

2.5 Dynamische Modelle neuronaler Netze

Wenn auch die neuronal codierte Informationen durch die Feuerrate nicht vollständig beschrieben wird, so können doch viele grundlegende neuronale Mechanismen mit dem Modell des Feuerraten-Neuron dargestellt werden. Es wird zunächst das aktivitätsbeschreibende Modell (??) betrachtet, bei dem alle Neurone die gleiche Sigmoidfunktion Θ_S dargestellt werden und die gleiche Zeitkonstante τ besitzen:

$$\frac{dv_i}{dt} = \frac{1}{\tau} \left(-v_i + \Theta_S \left[\sum_j w_{ij} v_j \right] \right)$$

Dabei wird jedes Neuron als ein eindimensionales nicht-lineares dynamisches System beschrieben. Ein neuronales Netz von n Neuronen entspricht dann einem n -dimensionalen nichtlinearen System. Auch für dieses System sind die stationären Zustände

$$\left. \frac{dv_i}{dt} \right|_{\mathbf{v}^*} = 0 \quad \forall i = 1, \dots, n \quad (2.78)$$

von besonderem Interesse, weil sie mögliche Zustände darstellen, in die sich das System auf Dauer begeben kann.

2.5.1 Systeme von zwei Neuronen

Die Dynamik eines gekoppelten Systems von zwei Neuronen kann in drei verschiedene Klassen unterteilt werden, welche durch die Wechselwirkung zwischen beiden Neuronen bestimmt sind.

Wechselseitige Erregung

Im Falle wechselseitiger Erregung

$$\begin{aligned} \frac{dv_1}{dt} &= \frac{1}{\tau} (-v_1 + \Theta_S(v_2)) \\ \frac{dv_2}{dt} &= \frac{1}{\tau} (-v_2 + \Theta_S(v_1)) \end{aligned} \quad (2.79)$$

existieren bei entsprechender Wahl von $\Theta_S(s)$ drei stationäre Zustände. Diese können im (für das vorliegende System zweidimensionalen) Phasenraum durch die Schnittpunkte der beiden *Nullklinen*

$$\begin{aligned} \frac{dv_1}{dt} &= 0 \\ \frac{dv_2}{dt} &= 0 \end{aligned} \quad (2.80)$$

wie in Abbildung 31 graphisch dargestellt werden.

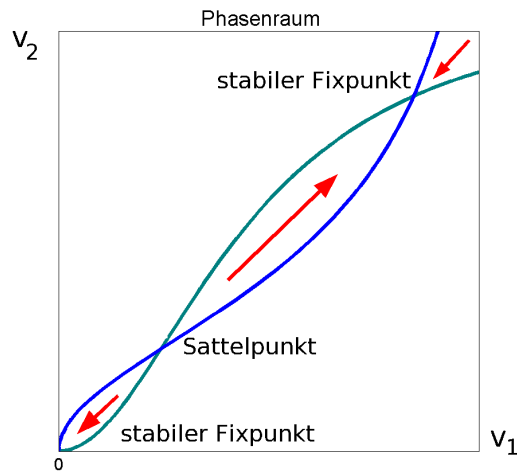


Abbildung 31: *Bistabilität im zweidimensionalen Phasenraum. Es sind die Nullkline für zwei sich wechselseitig erregende Neurone v_1 und v_2 dargestellt. Die Schnittpunkte der Nullklinen sind die stationären Zustände, von denen zwei stabil sind und der mittlere ein Sattelpunkt ist.*

Da in dem System zwei voneinander isolierte stabile Fixpunkte existieren, ist das System bistabil. Die durch die Sigmoidfunktion gegebene Nichtlinearität kann durch allmähliche Vergrößerung des Kontrollparameters Δ_s verringert werden. Im linearen Grenzfall werden die Nullklinen zu Geraden. Wenn diese nicht übereinander liegen, so hat eine sogenannte Pitchfork-Bifurkation stattgefunden, bei der nur noch ein stabiler Fixpunkt übrig bleibt. Bistabilität erfordert die Nichtlinearität des Systems.

Kennzeichnend für das System wechselseitiger Erregung ist, daß im stabilen Zustand entweder beide Neurone aktiv oder beide Neurone inaktiv sind.

Wechselseitige Hemmung

In einem System wechselseitiger Hemmung

$$\begin{aligned} \frac{dv_1}{dt} &= \frac{1}{\tau} (-v_1 + \Theta_S(-v_2 + k)) \\ \frac{dv_2}{dt} &= \frac{1}{\tau} (-v_2 + \Theta_S(-v_1 + k)) \end{aligned} \quad (2.81)$$

ist ein externer Stimulus k das einzig erregende Moment. Auch dieses System kann bei entsprechender Wahl von Θ_S drei stationäre Zustände besitzen, von denen zwei stabil sind. Diese sind bei wechselseitiger Hemmung jedoch dadurch gekennzeichnet, daß in jedem der beiden stabilen Zustände jeweils nur ein Neuron aktiv und das andere vollkommen inaktiv ist. (Abbildung 32, links). Man spricht daher auch von einer

Winner-Takes-All-Dynamik (WTA-Dynamik) oder einer exklusiven Entscheidungsdynamik.

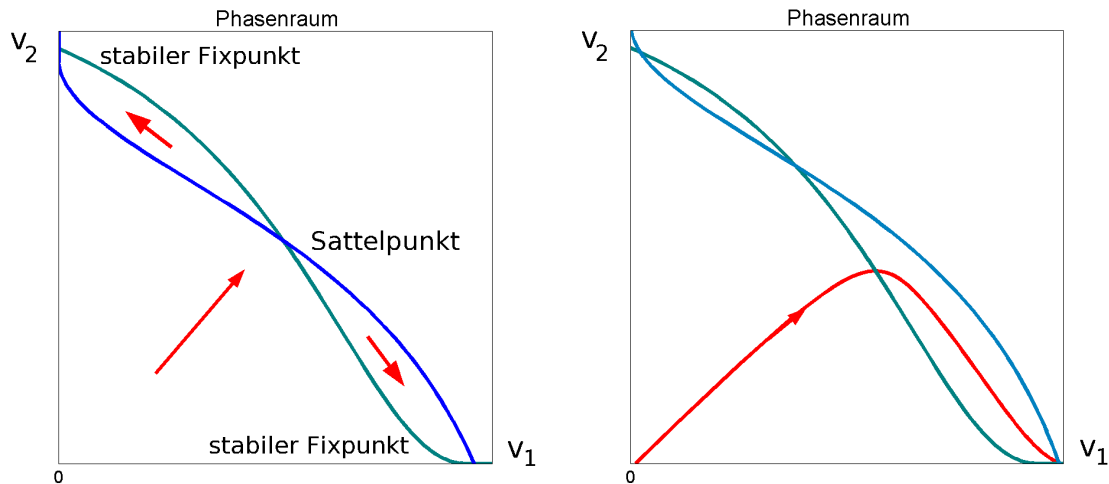


Abbildung 32: *Exklusive Entscheidungsdynamik.* Links sind die Nulllinien für die sich wechselseitig hemmenden Neurone v_1 und v_2 von (2.81) dargestellt. Im rechten Bild ist v_1 begünstigt, wodurch bei indifferentem Ausgangszustand eine Entscheidung zugunsten von v_1 fällt. Die bei $(0,0)$ beginnende Kurve zeigt die Zustandsabfolge (Trajektorie) bei einem Ausgangszustand, in dem beide Neurone inaktiv sind. Die Trajektorie beginnt im Ursprung und mündet in den stabilen Fixpunkt rechts unten.

Ist im initialen Zustand die Aktivität beider Neurone gleich Null, so würde sich in dem idealisierten System (2.81) der mittlere stationäre Zustand im zeitlichen Verlauf einstellen, in dem beide Neurone die gleiche gemäßigte Aktivität zeigen. Dieses System ist jedoch unrealistisch, da natürliche Systeme niemals rauschfrei sind, sodaß dort schon durch geringfügige Fluktuationen eine nicht-deterministische Entscheidung zugunsten einer der beiden stabilen Zustände herbeigeführt wird. In dem durch zwei verschiedene externe Stimuli k_1 und k_2 modifizierten System

$$\begin{aligned} \frac{dv_1}{dt} &= \frac{1}{\tau} (-v_1 + \Theta_S(-v_2 + k_1)) \\ \frac{dv_2}{dt} &= \frac{1}{\tau} (-v_2 + \Theta_S(-v_1 + k_2)) \end{aligned} \quad (2.82)$$

kann durch die Wahl des Vorzeichens von $(k_1 - k_2)$ einer der beiden stabilen Fixpunkte von außen begünstigt werden (Abbildung 32, rechts). Das System trifft stets eine Entscheidung zugunsten des Neurons mit dem stärkeren Stimulus.

Kombinierte Hemmung und Erregung

Ein kombiniertes System, bestehend aus einem exzitatorischen Neuron $v_1 = E$ und einem inhibitorischen Neuron $v_2 = I$, ist im einfachsten Falle durch

$$\begin{aligned}\frac{dE}{dt} &= \frac{1}{\tau_E} (-E + \Theta_E(-I)) \\ \frac{dI}{dt} &= \frac{1}{\tau_I} (-I + \Theta_I(E))\end{aligned}\tag{2.83}$$

gegeben. Da erregende und hemmende Neurone nicht vom gleichen Typ sind, wird formal zwischen $\Theta_E(x)$ und $\Theta_I(x)$ bzw. τ_E und τ_I unterschieden. Dieses System hat bei entsprechender Wahl der Sigmoidfunktionen $\Theta_E(x)$ und $\Theta_I(x)$ einen stabilen Fixpunkt, in dem E und I gleichermaßen gemäßigt aktiv sind. Jedes Neuron hat auf das jeweils andere einen regulierenden Effekt. Solche Selbstregulation wird in organischen Systemen auch als Homöostase bezeichnet.

Bei der synaptischen Übertragung werden einzelne Aktionspotentiale verstärkt oder abgeschwächt. Dieser Einfluß auf den postsynaptischen Stimulus kann in dem vereinfachten Modell des Feuerraten-Neurons durch eine Gewichtung der präsynaptischen Feuerraten beschrieben werden (Anhang ??):

$$\begin{aligned}\frac{dE}{dt} &= \frac{1}{\tau_E} (-E + \Theta_E(w_{IE} \cdot I)) \\ \frac{dI}{dt} &= \frac{1}{\tau_I} (-I + \Theta_I(w_{EI} \cdot E))\end{aligned}\tag{2.84}$$

Hemmende Synapsen entsprechen hierbei einem negativen Gewichtswert $w_{IE} < 0$ und erregende Synapsen einem positiven Gewichtswert $w_{EI} > 0$.

In Systemen, die aus hemmenden und erregenden Neuronen bestehen, können neben Fixpunkten auch sogenannte *Grenzzyklen* im Phasenraum entstehen. Ein Grenzzyklus ist eine isolierte, geschlossene Zustandsabfolge (*Trajektorie*). Isoliert bedeutet, daß Trajektorien der Umgebung nicht geschlossen sind, sondern entweder spiralförmig auf den Grenzzyklus zu (*stabiler Grenzzyklus*) oder von ihm weg laufen (*instabiler Grenzzyklus*). Daneben gibt es auch *halbstabile Grenzzyklen*, die in ihrer inneren Umgebung ein anderes Anziehungsverhalten auf Trajektorien zeigen, als in ihrer äußeren Umgebung. Stabile Grenzzyklen sind, wie auch die stabilen Fixpunkte, *Attraktoren*, da alle Trajektorien in einer hinreichend kleinen Umgebung der stabilen Grenzzyklen auf diese zu laufen (Abbildung 33).

Die Sequenz von Aktionspotentialen des stimulierten Neurons im Hodgkin-Huxley-Modell entspricht einem Grenzzyklus. Ein einfacheres Beispiel für einen Grenzzyklus stellt der sogenannte Buhmann-Oszillator [Mal00] dar: Ein exzitatorisches Neuron E und ein

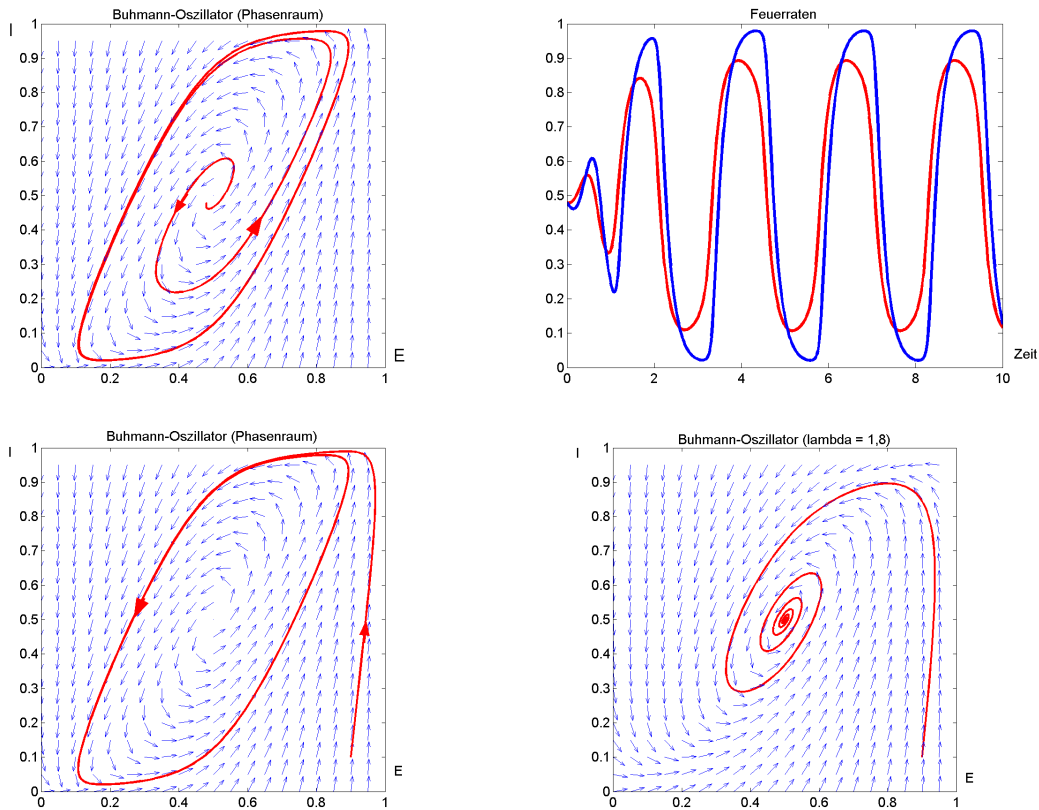


Abbildung 33: Grenzyklus beim Buhmann-Oscillator. Links ist der Phasenraum bei $\lambda = 3$ dargestellt. Die Trajektorien haben oben und unten verschiedene Anfangsbedingungen. Rechts oben sieht man den periodischen Zeitverlauf bei $\lambda = 3$ und der Anfangsbedingung $(0, 48; 0, 48)$. Rechts unten ist für $\lambda = 1, 8$ kein Grenzyklus mehr vorhanden, der Fixpunkt an der Stelle $(0, 5; 0, 5)$ ist hier stabil.

inhibitorisches Neuron I sind synaptisch gekoppelt, wobei das exzitatorische Neuron selbsterregend wirkt⁴¹:

$$\begin{aligned} \tau_E \frac{dE}{dt} &= -E(x) + \Theta_E \left(E - w \cdot I - \frac{1-w}{2} \right) \\ \tau_I \frac{dI}{dt} &= -I(x) + \Theta_I \left(E - \frac{1}{2} \right) \end{aligned} \quad (2.85)$$

⁴¹In einem biologisch realistischeren Modell müsste die Selbsterregung durch eine wechselseitige Erregung mit weiteren exzitatorischen Neuronen beschrieben werden [Wil99].

Die Zeitkonstanten mögen hierbei zu $\tau_E = \tau_I = 1$ gegeben sein und die Sigmoidfunktionen seien mit $\Theta_{max} = 1$, $x_{1/2} = 0$ und $\Delta_x = \frac{1}{4\gamma}$ jeweils:

$$\Theta(x) = \frac{\Theta_{max}}{1 + \exp\left(\frac{x_{1/2} - x}{\Delta_x}\right)} = \frac{1}{1 + \exp(-4\gamma x)} \quad (2.86)$$

Ein Fixpunkt befindet sich dann bei $E^* = I^* = \frac{1}{2}$, wie man durch Einsetzen in (2.85) überprüfen kann.

Zur *Stabilitätsanalyse* des Fixpunktes muß das zeitliche Verhalten kleiner Auslenkungen $e = E - E^*$ und $i = I - I^*$ vom Fixpunkt betrachtet werden. Bei hinreichend kleinen Auslenkungen ist deren Linearisierung um den Fixpunkt gerechtfertigt. Dazu wird (2.85) zunächst in einen linearen und einen nicht-linearen Anteil zerlegt

$$\begin{aligned} \frac{dE}{dt} &= -E(x) + \Theta\left(E - w \cdot I - \frac{1-w}{2}\right) = -E(x) + \Theta^E \\ \frac{dI(x)}{dt} &= -I(x) + \Theta\left(E - \frac{1}{2}\right) = -I(x) + \Theta^I \end{aligned} \quad (2.87)$$

und eine Linearisierung des durch die Sigmoidfunktion gegebenen, nicht-linearen Anteils vorgenommen:

$$\begin{aligned} \Theta^E(E^* + e, I^* + i) &\approx \Theta^E(E^*, I^*) + \left.\frac{\partial \Theta^E}{\partial E}\right|_{E^*, I^*} \cdot e + \left.\frac{\partial \Theta^E}{\partial I}\right|_{E^*, I^*} \cdot i \\ \Theta^I(E^* + e, I^* + i) &\approx \Theta^I(E^*, I^*) + \left.\frac{\partial \Theta^I}{\partial E}\right|_{E^*, I^*} \cdot e + \left.\frac{\partial \Theta^I}{\partial I}\right|_{E^*, I^*} \cdot i \end{aligned}$$

Die Ableitung der Sigmoidfunktion (2.86) nach x kann wie folgt geschrieben werden:

$$\frac{d}{dx}\Theta(x) = \frac{4\gamma}{(\exp(2\gamma x) + \exp(-2\gamma x))^2}$$

Demnach ist $\frac{d}{dx}\Theta(0) = \gamma$, sodaß die linearisierten Auslenkungen durch

$$\begin{aligned} \frac{d}{dt}e &= -e + \gamma e - \gamma w i \\ \frac{d}{dt}i &= -i + \gamma e \end{aligned}$$

beschrieben werden und die Jakobi-Matrix für dieses System

$$A = \begin{pmatrix} -1 + \gamma & -\gamma w \\ \gamma & -1 \end{pmatrix}$$

lautet. Aus der charakteristischen Gleichung $\det(A - \lambda) = 0$ ergeben sich die beiden Eigenwerte

$$\lambda_{1,2} = -1 + \frac{\gamma}{2} \pm \gamma \sqrt{\frac{1}{4} - w}. \quad (2.88)$$

Für $\gamma < 2$ fällt die Auslenkung mit der Zeit exponentiell in den Fixpunkt zurück und der Fixpunkt $E^* = I^* = \frac{1}{2}$ ist demnach stabil. Ist $w > \frac{1}{4}$, so werden die Eigenwerte komplex und die Lösung des linearisierten Systems bekommt einen oszillatorischen Anteil:

$$\begin{pmatrix} e \\ i \end{pmatrix} (t) = \begin{pmatrix} e \\ i \end{pmatrix} (0) \cdot \underbrace{\exp\left(\left(\frac{\gamma}{2} - 1\right)t\right)}_{\text{stabilitätsbestimmend}} \cdot \underbrace{\exp\left(j\left(\gamma\sqrt{\left|\frac{1}{4} - w\right|}\right)t\right)}_{\text{oszillatorisch}} \quad (2.89)$$

Hierbei ist $j^2 = -1$. Die Bedingung für das Entstehen von Grenzzyklen ist das Anwachsen der Oszillation, was von dem stabilitätsbestimmenden Faktor abhängt. So ergeben sich Grenzzyklen erst für Werte von $\gamma > 2$. Ein solcher ist in Abbildung 33 illustriert und stellt als stabiler Grenzzyklus einen Attraktor dar. Wenn der Systemparameter γ als *Bifurkationsparameter* fungiert, tritt an der Stelle $\gamma = 2$ eine sogenannte *Hopf-Bifurkation* mit der von Null verschiedenen Kreisfrequenz $\gamma \left|\frac{1}{4} - w\right|^{1/2}$ auf. Wie in Abbildung 30 veranschaulicht, wird die Nicht-Linearität der Sigmoidfunktion mit fallendem Δ_x , d.h. mit wachsendem γ , verstärkt. Das verdeutlicht, daß Nicht-Linearität eine Voraussetzung für Grenzzyklen darstellt. Zwar kann auch ein lineares System $\frac{d}{dt}\mathbf{x} = \mathbf{A}\mathbf{x}$ geschlossene Bahnen besitzen, diese können jedoch niemals isoliert sein, da für jede periodische Lösung $\mathbf{x}(t)$ gilt, daß auch $c\mathbf{x}(t)$ eine Lösung ist. Die Amplitude *linearer Oszillationen* ist stets durch die Anfangsbedingungen festgelegt und alle Änderungen dieser Amplitude bleiben bestehen.

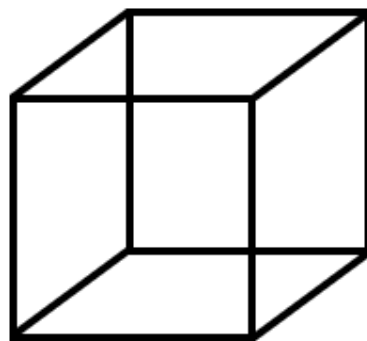


Abbildung 34: *Der Necker-Würfel. Ein mögliches Beispiel für einen Grenzzyklus bei der bewußten Wahrnehmung. Der Würfel erscheint abwechselnd aus der Perspektive von rechts oben oder von links unten.*

2.5.2 Kortikale Netzwerke

Eine aus mehreren Neuronen bestehende eindimensionale kortikale Schicht mit räumlicher Struktur kann in Anlehnung an (2.84) durch das vereinfachte Wilson-Cowan-Modell [Wil99] beschrieben werden:

$$\begin{aligned}\tau_E \frac{dE(x)}{dt} &= -E(x) + \Theta_E \left(\sum_{x'} w_{EE}(x, x') \cdot E(x') - \sum_{x'} w_{IE}(x, x') \cdot I(x') + P(x) \right) \\ \tau_I \frac{dI(x)}{dt} &= -I(x) + \Theta_I \left(\sum_{x'} w_{EI}(x, x') \cdot E(x') - \sum_{x'} w_{II}(x, x') \cdot I(x') + Q(x) \right)\end{aligned}\quad (2.90)$$

Diese Formulierung entspricht dem aktivitätsbeschreibenden Netzwerkmodell (2.77) für zwei verschiedene Neuronentypen, mit der Besonderheit, daß die Menge der Indizes zusätzlich eine Topologie besitzt. Hierbei bezeichnen $E(x)$ ein exzitatorisches und $I(x)$ ein inhibitorisches Neuron an der Stelle x , $P(x)$ und $Q(x)$ jeweils einen äußeren Stimulus auf diese Neurone. Die synaptische Kopplung zwischen einem Neuron vom Typ j an der Stelle x und einem Neuron vom Typ i an der Stelle x' wird durch $w_{ij}(x, x')$ beschrieben. Kortikale Schichten werden im allgemeinen als *homogen* ($w_{ij}(x, x') = w_{ij}(x - x')$) und *isotrop* ($w_{ij}(x) = w_{ij}(-x)$) modelliert, sodaß die synaptischen Kopplungen w_{ij} zweier Neurone an den Orten x und x' lediglich von deren räumlichen Abstand $|x - x'|$ abhängen und deshalb $w_{ij}(x, x') = w_{ij}(|x - x'|)$ ist. Die von Null verschiedenen Gewichte w_{IE} und w_{II} haben ihrem inhibitorischen Charakter entsprechend für alle $|x - x'|$ negative Werte, im Gegensatz dazu haben von Null verschiedene w_{EI} und w_{EE} grundsätzlich positive Werte. Damit entspricht (2.90) dem *Dale'schen Prinzip*, daß kortikale Neuronen entweder ausschließlich hemmend oder ausschließlich erregend sind. Zur analytischen Vereinfachung werden die neuronalen Gewebeschichten des Kortex vielfach mathematisch als kontinuierlich formuliert und die Summen durch Integrale ersetzt:

$$\begin{aligned}\tau_E \frac{dE(x)}{dt} &= -E(x) + \Theta_E \left[\int w_{EE}(|x - x'|) \cdot E(x') dx' \right. \\ &\quad \left. - \int w_{IE}(|x - x'|) \cdot I(x') dx' + P(x) \right] \\ \tau_I \frac{dI(x)}{dt} &= -I(x) + \Theta_I \left[\int w_{EI}(|x - x'|) \cdot E(x') dx' \right. \\ &\quad \left. - \int w_{II}(|x - x'|) \cdot I(x') dx' + Q(x) \right]\end{aligned}\quad (2.91)$$

Im Rahmen dieser Darstellung spricht man von zwei *neuronalen Feldern*, weil die Zustandsvariablen eine kontinuierliche Indexmenge mit einer Topologie besitzen. Diese Vereinfachung wird durch die hohe Neuronendichte im Kortex und die Annahme gerechtfertigt, daß die Neurone gleichmäßig verteilt sind. Die $w_{ij}(x, x')$ müssen beim Übergang von der Summendarstellung zu Integralen mit der Dichtefunktion $\rho(x)$ an der Stelle x und x' multipliziert werden. Diese Summe von N Dirac-Distributionen

$\rho(x) = \sum_{n=1}^N \delta(x - x_n)$, wobei x_n den Ort des n -ten Neurons bezeichnet, wird unter der Annahme einer gleichmäßigen und kontinuierlichen Verteilung zu $\rho(x) = 1$. Damit kann die neuronale Aktivität durch eine oder wenige *Integro-Differentialgleichungen* beschrieben und die Wechselwirkung durch Funktionen $w_{ij} : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^\pm$ dargestellt werden. In Abhängigkeit von diesen Funktionen können neben den durch wechselseitige Erregung bedingten homogenen Lösungen globaler Aktivität oder globaler Inaktivität auch verschiedene stabile Muster heterogener Aktivitätsverteilung infolge lateraler Hemmung (Abbildung 35) entstehen. Musterbildende Wechselwirkungen sind, wie bereits bei der Turing-Instabilität zu erkennen war, durch langreichweitige Hemmung und kurzreichweitige Erregung gekennzeichnet. Die lokale wechselseitige Erregung kann dazu führen, daß durch zeitlich beschränkte externe Stimuli hervorgerufene Aktivitätsmuster nach deren Wegfall bestehen bleiben. Zeitliche Verzögerungen ermöglichen sogar eine Zunahme der lokalen Erregung nach Wegfall des Stimulus. Ebenso, wie bei den Systemen zweier Neurone, können aber auch zyklische Attraktoren auftreten, wenn beispielsweise stationäre Lösungen, wie in (2.89), in linearer Näherung oszillatorische Instabilitäten aufweisen, denen nicht-lineare Komponenten entgegenwirken. In besonderen Fällen entstehen auch laufende Wellen [WC73].

Läßt man die Unterscheidung zwischen exzitatorischen und inhibitorischen Neuronen formal fallen, so vereinfacht sich (2.90) zu

$$\tau \frac{dv(x)}{dt} = -v(x) + \Theta_S \left(\sum_{x'} w(x, x') \cdot v(x') + \theta(x) \right). \quad (2.92)$$

Hemmung und Erregung werden durch das Vorzeichen von $w(x, x')$ festgelegt, Θ_S und die Zeitkonstante τ sind jedoch für alle Neurone identisch. In gleicher Weise vereinfacht sich (2.91) zu:

$$\tau \frac{dv(x)}{dt} = -v(x) + \Theta_S \left(\int w(x, x') \cdot v(x') dx' + \theta(x) \right). \quad (2.93)$$

Der bisher formulierten neuronalen Dynamik liegt das *aktivitätsbeschreibende Modell* (2.77) zugrunde, bei welchem die gewichtete Summe präsynaptischer Aktivität im Argument der Sigmoidfunktion steht. Im *spannungsbeschreibenden Modell* (2.76) haben die Zustandsvariablen nicht den Charakter von Feuerraten, sondern sind, wie in Anhang des vorangegangenen Abschnitts erläutert, durch die gemittelten Membranpotentiale $u(x)$ der Neurone an der Stelle x gegeben⁴². Die Feuerrate wird in diesem Modell durch

$$Z(x) = \Theta_S[u(x)]$$

⁴²Der wesentliche biophysikalische Unterschied beider Modelle (Anhang ??) liegt darin, das jene die Zeitkonstante der Zellmembran $\tau_m = RC$, diese die Zeitverzögerung bei der synaptischen Übertragung als vernachlässigbar annehmen [DA01] [Er98].

beschrieben. Das Membranpotential eines Neurons an der Stelle x ergibt sich dann aus der gewichteten Summe präsynaptischer Feuerraten $\Theta_S[u(x')]$ und einem externen Stimulus $I(x)$ gemäß:

$$\tau \frac{du(x)}{dt} = -u(x) + \sum_{x' \neq x} w(x, x') \cdot \Theta_S[u(x')] + I(x) \quad (2.94)$$

Dieses Modell ist eine spezielle Variante des Cohen-Grossberg-Modells [CG83] und wird oftmals als *Hopfield-Modell* bezeichnet, obgleich letzteres in seinem ursprünglichen Zusammenhang eine diskrete Indexmenge besitzt und erstmals zeitdiskret formuliert wurde [Ho82]. In der kontinuierlichen Darstellung als neuronales Feld lautet (2.94):

$$\tau \frac{du(x)}{dt} = -u(x) + \int w(x, x') \cdot \Theta_S[u(x')] dx' + I(x) \quad (2.95)$$

Modelle kortikaler Netzwerke zeigen das gleiche dynamische Verhalten unabhängig davon, ob ihnen ein Neuronenmodell wie in (2.92) oder wie in (2.94) zugrundeliegt. Im Rahmen beider Beschreibungsvarianten sind, abhängig von der Wechselwirkung $w(x, x')$, globale Aktivität, globale Inaktivität oder diverse stabile Muster lokaler Aktivität zu beobachten. Die stationären Zustände des Systems (2.94) und deren Stabilitätseigenschaften können, wie in [QP03] gezeigt, auf das System (2.92) abgebildet werden. Grundlegende dynamische Unterschiede werden jedoch durch eine Veränderung der Wechselwirkung erzielt. Beim Übergang von Systemen mit verschiedenen Neuronentypen, wie das durch (2.90) beschriebene, zu Systemen mit einem einzelnen Neuronentyp, wie in (2.92) oder (2.94), ist die Wechselwirkung entsprechend anzupassen. Erstgenannte werden oftmals als mehrschichtige Modelle, letztere als einschichtige Modelle bezeichnet. Geht man bei einem einschichtigen Modell von einer *symmetrischen Wechselwirkung*

$$w(x, x') = w(x', x) \quad (2.96)$$

aus, so hat dies wesentliche Auswirkungen auf die dynamische Charakteristik. Nach dem *Cohen-Grossberg-Theorem* [CG83] werden damit Grenzzyklen und Zentren ausgeschlossen, da ein solches System ausschließlich asymptotisch stabile oder nicht stabile Fixpunkte besitzen kann (Anhang ??). Diese Systemeigenschaft macht symmetrische Wechselwirkungen stets attraktiv, wenn ein stabiles System zu konstruieren ist und Oszillationen oder chaotisches Verhalten ausgeschlossen werden soll. Symmetrische Wechselwirkungen werden daher auch in einschichtigen Netzwerkmodellen häufig verwendet, obgleich sie in solchen gegen das Dale'sche Prinzip verstoßen, sobald sowohl hemmende als auch erregende Neurone existieren: Ist ein exzitatorisches Neuron $E(x)$ mit einem inhibitorischen Neuron $I(x')$ gekoppelt, so müssen $w(x, x')$ und $w(x', x)$ unterschiedliche Vorzeichen haben, was im Widerspruch zu (2.96) steht. In den durch (2.79) und (2.81) beschriebenen Systemen zweier Neurone liegt eine symmetrische Wechselwirkung vor. Wie Ermentrout in [Er95] zeigt, erlauben Systeme zweier Neurone bei wechselseitiger

Erregung oder wechselseitiger Hemmung generell keine anderen Attraktoren als Fixpunkte, auch wenn die Wechselwirkung, abgesehen vom Vorzeichen, nicht symmetrisch ist.

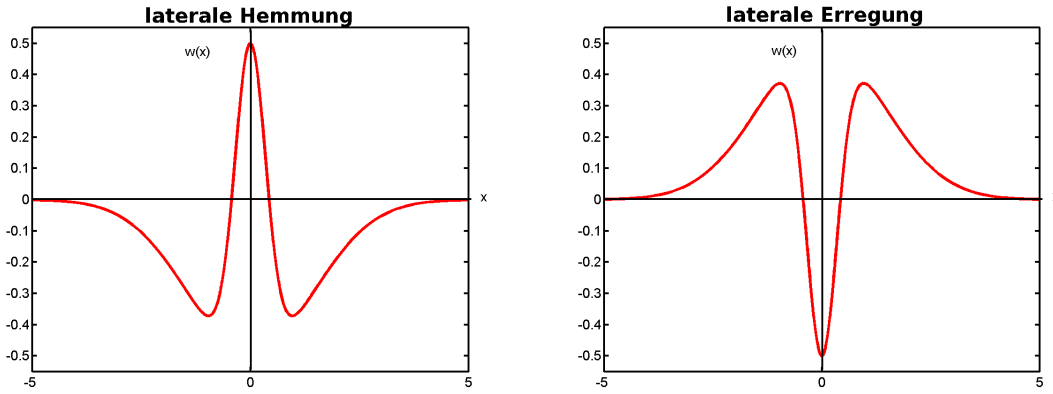


Abbildung 35: *Symmetrische Wechselwirkungsfunktionen mit lateraler Hemmung (links) und lateraler Erregung (rechts) durch die Differenz zweier Gaußfunktionen beschrieben.*

2.5.3 Neuronale Felder mit symmetrischer Wechselwirkung

Eine Stabilitätsanalyse eindimensionaler kortikaler Schichten mit symmetrischer Wechselwirkung ist in dem Modell von Amari relativ einfach [Ama77]. In diesem werden die neuronalen Gewebeschichten des Kortex mathematisch wie in (2.95) als *neuronale Felder* beschrieben, welche Erregungsmuster ausbilden und weiterleiten können. Der externe Stimulus wird um eine konstante globale Hemmung $-r$ ergänzt, welche dem negativen Ruhepotential der Neurone entspricht. Zerlegt man den externen Stimulus gemäß $I(x, t) = \bar{s} + s(x, t)$ in einen räumlich und zeitlich konstanten mittleren Stimulus \bar{s} und eine zeitabhängige, lokale Abweichung $s(x, t)$ davon, so kann eine globale und zeitunabhängige Konstante $h = \bar{s} - r$ definiert werden. Damit lautet das dynamische System für eine sich aus nur einem einzigen Neuronentyp konstituierende, eindimensionale kortikale Schicht

$$\tau \frac{\partial u(x, t)}{\partial t} = -u(x, t) + \int w(x, x') \cdot \Theta_H[u(x', t)] dx' + h + s(x, t) \quad (2.97)$$

Diese Darstellung beschreibt ein eindimensionales neuronales Feld. Das Feld wird in Amaris Modell als homogen und isotrop modelliert, weshalb für die Wechselwirkungsfunktion auch dort $w(x, x') = w(|x - x'|)$ gilt. Darüberhinaus hat diese einen lateral inhibitorischen Charakter (Abbildung 35, links), welcher beispielsweise durch die Differenz zweier *Gaußfunktionen* (A.25)

$$w(x) = a_1 \cdot e^{-(x/\sigma_1)^2} - a_2 \cdot e^{-(x/\sigma_2)^2} \quad (2.98)$$

dargestellt werden kann. Dazu muß die negativ-wertige Gaußfunktion breiter und flacher sein, d.h.

$$\begin{aligned} \sigma_1 &< \sigma_1 \\ a_1 &> a_2 \end{aligned} \tag{2.99}$$

für $a_1, a_2 > 0$ (Abbildung 41). Der Einfachheit halber wird $\Theta_H(s)$ als Stufenfunktion angenommen.

Stationäre Lösungen

Die stationären Lösungen $\frac{\partial u(x,t)}{\partial t} = 0$ von (2.97) sind in Abwesenheit äußerer Stimuli $s(x,t) = 0$ durch

$$u(x) = \int w(|x - x'|) \cdot \Theta_H[u(x)] dx' + h$$

gegeben und können in 3 verschiedene Klassen eingeteilt werden:

- Homogene Lösungen: \emptyset -Lösung (globale Inaktivität) oder ∞ -Lösung (globale Aktivität)
- a -Lösung (erregtes Intervall der Länge a)
- (a, b) -Lösung (mehrere erregte Intervalle der Länge a mit der Periode b)

Definiert man

$$W(x) = \int_0^x w(x') dx'$$

und W_m als das Maximum von W , sowie $W_\infty = W(\infty)$, so ist das Verhalten des Systems allein von den Relationen zwischen W_m , W_∞ und h abhängig: Wie in [Ama77] erläutert, sind die Lösungen des Systems durch 3 Fälle klassifizierbar.

$2W_\infty > W_m$:

$$\begin{aligned} h < -2W_\infty &\Rightarrow \{\emptyset\} \\ -2W_\infty < h < -W_m &\Rightarrow \{\emptyset, \infty\} \\ -W_m < h < -W_\infty &\Rightarrow \{\emptyset, a_1, a_2, \infty\} \\ -W_\infty < h < 0 &\Rightarrow \{\emptyset, a, \infty\} \\ 0 < h &\Rightarrow \{\infty\} \end{aligned}$$

$W_m > 2W_\infty > 0$:

$$\begin{aligned} h < -W_m &\Rightarrow \{\emptyset\} \\ -W_m < h < -W_\infty &\Rightarrow \{\emptyset, a_1, a_2\} \\ -2W_\infty < h < -W_\infty &\Rightarrow \{\emptyset, a_1, a_2, \infty\} \end{aligned}$$

$$-W_m < h < 0 \Rightarrow \{\emptyset, a, \infty\}$$

$$0 < h \Rightarrow \{\infty\}$$

$W_\infty < 0$:

$$h < -W_m \Rightarrow \{\emptyset\}$$

$$-W_m < h < 0 \Rightarrow \{\emptyset, a_1, a_2\}$$

$$0 < h < -2W_\infty \Rightarrow \{(a, b)\}$$

$$-2W_\infty < h \Rightarrow \{\infty\}$$

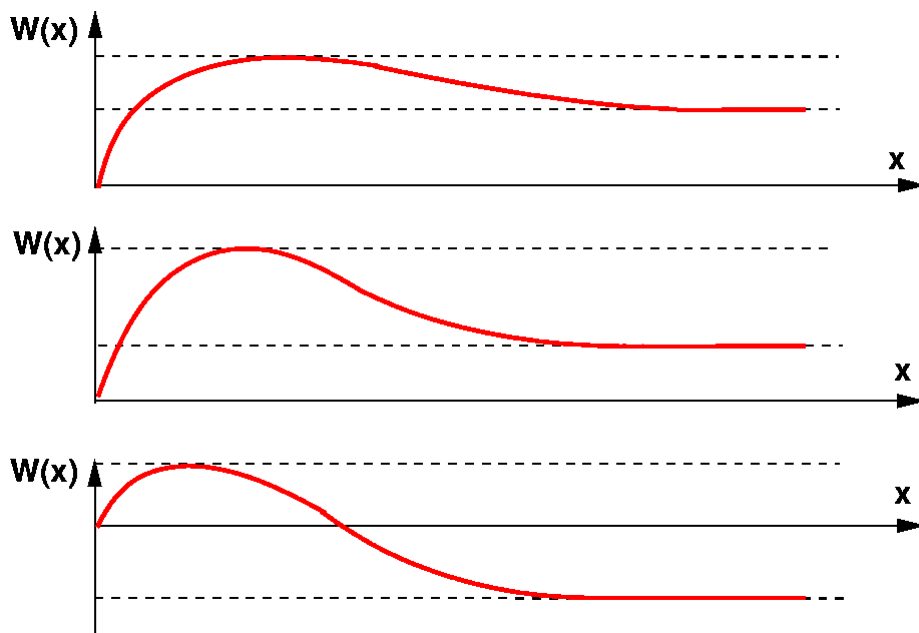


Abbildung 36: Drei mögliche Fälle in Abhängigkeit von $W(x)$: $2W_\infty > W_m$ (oben), $W_m > 2W_\infty > 0$ (Mitte) und $W_\infty < 0$ (unten). Die obere gestrichelte Linie kennzeichnet in allen drei Graphen W_m , die untere jeweils W_∞ .

Wegen der Homogenität der Wechselwirkung ist, wenn $u(x)$ eine stationäre Lösung ist, ebenso $u(x - a)$ eine stationäre Lösung. Ohne Einschränkung der Allgemeinheit kann daher bei einer a -Lösung angenommen werden, daß die Erregung im Intervall von 0 bis a lokalisiert ist. Für eine a -Lösung gilt dann

$$W(a) + h = 0 \tag{2.100}$$

und diese ist genau dann stabil (Anhang 2.5.3), wenn $\frac{dW}{da} < 0$ oder

$$w(a) < 0. \tag{2.101}$$

Die Lösungen $\{\emptyset, a, \infty\}$ sind bistabil, sodaß initial erregte Intervalle sich zu einer globalen Aktivität ausdehnen, falls ihre Ausdehnung größer als a ist, und zur globalen Inaktivität schrumpfen, falls ihre Ausdehnung kleiner als a ist. Ebenso sind die Lösungen $\{\emptyset, a_1, a_2, \infty\}$ bistabil, sodaß unter der Annahme $a_1 < a_2$ eine initiale Erregung mit einer Länge kleiner als a_1 zu einer \emptyset -Lösung schwindet und eine initiale Erregung mit einer Länge größer als a_1 zu einer a_2 -Lösung konvergiert.

Reaktion auf äußere Stimuli

Ein äußerer Stimulus $s(x)$ wird, wenn er im Vergleich zur wechselseitigen Hemmung und Erregung der Neurone groß ist, die Lösung der Gleichung (2.97) völlig dominieren. Der Effekt relativ schwacher Stimuli $\epsilon s(x)$, mit einem kleinen $\epsilon > 0$, bewirkt eine räumliche Verschiebung des angeregten Intervalls in Richtung des stärksten Stimulus $\max_{x \in \mathbb{R}} s(x)$, ähnlich wie bei der durch (2.82) beschriebenen *Entscheidungsdynamik* zweier Neurone. An der Stelle des stärksten Stimulus bleibt die Aktivität infolge lokaler Selbsterregung bei einer stabilen a -Lösung auch nach dem Wegfallen äußerer Stimuli bestehen, was vielfach als Modell für ein *Kurzzeitgedächtnis* aufgefaßt wird. Dies hat eine Analogie zu dem System (2.79) zweier sich wechselseitig erregender Neurone, welches, wenn es durch einen kurzzeitigen äußeren Stimulus in den aktiven Zustand versetzt wird, in diesem stabilen Fixpunkt längerfristig verharrt⁴³.

Das neuronale Feld ist somit kein bloßer Reaktionsmechanismus auf stimulierende sensorische Signale, sondern verhält sich auch in Abhängigkeit gespeicherter interner Zustände. Wenn infolge eines vorangegangenen kurzzeitigen Stimulus der Bereich von $-a$ bis 0 angeregt ist und ein externer Stimulus eine weitere Anregung in der Umgebung von x erzeugt, so wird diese Erregung von dem zuvor erregten Intervall $(-a, 0)$ beeinflusst. Letzteres wirkt selbst wie ein Stimulus

$$\tilde{s}(x) = \int_{-a}^0 w(x - x') dx' = \int_x^{x+a} w(x') dx'$$

auf ein anschließend erregtes Intervall in der Umgebung von x ein und versetzt dieses in Richtung auf $\max_{x \in \mathbb{R}} \tilde{s}(x)$ in Bewegung. Die gleiche Wirkung hat das später erregte Intervall auf das zuvor erregte Intervall. Ist der Abstand zwischen beiden Intervallen geringer als $\min_{x \in \mathbb{R}} w(x)$, so bewegen sich beide aufeinander zu, anderenfalls bewegen sich beide solange voneinander weg, bis sie sich außerhalb der Reichweite von $w(x)$ des jeweils anderen Intervalls befinden.

Das Konzept des neuronalen Feldes wird in [ES02] zur Beschreibung von Bewegungsplanung in Nervensystemen verwandt und ist in der Lage experimentelle Befunde bezüglich Reaktionszeiten zu erklären. In seiner Verallgemeinerung als *dynamisches Feld* kann es darüberhinaus zur Beschreibung und beim Entwurf künstlicher Systeme eingesetzt werden, wie beispielsweise in der autonomen Robotik [SDE95] [MS00] mit Verhaltensvariablen als Indexmenge.

⁴³In der Regel sorgen adaptive Prozesse im Nervensystem jedoch dafür, daß das System mit der Zeit von sich aus in den inaktiven Zustand zurückfallen wird.

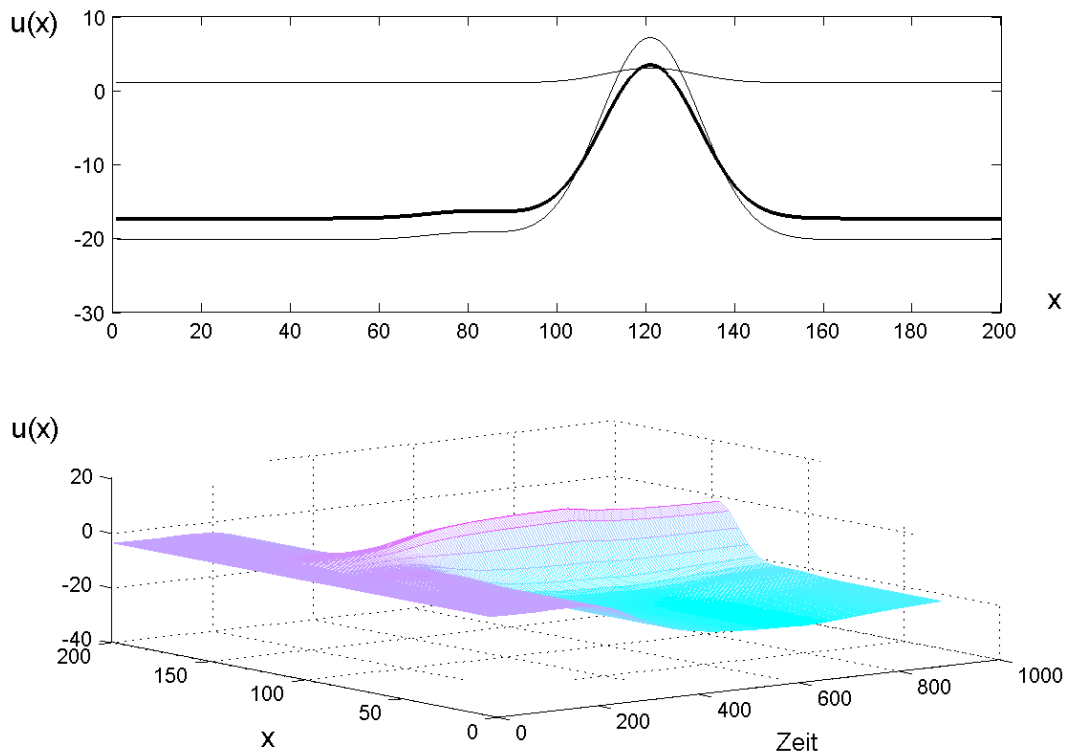


Abbildung 37: Zeitlicher Verlauf des dynamischen Feldes von Amari mit zwei initial geringfügig erregten Intervallen um 80 und 120 und einem kurzzeitigen Stimulus bei 120 (unten). Im oberen Graphen sind durch dünne Linien der Stimulus selbst und das dynamische Feld zum Zeitpunkt der Terminierung des externen Stimulus eingezeichnet, welcher von 100 bis 600 auf der Zeitskala anhielt. Die dicke Linie zeigt das dynamische Feld zum Zeitpunkt 1000. Der vorangegangene Stimulus wurde demnach im Feld gespeichert.

Anhang: Stationäre Lösungen des dynamischen Feldes von Amari

Ohne einen äußeren Stimulus haben stationäre Zustände des eindimensionalen dynamischen Feldes

$$\tau \frac{\partial u(x, t)}{\partial t} = -u(x, t) + \int w(|x - x'|) \cdot \Theta_H[u(x', t)] dx' + h + s(x, t) \quad (2.102)$$

die Potentialverteilung

$$u(x) = \int w(|x - x'|) \cdot \Theta_H[u(x')] dx' + h. \quad (2.103)$$

Hierbei sind die Neurone an der Stelle x genau dann als erregt definiert, wenn $u(x) > 0$. Die erregte Region des Feldes ist damit durch

$$R[u] = \{x | u(x) > 0\}$$

gegeben, und (2.103) kann als Integral über jene Region formuliert werden:

$$u(x) = \int_{R[u]} w(|x - x'|) dx' + h \quad (2.104)$$

Dies ist möglich, weil für die Stufenfunktion $\Theta_H(s \leq 0) = 0$ und $\Theta_H(s > 0) = 1$ gilt. Weiter sei

$$W(x) = \int_0^x w(y) dy$$

und $W_m = \max_{x>0} W(x)$, sowie $W_\infty = W(\infty)$.

Für eine \emptyset -Lösung mit $R[u] = \emptyset$ folgt aus (2.104), daß $u(x) = h$ und $h < 0$, was andererseits stets eine Lösung von (2.104) ist. Für eine ∞ -Lösung mit $R[u] = (-\infty, \infty)$ folgt entsprechend $u(x) = \int_{-\infty}^\infty w(|x - x'|) dx' + h = 2W_\infty + h$, sodaß für $u(x) > 0$, was einer Erregung entspricht, $2W_\infty > -h$ zu fordern ist. Andererseits existiert für $u(x) = 2W_\infty + h$ stets eine ∞ -Lösung, wenn $2W_\infty > -h$ ist.

Existiert eine a -Lösung, so kann die Erregung wegen der Homogenität des Feldes ohne Einschränkung der Allgemeinheit als im Bereich von 0 bis a liegend angenommen werden, sodaß $R[u] = (0, a)$. Dann ist nach (2.104) $u(x) = \int_0^a w(x - x') dx' + h = \int_{-x}^{a-x} w(x') dx' + h = W(a - x) - W(-x) + h = W(x) - W(x - a) + h$. Nach der Definition der a -Lösung muß diese $u(0) = u(a) = 0$ erfüllen. Da $W(0) = 0$ und $W(-x) = -W(x)$ folgt $W(a) = -h$.

Die Bedingung für die Existenz einer a -Lösung kann am Graphen von $W(x)$ veranschaulicht werden (Abbildung 38), indem man die Schnittpunkte $W(a) = -h$ betrachtet. Ebenso ist die Stabilität einer a -Lösung anhand dieser Schnittpunkte zu erkennen: Die a -Lösung ist stabil, wenn $W(x)$ an der Stelle a fällt

$$\left. \frac{dW(x)}{dx} \right|_{x=a} < 0$$

und instabil, wenn $W(x)$ an der Stelle a steigt

$$\left. \frac{dW(x)}{dx} \right|_{x=a} > 0.$$

Der Grund für dieses Stabilitätskriterium beruht auf der Idee, das dynamische Ver-

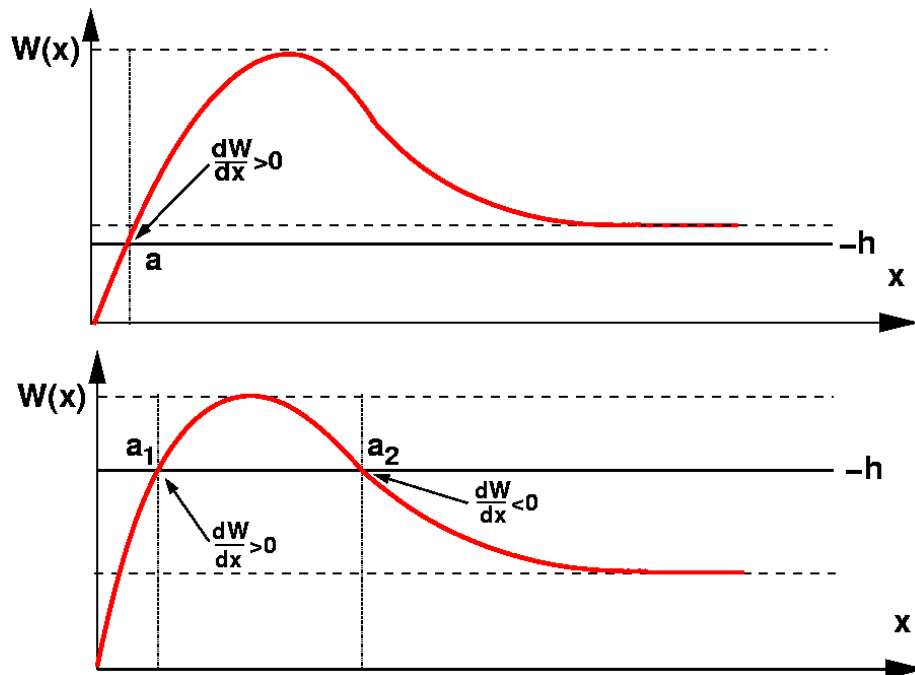


Abbildung 38: Im Falle von $W_m > 2W_\infty > 0$ gibt es zwei mögliche a -Lösungen in Abhängigkeit von h . Oben: Instabile a -Lösung für $-W_m < h < 0$. An der Stelle a steigt der Graph. Unten: Zwei a -Lösungen für $-W_m < h < -W_\infty$. An der Stelle a_1 steigt der Graph und die a_1 -Lösung ist instabil. An der Stelle a_2 fällt der Graph hingegen und die a_2 -Lösung ist stabil.

halten des erregten Intervalls allein an den Intervallgrenzen $x_1(t)$ und $x_2(t)$ zu untersuchen. Dort sind zum Zeitpunkt t die Gradienten $c_1 = \frac{\partial u(x_1, t)}{\partial x}$ und $c_2 = -\frac{\partial u(x_2, t)}{\partial x}$, wenn das erregte Intervall zu diesem Zeitpunkt durch $R[u(x, t)] = (x_1(t), x_2(t))$ gegeben ist. Falls der Zustand zum Zeitpunkt t nicht stationär ist, so ändert sich $u(x, t)$ nach einer kurzen Zeit dt zu $u(x, t + dt)$, wobei das erregte Intervall entsprechend in $R[u(x, t + dt)] = (x_1(t + dt), x_2(t + dt))$ übergeht.

Zum Zeitpunkt t ist an den Intervallgrenzen

$$u(x_i, t) = 0$$

und zum Zeitpunkt $t + dt$ ist

$$u(x_i + dx_i, t + dt) = 0,$$

wobei $x_i(t + dt) = x_i + dx_i$ mit $i = 1, 2$. Für infinitesimale dx_i und dt gilt die lineare Näherung:

$$0 = u(x_i + dx_i, t + dt) \approx \frac{\partial u(x_i, t)}{\partial x} dx_i + \frac{\partial u(x_i, t)}{\partial t} dt$$

Im ersten Summanden stehen die Gradienten c_i . Aus (2.102) läßt sich, da $u(x_i, t) = 0$ für den zweiten Summanden schließen:

$$\tau \frac{\partial u(x_i, t)}{\partial t} = \int_{x_1(t)}^{x_2(t)} w(|x - x'|) \cdot dx' + h = W(x_2 - x_1) + h$$

Zusammengefaßt erhält man

$$\frac{dx_1}{dt} = -\frac{\frac{\partial u}{\partial t}}{\frac{\partial u}{\partial x}} = -\frac{1}{\tau c_1} [W(x_2 - x_1) + h] \quad (2.105)$$

und analog

$$\frac{dx_2}{dt} = +\frac{1}{\tau c_2} [W(x_2 - x_1) + h] \quad (2.106)$$

als Geschwindigkeit der Grenzen des erregten Intervalls. Das zeitliche Verhalten der Intervalllänge wird durch

$$a(t) = x_2(t) - x_1(t)$$

beschrieben und ändert sich nach (2.105) und (2.106) mit der Zeit wie folgt:

$$\frac{da}{dt} = \frac{1}{\tau} \left(\frac{1}{c_1} + \frac{1}{c_2} \right) [W(a) + h] \quad (2.107)$$

Somit kann die Dynamik der Ausdehnung des erregten Intervalls durch eine einfache Differentialgleichung beschrieben werden. Dieses System befindet sich mit $[W(a) + h] = 0$ in einem stationären Zustand a^* . Sei $\eta = a - a^*$ eine kleine Auslenkung vom stationären Zustand, so ist

$$\frac{d}{dt} \eta = \frac{da}{dt} \approx \frac{1}{\tau} \left(\frac{1}{c_1} + \frac{1}{c_2} \right) \frac{W(a)}{da} \Big|_{a^*} \cdot \eta$$

und die a -Lösung stabil, falls $\frac{W(a)}{da} \Big|_{a^*} < 0$, da c_1 als auch $c_2 = \overbrace{-}^!$ $\frac{\partial u(x_2, t)}{\partial x}$ positiv sind, d.h. die Auslenkung fällt exponentiell auf Null zurück.