet sich die bimolekulare Reaktion (29) an:



In erster Näherung werden die S_i Moleküle als nicht deformierbare Spähren mit Durchmesser d_i und Masse m_i angenommen (idealiserter Fall). Der Abstand der beiden Spähren sei mit d_{12} und die Relativgeschwindigkeit von S_1 zu S_2 mit v_{12} bezeichnet. Eine Kollision der beiden Moleküle liegt vor, falls für d_{12} gilt:

$$d_{12} \leq \frac{d_1 + d_2}{2} \quad (35)$$

Im Zeitintervall δt durchwandert das Teilchen S_1 relativ zu S_2 das Kollisionsvolumen:

$$\delta V_{coll} = \pi d_{12}^2 v_{12} \delta t \quad (36)$$

Befindet sich das Zentrum von S_2 zu diesem Zeitpunkt innerhalb von δV_{coll} , so liegt eine Kollision vor. Um einen Ausdruck für die Kollisionswahrscheinlichkeit angeben zu können, ist eine weitere grundlegende Hypothese notwendig:

Annahme 1 (Räumliche Homogenität)

Es liegt eine zufällige Gleichverteilung der betrachteten Partikeltypen in V vor.

Daraus lässt sich dann direkt für die Wahrscheinlichkeit P_{coll} (S_2 liegt in δV_{coll}) folgern:

$$P_{coll} = \frac{\delta V_{coll}}{V} \quad (37)$$

Für die mittlere Wahrscheinlichkeit einer $S_1 - S_2$ -Kollision ergibt sich dann:

$$\langle P_{coll} \rangle = \left\langle \frac{\delta V_{coll}}{V} \right\rangle = \frac{\pi d_{12}^2}{V} \langle v_{12} \rangle \delta t \quad (38)$$

Annahme 2 (Thermisches Gleichgewicht)

Die Verteilung der kinetischen Energie der Moleküle ist Maxwell-Boltzmann verteilt.

Mit der Annahme 2 lässt sich für die Kollisionswahrscheinlichkeit der geschlossene Ausdruck

$$\left\langle \frac{\delta V_{coll}}{V} \right\rangle = V^{-1} \pi d_{12}^2 \sqrt{\frac{8k_B T}{\pi m_{12}}} \delta t \quad (39)$$

angeben. Dabei ist $m_{12} = m_1 m_2 / (m_1 + m_2)$ die reduzierte Masse. Es ist zu beachten, dass nach Gleichung (39) jede 1 – 2-Kollision zu einer R_μ -Reaktion führen würde. Da die meisten chemischen Reaktionen eine Aktivierungsenergie u_μ^* benötigen, sollen nur Teilchen

mit einer kinetischen Energie, die mindestens u_μ^* entspricht, gemäß Gleichung (38) reagieren dürfen. Mit dieser Modifikation ergibt sich für die Reaktionswahrscheinlichkeit:

$$c_\mu = V^{-1} \pi d_{12}^2 \sqrt{\frac{8k_B T}{\pi m_{12}}} \delta t e^{-\frac{u_\mu^*}{k_B T}} \quad (40)$$

Um die Gültigkeit der bisher gemachten Annahmen über die Zeit zu garantieren, ist es aus Konsistenzüberlegungen erforderlich, eine weitere Forderung an das System zu stellen:

Annahme 3

Elastische Stöße finden viel häufiger statt als inelastische Stöße.

In vielen natürlichen Systemen kann diese Annahme als gegeben betrachtet werden. Unter diesen Bedingungen findet eine kontinuierliche, zufällige Redistribution der Moleküle im Reaktionsvolumen V statt. Dieser Prozess garantiert die Gültigkeit von Annahme 1.

Weiterhin erfolgt durch die große Zahl elastischer Stöße eine Restauration der Maxwell-Boltzmann-Geschwindigkeitsverteilung, deren oberes Ende durch Reaktionen stetig abgeschnitten wird.

Um eine Beziehung zwischen der deterministischen und stochastischen Beschreibung eines Reaktionssystems zu erhalten, wird im Folgenden auf die Beziehung zwischen dem Reaktionsparameter c_μ und der im gewöhnlichen DGL-Formalismus verwendeten Geschwindigkeitskonstanten r_μ eingegangen.

Sei $X_1(X_2)$ die Anzahl von $S_1(S_2)$ Molekülen in V zu einem beliebigen Zeitpunkt, dann gibt es $X_1 X_2$ verschiedene Kombinationen reaktiver Moleküle, und es gilt für die Wahrscheinlichkeit der Reaktion R_μ im Zeitintervall dt :

$$p_\mu = X_1 X_2 c_\mu dt \quad (41)$$

und für die mittlere Reaktionsrate:

$$\langle p_\mu \rangle = \langle X_1 X_2 \rangle c_\mu \quad (42)$$

Man beachte, dass $\langle \rangle$ hier für das Mittel über ein Ensemble stochastisch identischer Systeme steht. Für das System muss also die Quasiergodenhypothese gelten:

Jede Trajektorie, abgesehen von einer Menge vom Maß Null, kommt jedem Punkt des Phasenraums beliebig nahe[5, S. 505]. Es folgt daraus, dass die Zeitmittelung gleich der Ensemblemittelung ist. Für Funktionen im Phasenraum gilt dann:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(q(t), p(t)) dt = \frac{1}{\Omega(E)} \int f(q, p) d\Gamma \quad (43)$$

Die Einführung von molekularen Konzentrationen $x_i \equiv X_i/V$, $\langle x_1 x_2 \rangle V c_\mu$ erlaubt es, den Zusammenhang von c_μ mit der Reaktionskonstanten k_μ anzugeben:

$$k_\mu = \frac{\langle x_1 x_2 \rangle V c_\mu}{\langle x_1 \rangle \langle x_2 \rangle} \quad (44)$$

Für eine bimolekulare Reaktion entspricht dies der üblichen Definition von k_μ :

$$\frac{dx}{dt} = k_\mu \cdot [S_i] [S_j] \quad (45)$$

An dieser Stelle lässt sich ein deutlicher Unterschied zwischen den verschiedenen Formulierungen der chemischen Kinetik beobachten.

Die deterministische Beschreibung unterscheidet nicht zwischen dem Mittelwert des Produkts der Konzentrationsvariablen und dem Produkt der Mittelwerte, d. h. es gilt $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$. Für $i \neq j$ werden damit die Effekte von Korrelationen und für $i = j$ die Effekte von Fluktuationen vernachlässigt.

Da Korrelationen nach Kurt [4] im thermodynamischen Grenzwert (Anzahl der Moleküle sowie Reaktionsvolumen konvergiert gegen unendlich) verschwinden, ist diese implizite Annahme im deterministischen Formalismus in vielen Fällen erfüllt. Für nichtlineare Systeme in der Nähe chemischer Instabilitätspunkte haben Korrelations- und Fluktuationphänomene hingegen einen großen Einfluss auf das makroskopische Verhalten des Systems.

Da nur im thermodynamischen Grenzwert die deterministische Formulierung der chemischen Kinetik über die *Liouville-Gleichung* (vgl. Abschnitt 3.3) aus der stochastischen Beschreibung herleitbar ist, stellt sich die Frage nach weiteren Kriterien für die Anwendbarkeit der stochastischen Formulierung. Gillespie [1, S. 409] gibt hier als hinreichende Bedingung an, dass erheblich mehr nichtreaktiver Molekülstöße gegenüber reaktiven Stößen vorliegen müssen (vgl. Annahme 3).

Da der gewöhnliche deterministische Ansatz ebenfalls eine gleichmäßige Verteilung der Moleküle voraussetzt, ist es unter Bedingungen, die mit der Annahme 3 nicht vereinbar sind, erforderlich, einen Beschreibungsansatz näher an der *Liouville-Gleichung* zu wählen. Dieser kann sowohl stochastischer als auch deterministischer Natur sein wie etwa die Beschreibung mittels partieller Differentialgleichungen.

4.2 Mathematischer Formalismus

Die traditionelle stochastische Herangehensweise an das Problem des gekoppelten chemischen Reaktionssystems betrachtet, basierend auf den physikalischen Überlegungen aus

dem vorigen Abschnitt, die Master-Equation für die Wahrscheinlichkeitsdichtefunktion:

$$\rho(X_1, \dots, X_N; t) \equiv \text{Wahrscheinlichkeit, im Volumen } V \text{ zum Zeitpunkt } t \\ X_1 \text{ Moleküle der Sorte } S_1, \dots, X_N \text{ Moleküle der Sorte } S_N \\ \text{zu finden}$$

Zur Lösung einer allgemeinen Master-Equation ist die Kenntnis der verschiedenen Momente von ρ erforderlich:

$$X_i^{(k)}(t) \equiv \sum_{X_1=0}^{\infty} \dots \sum_{X_N=0}^{\infty} X_i^k \rho(X_1, \dots, X_N; t) \quad (46)$$

Physikalisch repräsentiert $X_i^{(k)}(t)$ die mittlere Anzahl von Molekülen des Typs S_i zum Zeitpunkt t im Volumen V zur k -ten Potenz. Von besonderer Bedeutung sind die ersten und zweiten Momente. Mit diesen Größen lässt sich die erwartete Abweichung (Varianz) von der mittleren Zahl der Moleküle des Typs S_i zum Zeitpunkt t ermitteln:

$$\Delta_i(t) \equiv \sqrt{X_i^{(2)}(t) - (X_i^{(1)}(t))^2} \quad (47)$$

Zum Zeitpunkt t ist es also wahrscheinlich,

$$y_i \in [X_i^{(1)}(t) - \Delta_i(t), X_i^{(1)}(t) + \Delta_i(t)] \quad (48)$$

S_i -Moleküle vorzufinden.

Wie im Abschnitt 3.2 und im Anhang bereits ausgeführt, ist nur in den einfachsten Fällen eine analytische Lösung der *Master-Equation* möglich. Eine direkte numerische Lösung führt ebenfalls lediglich in den wenigsten Fällen zum Erfolg.

Es ist Gillespies Verdienst, eine zur Master-Equation äquivalente stochastische Beschreibung der chemischen Kinetik in Verbindung mit einer einfachen numerischen Lösungsmethode angegeben zu haben.

Ähnlich wie die Beschreibung der Kinetik durch die Master-Equation basiert auch Gillespies Ansatz auf der Fundamental-Hypothese (1)

4.3 Gillespie-Ansatz

Ausgangspunkt der Analyse ist im Gegensatz zur Wahrscheinlichkeitsdichte ρ des traditionellen Ansatzes die neu eingeführte Reaktionswahrscheinlichkeitsdichtefunktion $P(\tau, \mu)$:

$$P(\tau, \mu)d\tau \equiv \text{Wahrscheinlichkeit, dass zum Zeitpunkt } t \text{ im nächsten differentiellen} \\ \text{Zeitintervall } [t + \tau, t + \tau + d\tau] \text{ eine Reaktion stattfinden} \\ \text{und diese vom Typ } R_\mu \text{ sein wird.}$$

Offensichtlich handelt es sich bei $P(\mu, \tau)$ um eine Verbundwahrscheinlichkeitsdichte auf $\mathbb{R} \times \Omega$ mit $\Omega = \{1, 2, \dots, M\}$. Es seien weiterhin die Zustandsvariablen h_μ der M Reaktionskanäle definiert durch:

$$h_\mu \equiv \text{Anzahl verschiedener Molekülkombinationen} \\ \text{für Reaktion } R_\mu \text{ zum Zeitpunkt } t \text{ in } V$$

Aus (1) folgt dann:

$$h_\mu c_\mu \delta t = \text{Wahrscheinlichkeit, zur ersten Ordnung in } \delta t, \\ \text{dass im nächsten Zeitintervall } \delta t \text{ eine Reaktion des Typs } R_\mu \text{ stattfindet.}$$

Allgemein ergibt sich für die h_μ der einzelnen Reaktionstypen:

1. $h_\mu = 1$ für Reaktionen vom Typ (27)
2. $h_\mu = X_j$ für Reaktionen vom Typ (28)
3. $h_\mu = X_j X_k$ für Reaktionen vom Typ (29)
4. $h_\mu = X_j(X_j - 1)/2$ für Reaktionen vom Typ (30)
5. $h_\mu = X_j X_i X_k$ für Reaktionen vom Typ (31)
6. $h_\mu = X_j X_k(X_k - 1)/2$ für Reaktionen vom Typ (32)
7. $h_\mu = X_j(X_j - 1)(X_j - 2)/6$ für Reaktionen vom Typ (33)

$P(\tau, \mu)$ berechnet sich nun aus der Wahrscheinlichkeit $P_0(\tau)$ zum Zeitpunkt t , keine Reaktion im nächsten Zeitintervall zu beobachten, und aus der Wahrscheinlichkeit für das Auftreten der Reaktion R_μ :

$$P(\tau, \mu) d\tau = P_0(\tau) \cdot h_\mu c_\mu d\tau \quad (49)$$

$P_0(\tau)$ lässt sich mit der folgenden Überlegung aus der Binomialverteilung berechnen: Es sei das Intervall $[t, t+\tau]$ in K Teilintervalle der Länge $\epsilon = \tau/K$ zerlegt. Die Wahrscheinlichkeit, keine der R_1, \dots, R_M Reaktionen im ersten Teilintervall $[t, t + \epsilon]$ vorzufinden, ist dann gegeben durch:

$$\prod_{v=1}^M \left(1 - h_v c_v \epsilon + o(\epsilon)\right) = 1 - \sum_v h_v c_v \epsilon + o(\epsilon) \quad (50)$$

Die R_μ -Ereignisse sind hier als unabhängig voneinander angenommen.
Für K derartige Teilintervalle gilt also:

$$P_0(\tau) = \left(1 - \sum_v^M h_v c_v \epsilon + o(\epsilon)\right)^K \quad (51)$$

$$= \left(1 - \sum_v^M \frac{h_v c_v \tau}{K} + o(K^{-1})\right)^K \quad (52)$$

Damit folgt im Grenzwert für $K \rightarrow \infty$:

$$P_0(\tau) = \lim_{K \rightarrow \infty} \left(1 - \sum_v^M h_v c_v \tau + o(K^{-1})\right)^K = \exp\left(-\sum_{v=1}^M h_v c_v \tau\right) \quad (53)$$

Einsetzen in Gleichung (49) ergibt schließlich den exakten Ausdruck für die Reaktionswahrscheinlichkeitsdichtefunktion:

$$P(\tau, \mu) = h_\mu c_\mu \exp\left(-\sum_{v=1}^M h_v c_v \tau\right) \quad (54)$$

Es ist zu bemerken, dass die Dichtefunktion auf ihrem Definitionsbereich bereits korrekt normiert ist.

Die Wahrscheinlichkeit, zum Zeitpunkt t im nächsten differentiellen Zeitintervall $[t + \tau, t + \tau + d\tau]$ eine Reaktion vom Typ R_μ zu beobachten, ist also sowohl von den Reaktionsparametern aller M Reaktionskanäle als auch von der Anzahl aller reaktiven Moleküle zum aktuellen Zeitpunkt abhängig.

5 Der Gillespie-Algorithmus

Die zentrale Idee einer stochastischen Simulation der chemischen Kinetik ist es, den durch die Wahrscheinlichkeitsverteilung $(P(\tau, \mu))$ beschriebenen stochastischen Prozess mittels geeigneter Monte-Carlo-Methoden zu approximieren. Der dazu von Gillespie angegebene Algorithmus operiert wie folgt:

1. Setze $t = 0$.
 Setze die X_1, \dots, X_N auf die Startwerte der Molekülzahl der verschiedenen Substanzen S_i .
 Setze c_1, \dots, c_M auf die Reaktionsparameter der entsprechenden Reaktion $\{R_\mu\}$.
 Berechne die $h_1 c_1, \dots, h_M c_M$ gemäß Gleichungen (27 bis 33).
 $P(\tau, \mu)$ ist nun eindeutig festgelegt.
 Generiere Stützstellen $t_1 < \dots < t_{\text{Stop}}$.

2. Ermittle unter Verwendung einer Monte-Carlo-Methode ein Paar (τ, μ) gemäß der Verteilung $P(\tau, \mu)$
3. Setze $t := t + \tau$ und modifiziere die Menge $\Gamma = \{X_i\}$ der an Reaktion R_μ beteiligten Stoffe entsprechend R_μ .
Berechne die neuen $h_\nu, c_\nu \forall \nu \in \{i \mid X_i \in \Gamma\}$.
4. Falls $t > t_i$, speichere alle X_1, \dots, X_N . Für $t < t_{\text{stop}}$ und $\exists i : X_i \neq 0$ gehe zu 2, terminiere andernfalls.

Eine einfache Ausführung des Algorithmus liefert lediglich eine mögliche Realisation des stochastischen Prozesses $P(\tau, \mu)$. Um ein statistisch vollständiges Bild der zeitlichen Entwicklung des Prozesses zu bekommen, ist es erforderlich, unendlich viele Realisationen mit den gleichen Anfangsbedingungen zu betrachten.

Setzt man

$$X_i(k, t) \equiv \text{Anzahl der Moleküle vom Typ } S_i \text{ im } k\text{-ten Lauf zur Zeit } t \quad (55)$$

so gilt im Fall einer endlichen Zahl K von Ausführungen des Algorithmus für die erwartete Anzahl von S_i Molekülen zum Zeitpunkt t

$$X_i^{(1)} \simeq \frac{1}{K} \sum_{k=1}^K X_i(k, t) \quad (56)$$

sowie für die mittlere Schwankung:

$$\Delta_i(t) \simeq \left(\frac{1}{K} \sum_{k=1}^K (X_i(k, t))^2 - \left(\frac{1}{K} \sum_{k=1}^K X_i(k, t) \right)^2 \right)^{\frac{1}{2}} \quad (57)$$

Ist $\Delta_i(t) \ll X_i^{(1)}$, so schwankt $X_i(k, t)$ in den verschiedenen Realisationen des stochastischen Prozesses nicht sehr stark, und es reicht aus ein relativ kleines K zu betrachten. Liegen starke Schwankungen vor, so sind entsprechend größere K zu wählen. Lässt sich keine ausreichend starke Konvergenz feststellen, ist das Schwankungsintervall der $X_i(k, t)$ für die verschiedenen Realisationen von Interesse.

5.1 Der Monte-Carlo-Schritt

Der obige Algorithmus gestattet es nun, eine vollständige Simulation eines chemischen Reaktionsnetzwerks auszuführen. Die einzige bisher offen gelassene Frage ist die der Implementierung des Monte-Carlo-Schritts zur Generierung des (τ, μ) -Tupels aus $P(\tau, \mu)$. Nachfolgend wird hier die analytisch stringente Direkte Methode angegeben. Es ist an dieser Stelle jedoch zu beachten, dass zahlreiche weitere Verfahren zur Bestimmung des

Tupels (τ, μ) aus $P(\tau, \mu)$ existieren.

Als Ausgangspunkt verwendet diese Methode eine alternative Darstellung der Dichtefunktion $P(\tau, \mu)$:

$$P(\tau, \mu) = P_1(\tau) P_2(\mu|\tau) \quad (58)$$

$P_1(\tau)d\tau$ bezeichnet hier die Wahrscheinlichkeit, im nächsten Zeitintervall $[t + \tau, t + \tau + d\tau]$ eine beliebige Reaktion vorzufinden. $P_2(\mu|\tau)$ ist die Wahrscheinlichkeit, dass die nächste Reaktion vom Typ R_μ sein wird – unter der Bedingung, dass diese Reaktion zum Zeitpunkt $t + \tau$ stattfindet.

$P_1(\tau)$ ergibt sich durch Summation über alle Reaktionsmöglichkeiten:

$$P_1(\tau) = \sum_{\mu=1}^M P(\tau, \mu) \quad (59)$$

Einsetzen in (58) liefert für $P_2(\mu|\tau)$:

$$P_2(\mu|\tau) = \frac{P(\tau, \mu)}{\sum_{\nu=1}^M P(\tau, \nu)} \quad (60)$$

Explizit gilt dann mit Gleichung (54):

$$P_1(\tau) = a e^{-a\tau} \quad (0 \leq \tau \leq \infty) \quad (61)$$

$$P_2(\tau|\mu) = \frac{a_\mu}{a} \quad (\mu = 1, \dots, M) \quad (62)$$

Dabei wurden die Abkürzungen

$$a_\mu \equiv h_\mu c_\mu \quad (63)$$

$$a \equiv \sum_{\mu=1}^M a_\mu = \sum_{\mu=1}^M h_\mu c_\mu \quad (64)$$

verwendet. Für beide Dichtefunktionen muss die Normalisierungsbedingung gelten:

$$\int_0^\infty P_1(\tau) d\tau = \int_0^\infty a e^{-a\tau} d\tau = 1 \quad (65)$$

$$\sum_{\mu=1}^M P_2(\mu|\tau) = \sum_{\mu=1}^M \frac{a_\mu}{a} = 1 \quad (66)$$

Steht ein Zufallszahlengenerator für das Einheitsintervall zur Verfügung, so kann $P(\tau, \mu)$ aus den beiden Dichtefunktionen berechnet werden.

Seien $r_1, r_2 \in [0, 1]$ gleichverteilte Zufallszahlen aus dem Einheitsintervall, dann lässt

sich τ gemäß $P(\tau)$ mittels der *Inversionsmethode* [1, Anhang A] aus r_1 bestimmen. Es gilt:

$$\tau = \frac{1}{a} \cdot \ln\left(\frac{1}{r_1}\right) \quad (67)$$

Analog lässt sich μ als ganzzahliger Wert bestimmen, für den gilt:

$$\sum_{\nu=1}^{\mu-1} a_{\nu} \leq r_2 a \leq \sum_{\nu=1}^{\mu} a_{\nu} \quad (68)$$

Diese Operationen liefern nach Gleichung (58) ein gemäß $P(\tau, \mu)$ verteiltes Tupel, das im Schritt 2 des Gillespie-Algorithmus verwendet werden kann.

6 Zusammenfassung

Der deterministische Modellierungsansatz gewöhnlicher Differentialgleichungen lässt sich mittels der *Liouville-Gleichung* als Spezialfall der stochastischen Beschreibung des gleichen Reaktionssystems auffassen.

Da es bisher noch nicht gelungen ist, den stochastischen Formalismus aus der deterministischen Beschreibung herzuleiten, kann der stochastische Ansatz in diesem Sinne als mächtiger betrachtet werden.

Spielen stochastische Effekte in einem Reaktionsnetzwerk eine untergeordnete Rolle, was in vielen Fällen erfüllt zu sein scheint, so ist eine mathematische Charakterisierung durch gewöhnliche Differentialgleichungen völlig ausreichend.

Sind diese Bedingungen nicht erfüllt, ist die Aussagekraft eines deterministischen Modells von begrenztem Wert.

Weist das untersuchte System etwa mehrere Steady-States auf, so ist eine zusätzliche Verwendung des stochastischen Formalismus angeraten.

In jedem Fall ist es sinnvoll, vor Beginn einer Modellierung eines beliebigen Systems zunächst eine genaue Analyse der stochastischen Systemelemente vorzunehmen.

Da es sich als unpraktikabel herausgestellt hat, chemische Reaktionssysteme durch eine Lösung der *Master-Equation* zu beschreiben, ist es sinnvoll, die von Gillespie erarbeitete äquivalente Charakterisierung durch die Reaktionsratendichtefunktion zu verwenden. Trotz zahlreicher, nur unter Idealbedingungen erfüllter Annahmen besticht der Formalismus durch seine leichte Anwendbarkeit sowie das simple, durch den Gillespie-Algorithmus gegebene numerische Lösungsverfahren.

In komplexeren Reaktionsnetzwerken, beispielsweise aus der Biochemie, ist eine Anpassung der dem stochastischen Formalismus zugrundeliegenden Annahmen an die Gegebenheiten des biochemischen Systems vorzunehmen. So ist die Hartschalen-Hypothese für Moleküle im Fall von Proteinen sicherlich nicht gerechtfertigt. Auch räumliche Homogenität kann in der Zelle nur bedingt vorausgesetzt werden.

Unter biologisch realistischen Bedingungen sind demnach zahlreiche Modifikationen des mathematischen Formalismus erforderlich.

Literatur

- [1] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.
- [2] Josef Honerkamp. *Statistical Physics*. Springer, 2002.
- [3] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, 1992.
- [4] T.G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *Journal of Chemical Physics*, 57:2976, 1972.
- [5] Franz Schwabl. *Statistische Mechanik*. Springer, 2002.