theoretische Biologie (SS 2017)

Inhaltsverzeichnis

1 Vorlesung 06.04.2017

1.1 Begriffe und Konzepte

- Begriffsbildung am Beispiel Information (Was ist Information? [Prüfungsrelevant!])
- Vorlesungsunterlagen siehe [1]
- Begriffsbildung am Beispiel Gen [Prüfungsrelevant!]
  - Welche Überschneidungen, welche Differenzen?
  - Welche Genkonzepte gibt es? (zu lesen: siehe [2] und [3])

2 Vorlesung 13.04.2017

2.1 Begriffe und Konzepte

- GWAS (Prof. Markus Scholz)
- Diskussion zum Begriff Struktur

3 Vorlesung 20.04.2017

- Gendefinition im Kontext der Messtechnik[4]
- random mating, rezessive und dominante Epistasis ???

[1] http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/concepts.pdf

[2] http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/Gerstein07_gene_definition.pdf

[3] http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/Stadler09_gene_definition.pdf

[4] http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/gene_definition.pdf

4 Vorlesung 27.04.2017

4.1 Teil 1: Dynamische Systeme

- **diskrete Zeit: "Generationen"**
  $X_1, X_2, ...$
  aktueller Zustand $X_n = F(X_{n-1}) := \overbrace{X_{n-1}}^{\text{Zustand vorher}} + \overbrace{f(X_{n-1})}^{\text{Änderung des Zustandes}}$
  $X_{n+1} = F(X_n) = F(F(X_{n-1}))$
  Beispiel:
  $X_n = (1 + a) \cdot x_{n-1}$
  a = effektive Vermehrungsrate = Geburtenrate - Sterberate
  Anfangsbedingung: $X_{t_0} = X_0$
  Bedingung: effektive Vermehrungsrate a verändert sich nicht
  Lösung: $X_n = (1 + a)^n \cdot x_0$
  im allgemeinen: mit zeitlich variablen Vermehrungsraten:

  $$X_n = \prod_{i=0}^{n-1} (1 + a_i) \cdot x_0$$

  3 verschiedene Resultate:

  - $+\infty$ für $a > 0$
  - x_0 für $a = 0$
  - 0 für $a < 0$

  $X_n = X_{n-1} + \underbrace{a \cdot X_{n-1}}_{f(X_{n-1})}$
  $f(X_{n-1}) = X_{n-1} \cdot r(X_{n-1})$
  mit r(0)=const. entspricht autonomer Wachstumsrate $[lim(x \to 0)r(x) \in R_0^+]$

- **kontinuierliche Zeit**
  $x(t + \Delta t) = x(t) + \underbrace{f(x(t)) \cdot \Delta t}_{\text{Änderung}}$
  t=aktuelle Zeit, Δt=zeitliche Änderung $\dfrac{x(t + \delta t) - x(t)}{\delta t} = f(x(t))$
  $lim(\delta t \to 0)\dfrac{x(t + \delta t) - x(t)}{\delta t} = \dfrac{\delta x}{\delta t} \mathrel{\widehat{=}}$ zeitlicher Ableitung von x
  $= \dot{x} = f(x)$

Beispiel:
$\dot{x} = a \cdot x$, $x(0) = x_0$
$\dfrac{dx}{dt} = a \cdot x$

$\dfrac{dx}{a \cdot x} = dt$

$\int\limits_{x_0}^{x(t)} \frac{1}{a \cdot x} \cdot dx = \int\limits_{0}^{1} 1 \cdot dt$

$\dot{x} = f(x) \Rightarrow \int\limits_{x_0}^{x(t)} \frac{1}{f(x)} = \int\limits_{0}^{1} dt = t$

$\frac{1}{a} \int \frac{1}{x} dx = \frac{1}{a} \cdot ln(x)$ /· a

$\cancel{\frac{1}{a}} ln(x(t)) - \cancel{\frac{1}{a}} ln(x_0) = a \cdot t$

$ln(x(t)) = at + ln(x_0)$

$x(t) = e^{at} \cdot x_0$

Grenzen für a wieder wie vorher, 3 verschiedene Resultate:

- $+\infty$ für $a > 0$
- x_0 für $a = 0$
- 0 für $a < 0$

Wie machen wir das Model realistischer?
f(x) und r(x) muss für sehr große x dann ≤ 0 werden.
$\dot{x} = f(x) = x \cdot (a - bx)$

Übungsaufgabe:

1. Löse $\dot{x} = x(a - bx)$ mit $x(0) = x_0$
2. Löse $x' = x + x(a - bx)$ mit $x(0) = x_0$

4.2 Qualitative Analyse von DS

1. Fixpunkte: keine zeitliche Veränderung $(x' = x, \dot{x} = 0)$
   d.h. diskret und kontinuierlich, f(x)=0
   Welche Fixpunkte gibt es? im Beispiel x(a-bx)=0

   (a) x=0 $\rightarrow$ Population ausgestorben
   (b) a-bx=0 $\rightarrow x = \frac{a}{b}$

Störung: $x(0) = \underbrace{\hat{x}}_{Fixpunkt} + \epsilon$ mit sehr kleinem ϵ

$\dot{x} = f(x) \rightarrow f(\hat{x} + \epsilon) = \dot{\epsilon}$

mit $x = \hat{x} + \epsilon$

$\dot{x} = \frac{\delta \hat{x}}{\delta t} + \dot{\epsilon}$

$\dot{\epsilon} = f(\hat{x} + \epsilon)$

mit Taylorreihenentwicklung: $0 = \underbrace{f(\hat{x})}_{=0} + \epsilon\frac{\delta f}{\delta x}(\hat{x}) + O(\epsilon^2)$

Für sehr kleine Störungen:
$\dot{\epsilon} = \frac{\delta f}{\delta x}(\hat{x}) \cdot \epsilon + \cancel{O(\epsilon^2)}$
Linearisierung der Differentialgleichung $x = f(x)$ in der Nähe eines Fixpunktes $\hat{x}$: $\epsilon(t) = e^{[\frac{\delta f}{\delta x}(\hat{x})]\cdot t} \cdot \epsilon_0$
$\epsilon_0 = x_0 - \hat{x}$
$\epsilon_0 \leftarrow$ initiale Störung

- Störung wird gedämpft wenn $\frac{\delta f}{\delta x}(\hat{x}) < 0 =$ STABIL
- Störung eskaliert wenn $\frac{\delta f}{\delta x}(\hat{x}) > 0 =$ INSTABIL

<u>im Diskreten Fall?</u>
$x' = x + f(x)$ mit $x = \hat{x} + \epsilon$
$\cancel{\hat{x}} + \epsilon' = \cancel{\hat{x}} + \epsilon + f(\hat{x} + \epsilon) = f(\hat{x}) + \epsilon \cdot \frac{\delta f}{\delta x}(\hat{x}) + \cancel{Rest(\epsilon)}$
$\epsilon' = \epsilon(1 + \frac{\delta f}{\delta x}(\hat{x}))$ mit:

- $\epsilon \to 0$ wenn $|1 + \frac{\delta f}{\delta x}(\hat{x})| < 1 \hat{=}$ stabil $(1 > |\frac{df}{dx}(\hat{x})|)$
- $\epsilon \to \infty$ wenn $|1 + \frac{\delta f}{\delta x}(\hat{x})| > 1 \hat{=}$ instabil $(1 < |\frac{df}{dx}(\hat{x})|)$

jetzt Mehrdimensional:

- Räuber x: $f_x(x, y) = x(-a + by - cx)$
- Beute y: $f_y(x, y) = y(+d - ex - gy)$

a,d $\rightarrow$ Wachstumsraten
c,g $\rightarrow$ intraspezifische Konkurrenz
be, $\rightarrow$ interspezifische Beeinflussung

<u>4 Fixpunkte möglich:</u>

1. Fall: x=0, y=0 (trivialer Fall)
2. Fall: x≠0, y=0
3. Fall: x=0, y≠0
4. Fall: x≠0, y≠0

Nicht alle Fälle müssen biologisch relevant sein, es können sich zum Beispiel auch Fixpunkte im negativem befinden $\rightarrow$ wird nicht betrachtet

Stabilität: gegeben durch

- $\frac{\delta f_x}{\delta x}(\hat{x}, \hat{y})$ $\frac{\delta f_x}{\delta y}(\hat{x}, \hat{y})$
- $\frac{\delta f_y}{\delta x}(\hat{x}, \hat{y})$ $\frac{\delta f_y}{\delta y}(\hat{x}, \hat{y})$

Übungsaufgabe 2:
Bestimme die Fixpunkte von Räuber-Beute-Modell für a,b,c,d,e,g >0
Welche Fixpunkte gibt es immer? Wieviele sind das?

4.3 Teil 2: Genkonzept

- Unterschiede und Überscheidungen zwischen den beiden in den Papern vorgestellten Genkonzepten (siehe Vorlseung 13.04.2017) [Prüfungsrelevant]

5 Vorlesung 04.05.2017

- Vorlesung entfallen wegen: Mitteldeutschen Bioinformatik-Meeting 2017[5]

[5] `http://me17.bioinf.uni-leipzig.de/`

6 Vorlesung 11.05.2017

6.1 Teil 1: Populationsdynamik

Eindimensional:
$\dot{x} = f(x), x(0) = x_{(0)}$
Fixpunkte: $\dot{x} = 0 = f(\hat{x})$

$x - \hat{x} = \epsilon \leftrightarrow x = \hat{x} + \epsilon$
$\hat{x} = \dot{\epsilon}?$
$\dot{x} = \dot{\epsilon} = f(x) = f(\hat{x} + \epsilon)$
durch Taylorreihe folgt: $\underbrace{f(\hat{x})}_{=0} + \frac{\delta f(\hat{x})}{\delta x} \cdot \epsilon + \text{Rest}(\epsilon^2)$

$\dot{\epsilon} = \frac{\delta f(\hat{x})}{\delta x} \cdot \epsilon$ für $|\epsilon|$ klein
$\epsilon(t) = e^{\frac{\delta f(\hat{x})}{\delta x} \cdot t} \cdot \epsilon(0)$
$\frac{\delta f(\hat{x})}{\delta x} < 0 \Rightarrow$ stabil $\hat{=}$ anziehend
$\frac{\delta f(\hat{x})}{\delta x} > 0 \Rightarrow$ instabil $\hat{=}$ abstoßend

Mehrdimensional:
DGL. System 1. Ordnung
$\dot{x_1} = f_1(x_1, x_2, ..., x_n)$
...
$\dot{x_n} = f_n(x_1, x_2, ..., x_n)$

Anfangsbedingungen (AB):
$x_1(0) = \dot{x_1}$
...
$x_n(0) = \dot{x_n}$
wo kommen die Punkte hin? $\dot{x_n}$ oder $x_{\dot{n}}$

Fragen:

1. Existenz von Lösungen?
2. Eindeutigkeit?
3. Wie schaut die Lösung überhaupt aus?

Betrachten unser DGL. System auf einem Gebiet $\Omega \subseteq R^n$
$f = (f_1 ... f_n)^T$
$f : x \mapsto f(x)$ Vektorfeld
$x \in \Omega, R^n$... n-dim Vektor von Änderungen

Lösung:

Übungsaufgabe: Visualisiere die Vektorfelder für Reuber-Beute Modelle

Annahme: Vektorfeld f ist mindestens 1 mal stetig differenzierbar, d.h.
$\frac{\delta f_i}{\delta x_j}$ existiert auf ganz Ω, sind auf ganz Ω stetig

$\Rightarrow$ Existenz: für jede Anfangsbedingung x_0 in Ω gibt es eine Zeitspanne $T(x_0)$ sodass $x(t|x_0)$ für alle $0 \leq t \leq T(x_0)$ existiert, eindeutig ist und $t \mapsto x(t|x_0)$ stetig ist

Trajektorien kreuzen sich nie!

Kreuzungspunkt nicht eindeutig $\Rightarrow$
$z_0 = x_0(t_1) = y_0(t_2)$
$z(t|z_0) = x_0(t - t_1) = y_0(t - t_2)$
laut Eindeutigkeitssatz nicht möglich $\rightarrow$ Situation kann nicht vorkommen

Fixpunkte: $\hat{x} \in \Omega$ sodass $f(\hat{x}) = 0$

6.1.1 Beispiel: Räuber-Beute-Modell

Lotka-Volterra-Gleichung: $f_i(x) = (r_i + \sum_j b_{ij} \cdot x_j)x_i$ mit

r_i ... Spezies-spezifische autonome Wachstumsrate
$\sum_j b_{ij} \cdot x_j$ ... Interaktion mit allen (anderen) Spezies + intraspezifische Konkurenz $b_{ii} \cdot x_i$
x_i ... Wachstum proportional zur Populationsgröße

Fixpunkte in LV-Systemen:

1. Trivialer Fixpunkt:
   $x_1 = x_2 = \ldots = x_n = 0, \hat{x} = 0 \Rightarrow$ alle Spezies ausgestorben

2. innerer Fixpunkt: alle $x_i \neq 0$
   $f_i(x) = (r_i + (Bx)_i)x_i = 0$ mit $(Bx)_i = \sum_j b_{ij} \cdot x_j$
   $\Leftrightarrow r_i + (Bx)_i = 0$
   $\Leftrightarrow Bx = -r$ (lineares GLS)
   $x = -B^{-1} \cdot r$

3. Fixpunkte von Teilsystemen: $S = \{1, \dots, n\}$
   $A \subseteq S \to x_i = 0$ für $i \in S \backslash A$: verstorben, $x_i \neq 0$ für $i \in A$

es existieren insgesamt 2^n mögliche Fixpunkte!

Status um Fixpunkte?

$\epsilon = x - \hat{x}, \epsilon_i = x_i - \hat{x}_i$

$\dot{\epsilon_i} = \dot{x_i} = f_i(x_1, \dots, x_n) = f_i(\hat{x} + \epsilon) = f_i(\hat{x_1} + \epsilon_1, \dots, \hat{x_n} + \epsilon_n)$

Taylorreihenentwicklung: $= f_i(\hat{x_1}, \dots, \hat{x_n}) + \sum\limits_k \frac{\delta f_i(\hat{x})}{\delta x_k} \cdot \epsilon_k + \text{Rest}(\epsilon^2)$

für $|\epsilon|$ sehr klein folgt:

Jacobimatrix $J_{ik} := \sum\limits_k \frac{\delta f_i(\hat{x})}{\delta x_k}$

$\epsilon_i = (J_\epsilon)_i$ zeitliche Entwicklung der Störung $\epsilon \Rightarrow \dot{\epsilon} = J(\hat{x}) \cdot \epsilon$

Das Verhalten von $\dot{x} = f(x)$ wird in der Nähe einer Fixpunktes $\hat{x}$ durch die Linearisierung $\dot{\epsilon} = J(\hat{x}) \cdot \epsilon$ beschrieben.

$\Rightarrow$ müssen verstehen, wie die Lösung von linearen DGL-Systemen aussehen
$\dot{x} = Ax, x \in R^n, A \in R^{n,n}$

1. Dimension:

$\dot{x} = a \cdot x \Rightarrow x(t) = e^{a \cdot t} \cdot x_0$

Formale Lösung in $R^n : x(t) = exp(t \cdot A) \cdot x_0$

Darstellung Exponentialfunktion Matrix:

Zahl: $e^a = 1 + \frac{1}{1!}a^1 + \frac{1}{2!}a^2 + \dots$

$$exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

$$[exp(A)]_{ij} = \underbrace{\sum_{k=0}^{\infty} \frac{1}{k!} [A^k]_{ij}}_{\text{konvergiert für alle A}}$$

$$\Rightarrow exp(t \cdot A) = \sum_{k=0}^{\infty} \frac{1}{k!} (t \cdot A)^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$$

$$\Rightarrow \frac{d}{dt} [exp(t \cdot A)]_{ij} = \sum_{k=0}^{\infty} \underbrace{\frac{dt^k}{dt}}_{k \cdot t^{k-1}} \cdot \frac{1}{k!} [A^k]_{ij}$$

$$= \sum_{k=0}^{\infty} t^{k-1} \cdot \underbrace{\frac{k}{k!}}_{k \cdot (k-1)!} [A^k]_{ij}$$

$$= \sum_{k=0}^{\infty} \frac{1}{(k-1)!} \cdot t^{k-1}[A^k]_{ij}$$

mit $k-1=l$ folgt: $\sum_{l=0}^{\infty} \frac{1}{l!} \cdot t^l [A^{l+1}]_{ij}$

$$\Rightarrow \frac{d}{dt} exp(t \cdot A) = \sum_{l=0}^{\infty} \frac{t^l}{l!} \cdot A^{l+1} = A \sum_{l=0}^{\infty} \underbrace{\frac{t^l}{l!} \cdot A^l}_{exp(t \cdot A)} = A \cdot exp(t \cdot A)$$

$\dot{x} = A \cdot x \Rightarrow$ Ansatz:
$x(t) = exp(t \cdot A) \cdot x_0$
$\dot{x}(t) = A \cdot \underbrace{exp(t \cdot A) \cdot x_0}_{x(t)}$
$\Rightarrow x(t) = exp(t \cdot A) \cdot x_0$ löst tatsächlich die lineare DGL

Spezialfall: Matrix A ist eine Diagonalmatrix

$$A = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \Rightarrow A^k = \begin{pmatrix} \lambda_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^k \end{pmatrix}$$

$$\sum_{k=0}^{\infty} \frac{1}{k!} \cdot A^k = \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \begin{pmatrix} \lambda_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^k \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \lambda_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \lambda_n^k \end{pmatrix} =$$
$$\begin{pmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_n} \end{pmatrix}$$

$\Rightarrow x(t) = exp(tA)x_0?$
$x_i(t) = \sum_j [exp(tA)]_{ij} \cdot x_{\mathring{j}} = \sum_j e^{t \cdot \lambda_j} \cdot \delta_{ij} \cdot x_{\mathring{j}}$
mit $x_{\mathring{j}}$ =j-Koordinate der $A_i B_i$, δ_{ij}=Kronecker-Delta was bedeutet $\mathring{j}$?

$x_i(t) = e^{t\lambda_i} \cdot x_{\mathring{j}}$

$x_i(t) \to 0$ wenn $\lambda_i < 0$ stabile Richtung
$x_i(t) \rightsquigarrow \pm\infty$ wenn $\lambda_i > 0$ instabile Richtung

wenn alle Richtungen stabil: Senke
wenn alle Richtungen instabil: Quelle
sonst: Sattelpunkt

Was wenn Jacobimatrix nicht diagonal?
Eigenwerte und Eigenvektoren
$Au = \lambda u \leftarrow u \neq 0$ Eigenvektor von A zum Eigenwert λ
$Au^{(j)} = \lambda_j u^{(j)}$
$U \cdot A \to A = \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix}$
$A \underbrace{[u^{(1)}, \ldots, u^{(n)}]}_{*} = [\lambda_1 u^{(1)}, \ldots, \lambda_n u^{(n)}]$
*... U-Matrix deren Spalten die Eigenvektoren von A sind

$A \cdot U = U \cdot A$ vorausgesetzt es gibt n linear unabhängige Eigenwerte von A
$U^{-1} exp(tA) U = U^{-1} \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k U = U^{-1} \sum_{k=0}^{\infty} \frac{t^k}{k!} U A^k = \underbrace{U^{-1} U}_{I} \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$
$= exp(tA) = e^{tA}$

$x(t) = exp(tA) \cdot x_0 = U exp(tA) U^{-1} \cdot x_0$
neue Koordinate: $y := U^{-1} x \Leftrightarrow x = U \cdot y$
$\dot{y} = U^{-1} \dot{x} = U^{-1} A x = \underbrace{U^{-1} \cdot A \cdot U}_{*} \cdot y = Ay$
* wenn A symmetrisch, dann gibt es n linear unabhängige Eigenvektoren, die sind sogar orthogonal $U^{-1} = U^T$ alle Eigenwerte $\lambda_i \in R$

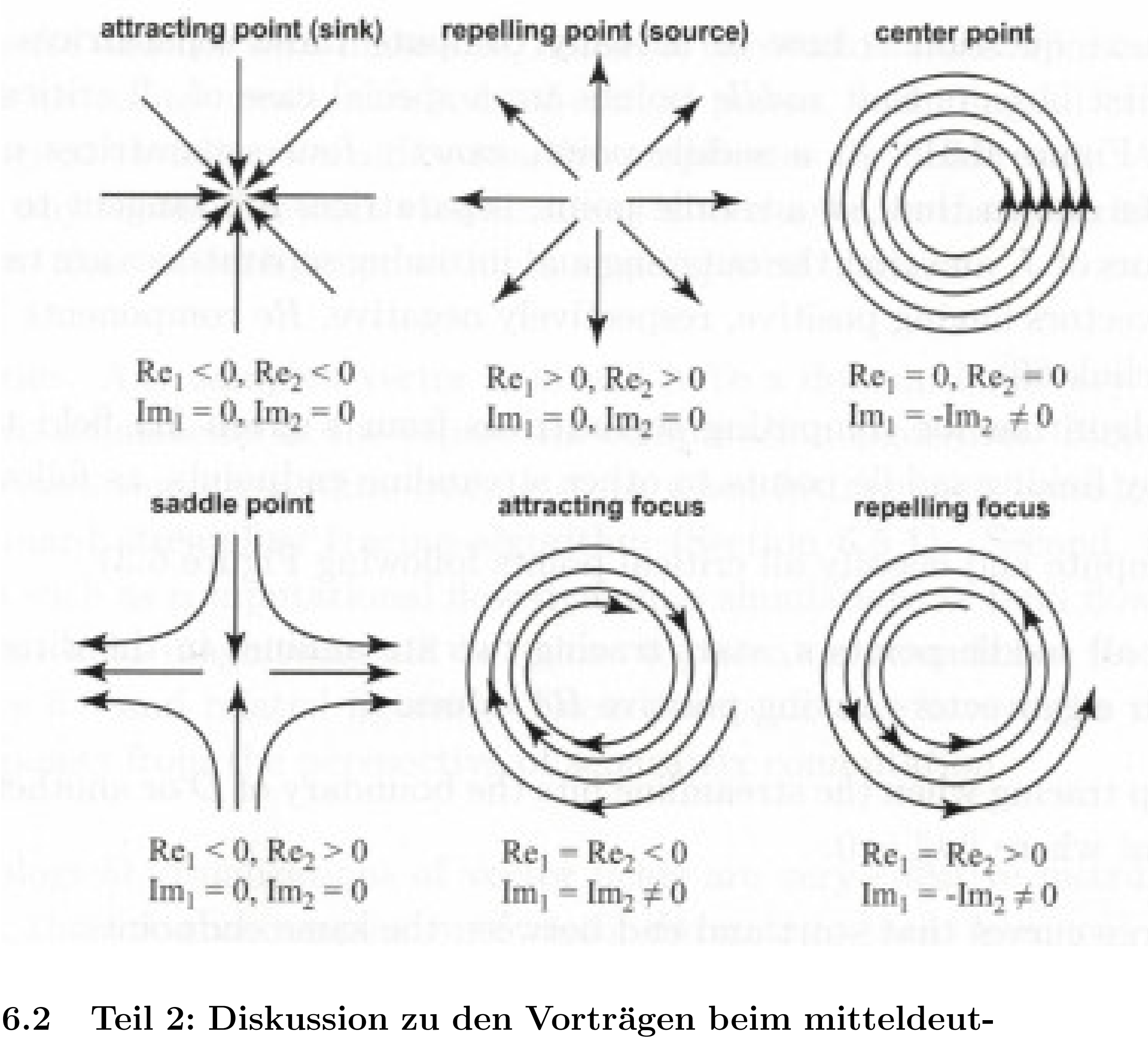
hier fehlt was im Skript von Christian

Eigenwerte i. A. nicht reell:
macht nix, nehmen Eigenwerte über komplexe Zahlen
für jeden Eigenwert $\lambda = Re\lambda + i \cdot Im\lambda$ ist auch konjugiert komplexe Zahl $\lambda = Re\lambda - i \cdot Im\lambda$ ein Eigenwert

Satz von Moivre: $e^{t\lambda} = e^{tRE\lambda} \cdot e^{itIm\lambda} = e^{tRe\lambda}(cos tIm\lambda + i sin tIm\lambda)$
$e^{t\overline{\lambda}} = e^{tRE\lambda} \cdot e^{-itIm\lambda} = e^{tRe\lambda}(cos tIm\lambda - i sin tIm\lambda) \leftarrow$ Wirbel

λ_i ... reell in diese Eigenrichtung $u_i e^{t\lambda_i}$
$\lambda_i, \overline{\lambda_i}$ ... Paar von konjugiert komplexen Eigenwerten dann gibt es 2D-Ebene, in der
$y_{1,2}(t) = e^{tRe\lambda} \cdot cos tIm\lambda$

Allgemeiner Fall: Stabilität durch $Re\lambda_i$ gegeben.
$Re\lambda_i < 0$ anziehende Richtung
$Re\lambda_i > 0$ abstoßende Richtung
Quelle: $Re\lambda_i > 0$ für alle i
Senke: $Re\lambda_i < 0$ für alle i
Sattelpunkt: sowohl $Re\lambda_i < 0$ als auch $Re\lambda_i > 0$ existieren.



6.2 Teil 2: Diskussion zu den Vorträgen beim mitteldeutschen Bioinformatik-Meeting 2017

7 Vorlesung 18.05.2017

7.1 Musterbildung

8 Vorlesung 01.06.2017

8.1 Teil 1: Musterbildung

8.2 Teil 2: Cat Coat Colors

8.3 Vergleich Übungsaufgaben

9 Vorlesung 08.06.2017

9.1 Teil 1: Fitnesslandschaften

evol. Theorie: Wachstumsrate in Räuber-Beute-Modellen + Reproduktion mit Variation + "Survival of the fittests" → Wachstumsrate einer Population in einer gege. Umgebung

Günther Wagner: Messtheorie von Fitness (measurement theorie)

X… Suchraum (genotyp—phänotyp), allgemein irgendeine Repräsentation der betrachteten Taxa
Ähnlichkeitsstruktur σ
Fitnessfunktion: $f : x \to R$
mit R=totale geordnete Menge ($f_1, f_2 \in R : f_1 < f_2, f_1 > f_2, f_1 = f_2$)

Begründer: Sewall Wright ($\sim$ 1930)
siehe zurückliegendes Bild: Individuum hat höhere Wahrscheinlichkeit Erbgut in nächste Generation zu übertragen ("Verbesserung")

9.1.1 Genetische Algorithmen

Idee: benutze künstliche Evolution um Optimierungsprobleme zu lösen

1. Population $A \subseteq X$
2. Nachfolgerpopulation von Kandidaten C(A)
3. Selektiere die Besten bezüglich Fitnessfunktion $x \in C(A)$
4. zurück zu 1.

genetische Algorithmen, evolutionäre Progammierung (Rechenberg, Schwefel $\sim$ 1960/70)

geg.: RNA oder Proteinsequenz α ges.: Alle möglichen Strukturen x, die α einnehmen kann → Menge x von Konfigurationen
Energiefunktion $f : X \to R$
z.B. Loop basiertes Energiemodell für RNA Sekundärstrukturen

Lenskis E.Coli Zucht[6]
X… Menge von Gen oder Genomsequenzen
σ… Mutationen (hauptsächlich Substitution, Insertion, Deletion)
(X,σ)… Suchraum ⇔ Graphen über $\{A,G,T,C\}^n$
mit n=Sequenzlänge
mit Kanten=Hammingdistanz 1 (| Levensteindistanz 1)

9.1.2 3D-Strukturen

Proteinstruktur = $(\overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_n})$ mit $\overrightarrow{x_1}$=3D-Koordinaten für Atom 1

Constraint: Bindungswinkel, Bindungslängen
X... alle möglichen 3D-Einbettungen des Proteins
Nachbar: $||\overrightarrow{x} - \overrightarrow{x}'|| = \sum_i |\overrightarrow{x_i} - \overrightarrow{x_i}'| < \epsilon$ für gegebenes $\epsilon > 0$

Wenn RNA Sekundärstrukturen?
X... Menge aller erlaubten Strukturen [(,),.]
x∼y wenn x und y sich durch ein Basenpaar unterscheiden ⇒ Graph

Beispiel:

9.1.3 Optimierung auf Landschaften

→ max, min finden

Wie misst man Rauheit?
Minimum: $\hat{x} \in X$ sodass $\forall y$ Nachbar von $\hat{x} : f(\hat{x}) \leq f(y)$
für metrischen (kontinuierlichen) Raum: $\forall y : |\hat{x} - y| < \epsilon$
Maximum: $f(\hat{x}) \geq f(y)$

Was möchte man messen?

- # lokale Minima, nur gut bei kleinen Instanzen, daher sampeln! (zufällige x wählen und bestimmen ab Minimum)
- mittlere Länge von adaptiven walks
  $x_0, x_1, x_2, \ldots, x_l$ sodass x_i Nachbar von x_{i-1} ist und $\underbrace{f(x_i) > f(x_{i-1})}_{*}$ i=1...l

  * für Fitness (für Energie <)
- Alternative: gradient walks (Weg des stelsten Anstiegs), Distanz zum "nächstgelegenenll-okalen Min/Max

9.1.4 Autokorrelationsfunktionen

$x_0, x_1, x_2, \ldots, x_l$ sodass x_i Nachbar von x_{i-1}
betrachte Folge der Funktionswerte $f(x_0), f(x_1), \ldots$ betrachte das als Signal (Zeitserie)

$$\varrho(\tau) = \frac{< f(x_t) \cdot f(x_{t+\tau}) >_t - < f >_t^2}{< f^2 >_t - < f >_t^2}$$

$< f_t > \ldots$ Mittelwert über die $f(x)$
$< f_t > := \lim\limits_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} f(x_t)$
$< f(x_t) \cdot f(x_{t+\tau}) > := \lim\limits_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) f(x_{t-\tau})$

X... Menge von Gen oder Genomsequenzen
$\sigma \ldots$ Graph, regulär (jedes x hat gleich viele Nachbarn D)
A... Adjazenz von (X,σ)

$$\varrho(\tau) = \frac{(f(\frac{1}{D} \cdot A)^\tau \cdot f) - (f)^2}{(f^2) - (f)^2}$$
$\Rightarrow$ leichter so auf Graphen als direkt auf Fitness (Funktion)

<u>Korellationslänge:</u>

Funktion in Abhängigkeit der Verschiebung von τ
$$L_c = \sum_{\tau=0}^{\infty} \varrho(\tau)$$
$(f) = \frac{1}{|X|} \sum_{x\in X} f(x)$
$(f^2) = \sum_{x\in X} f(x)^2$
$(f, \frac{1}{D} \cdot A \cdot f) := \sum_{x\in X} \sum_{y\in X} \frac{1}{D} f(y) \cdot A_{yx} \cdot f(x)$
wenn $(f) = 0$ folgt vereinfachte Gleichung
$$(f^2) = 1 \rightarrow \varrho(\tau) =< f(x_t), f(x_{x_t - \tau}) >= (f, \underbrace{\frac{1}{D} \cdot A \cdot f}_{\text{Graphstruktur}})$$

<u>Beispiel:</u>

Kostenfunktion in Fall 1 ändert sich stärker als in Falls 2
$\Rightarrow L_R \simeq 2L_T$

?(Länge Korrelationslängen $\rightarrow$ Lange Wege zum nächsten Minimum $\rightarrow$ gut!)

9.2 Übung farbliche Ausprägung Katzenfell und beteiligte Gene