Stabilität: "kleine Störungen in den Daten ändern die persistente Homologie nur unwesentlich"

(1) Einparameterfamilien

Daten: K Simplizialkomplex

$f, g : K \to \mathbb{R}$ zwei monotone Funktionen.

(d.h. $\sigma \subseteq \tau$ in K $\Rightarrow$ $f(\sigma) \leq f(\tau)$, $g(\sigma) \leq g(\tau)$)

Notation: $F : K \times [0,1] \longrightarrow \mathbb{R}$

$$(\sigma, t) \longmapsto (1-t) f(\sigma) + t \cdot g(\sigma) =: f_t(\sigma)$$

Beachte: Aus der Monotonie von f und g folgt:

$$\sigma \subseteq \tau \Rightarrow f_t(\sigma) \leq f_t(\tau)$$

d.h. die f_t sind "gleichmäßig monoton"

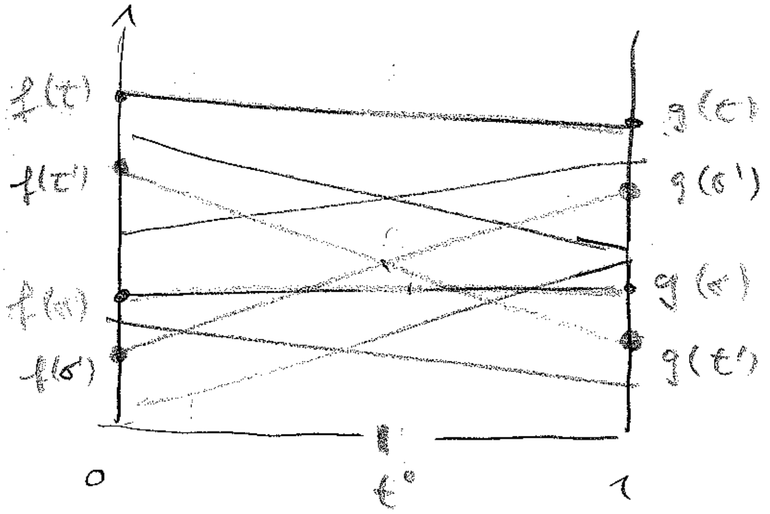
$\Rightarrow$ auf K existiert eine Ordnung, die mit der durch die f_t definierten partiellen Ordnung und der Seitenrelation verträglich ist.

$\rightsquigarrow$ für jedes f_t können wir dann (separat) die persistente Homologie berechnen

Überlegung: unterscheiden sich f und g nicht wesentlich, so kann man vielleicht die persistente Homologie für g bzw. die f_t aus der von f einfach erhalten.

Annahme (zur Vereinfachung): f und g seien injektiv.
$\rightsquigarrow$ f und g definieren Ordnung auf K.

Frage: Wie ändert sich die Ordnung mit t?



Für alle $t \in [0,1]$ ist $\sigma < \tau$.

Für $t \in [0, t_0)$ gilt $\sigma' < \tau'$, für $t \in (t_0, 1]$ gilt $\tau' < \sigma'$

Notation: Die Werte in $[0,1]$, bei denen Kreuzungen im obigen Diagramm vorliegen, nennen wir krit. Werte

Annahme (weitere Vereinfachung): Es gibt keine zwei Paare von "Linien" die sich an ein und denselben kritischen Wert kreuzen.

D.h. bei jedem kritischen Wert gibt es genau eine Transposition.

Folgerungen:

- es gibt nur endlich viele kritische Werte
- außer an die kritischen Werten legt f_t eine Ordnung fest; durchläuft t einen kritischen Wert ändert sich die Ordnung um (genau) eine Transposition.

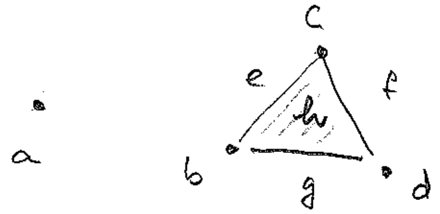
$\rightsquigarrow$ Untersuche die Auswirkungen einer Transposition auf die persistente Homologie

<u>Bem.</u> Ist m die Anzahl der Simplizes in K, so können max. $\binom{m}{2}$ Transpositionen auftreten. In diesem "worst case"-Szenario ist der Rechenaufwand mit den im Folgenden angedeuteten Rechenverfahren zur Verarbeitung von Transpositionen ebenso aufwändig wie die Berechnung der persistenten Homologie bzgl. g unabhängig von f.

$\rightsquigarrow$ Jedoch erhält man interessante Daten als Zwischenergebnis.

② Berechnung der persistenten Homologie (Wiederholung am Beispiel)

Verfahren:
- Bestimme aus den Daten die "Randoperatormatrix" ∂
- Reduziere ∂, und bestimme aus der reduzierten Matrix leicht das Persistenzdiagramm.



$f : K \longrightarrow \mathbb{R}$

	σ_1 a	σ_2 b	σ_3 c	σ_4 d	σ_5 e	σ_6 f	σ_7 g	σ_8 h
f	1	2	3	4	5	6	7	8

$$\partial = \begin{array}{c} \begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\ \left(\begin{array}{cccccccc} 0 & 0 & 0 & 0 & & & & \\ & & & & 1 & 0 & 1 & \\ & & & & 1 & 1 & 0 & \\ & & & & 0 & 1 & 1 & \\ & & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & 0 \end{array}\right) \end{array} \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{array}$$

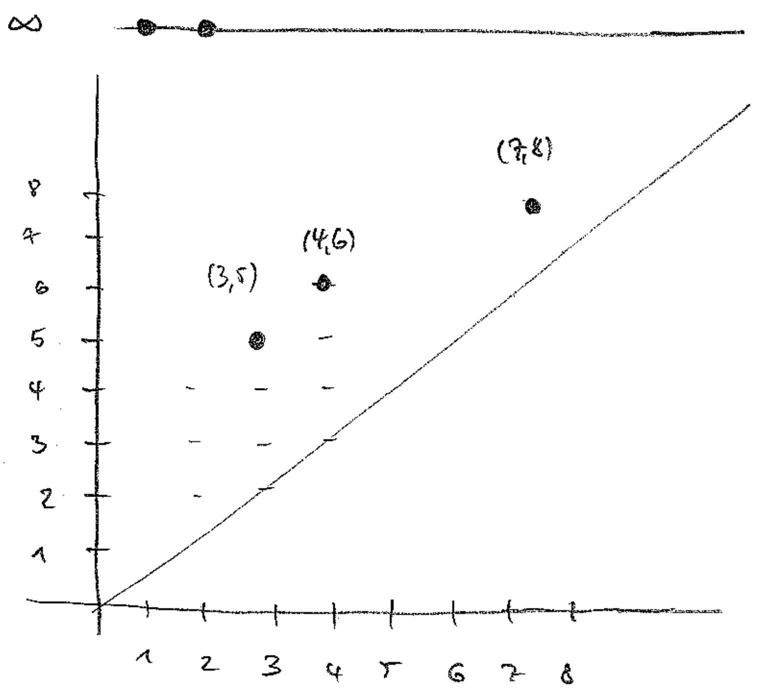
nicht reduziert!

$$\rightsquigarrow \quad \partial' = \left(\begin{array}{cccccccc} 0 & 0 & 0 & 0 & & & & \\ & & & & 1 & 0 & 0 & \\ & & & & 1 & 1 & 0 & \\ & & & & & 1 & 0 & \\ & & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & 0 \end{array}\right)$$

Regeln zur Erstellung des Persistenzdiagramms.

- erhalte einen Punkt bei $(f(\sigma_i), \infty)$, falls i eine Nullenspalte indiziert und die i-te Zeile keine lowest 1 hat.

- erhalte einen Punkt bei $(f(\sigma_i), f(\sigma_j))$, falls i eine Nullenspalte indiziert und $i = \mathrm{low}(j)$



← permanente Zykel.

Wesentliche Punkte:

- Man erhält eine Zerlegung $R = \partial \cdot V$ mit: R ist reduziert, V ist eine obere Dreiecksmatrix

kodiert die Additionen, die zur Reduktion führen

- Die Zerlegung ist nicht eindeutig, die Daten, die wir aus ihr ableiten aber schon!

(3) Auswirkungen einer Transposition auf die Berechnung der persistenten Homologie

Der "neue" Randoperator ist

$$\partial' = P \partial P \quad , \text{ mit } \quad P = \begin{pmatrix} 1 & & & & & & & & & \\ & 1 & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & 1 & & & & & & \\ & & & & 0 & 1 & & & & \\ & & & & 1 & 0 & & & & \\ & & & & & & 1 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{pmatrix}$$

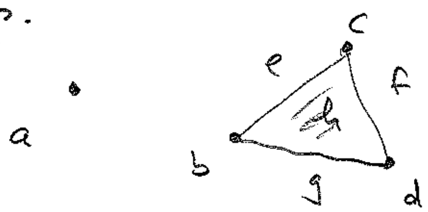
$$= P R V P$$

$$= \underbrace{(P R P)}_{R'} \underbrace{(P V P)}_{V'}$$

- im Allg. ist $R' = PRP$ nicht reduziert
- " " " $V' = PVP$ keine obere Dreiecksmatrix

⇝ aber das kann man beides durch geeignete Modifikationen leicht ändern; und zwar durch sehr einfachen Algorithmus.

<u>Bsp.</u>



	a	b	c	d	e	f	g	h
f	1	2	3	4	5	6	7	8
g	1	2	3,5	4	5,5	5,4	7	8

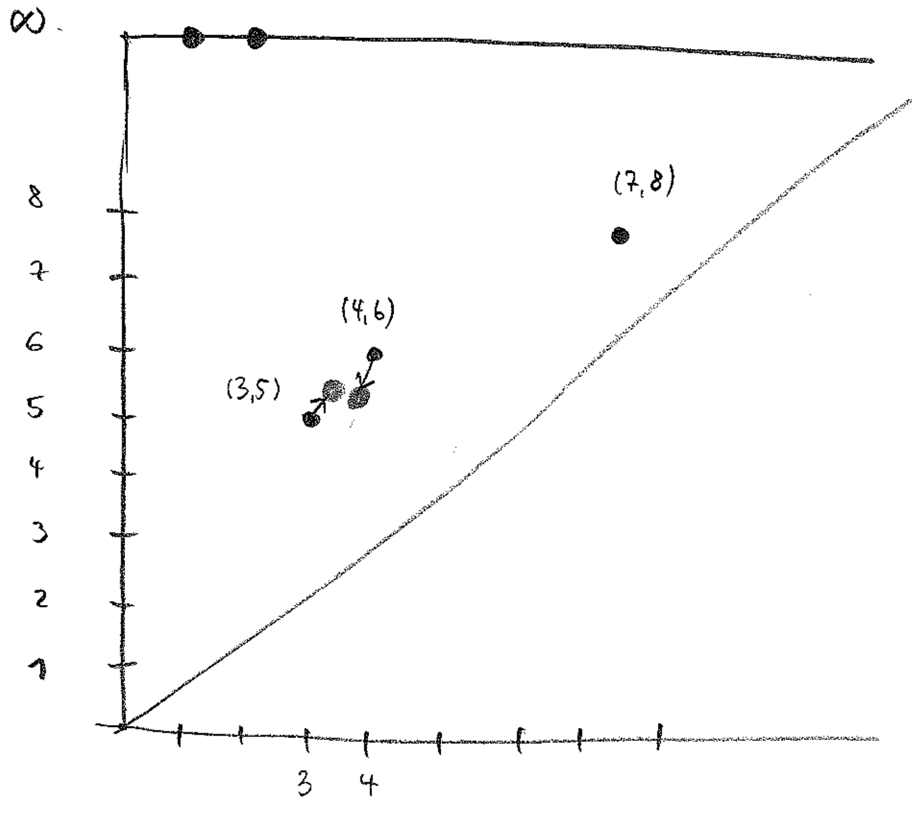
⇝ Die Ordnungen unterscheiden sich durch die Transposition, die e und f vertauscht.

$i=5$

$$P R' P = \begin{pmatrix} \boxed{0\ 0\ 0\ 0} & & \\ & \boxed{\begin{matrix}1&0&0\\0&1&0\\1&1&0\end{matrix}} & \\ & & \boxed{\begin{matrix}1\\1\\1\end{matrix}} \\ & & 0 \end{pmatrix}$$

Algorithmus liefert: addiere i-te Spalte zur $(i+1)$-ten Spalte.

$$\rightsquigarrow \quad = \begin{pmatrix} \boxed{0\ 0\ 0\ 0} & & \\ & \boxed{\begin{matrix}1&1&0\\0&1&0\\1&0&0\end{matrix}} & \\ & & \boxed{\begin{matrix}1\\1\\1\end{matrix}} \\ & & 0 \end{pmatrix}$$



$\leftarrow$ zugehöriges Persistenzdiagramm

④ Der Flaschenhalsabstand (bottleneck distance)

Ziel des Abschnitts: Maß für unterschiedliche Persistenzdiagramme

Annahme (zur Vereinfachung):

1) wir arbeiten mit reduzierter Homologie

2) wir nehmen an: $\tilde{H}_*(K) = 0$.

Beachte: unter diesen Voraussetzungen gibt es

- keine permanenten Zykel; zugehörige Persistenzdiagramme liegen also in $\mathbb{R}^2$.

- Jedes Persistenzdiagram besitzt $\frac{|K|}{2}$ viele Punkte. Denn:

<u>Erinnerung</u>: Sei $f: K \to \mathbb{R}$ monoton und injektiv, und es sei $\sigma_1 < \sigma_2 < \dots < \sigma_m$ die zugeh. Ordnung der Simplizes.

<u>Def</u> (σ_i, σ_j) ist ein Paar, falls in R die i-te Spalte eine Nullenspalte ist, und $i = \text{low}(j)$.

Ist (σ_i, σ_j) ein Paar so repräsentiert σ_i eine Homologieklasse, die bei $f(\sigma_i)$ geboren wird und bei $f(\sigma_j)$ stirbt. Wegen $\tilde{H}_*(K) = 0$ läßt sich K in Paare aufteilen.

Def. (i) Für $x, y \in \mathbb{R}^2$ setze $\|x - y\|_\infty = \sup\{|x_1 - y_1|, |x_2 - y_2|\}$

(ii) Sind X und Y zwei bijektive endl. Mengen in $\mathbb{R}^2$, setze

$$W_\infty(X, Y) = \inf_{\substack{\eta : X \to Y \\ \text{bij.}}} \ \sup_{x \in X} \|x - \eta(x)\|_\infty$$

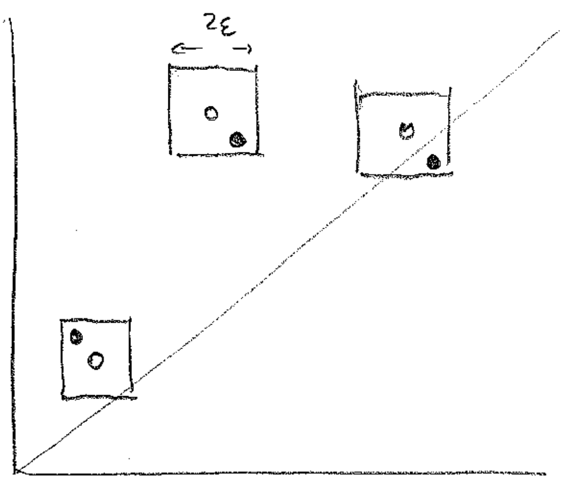
Der <u>Flaschenhalsabstand</u> zweier Persistenzdiagramme $Dgm'(f)$ und $Dgm'(g)$ ist definiert durch

$$W_\infty(Dgm'(f), Dgm'(g))$$

<u>Anschauung:</u>

$W_\infty(X, Y) < \varepsilon$

$\iff$ es gibt eine Bijektion $\eta : X \to Y$, so dass $\forall x \in X$ gilt: $\eta(x)$ liegt im Quadrat mit Kantenlänge 2ε um x.



Prop. Es gilt

- $W_\infty(X, Y) = 0 \iff X = Y$
- $W_\infty(X, Y) = W_\infty(Y, X)$
- $W_\infty(X, Z) \leq W_\infty(X, Y) + W_\infty(Y, Z)$

für drei bijektive endliche Teilmengen X, Y, Z in $\mathbb{R}^2$.

Stabilitätssatz: Er sei K ein Simplizialkomplex mit $\tilde{H}_*(K) = 0$. Weiter seien $f, g: K \to \mathbb{R}$ zwei monotone Funktionen, die den obigen Annahmen genügen. Dann gilt:

$$W_\infty(\mathrm{Dgm}' f, \mathrm{Dgm}' g) \leq \| f - g \|_\infty$$

wobei $\| f - g\|_\infty = \sup_{\sigma \in K} \| f(\sigma) - g(\sigma) \|$.

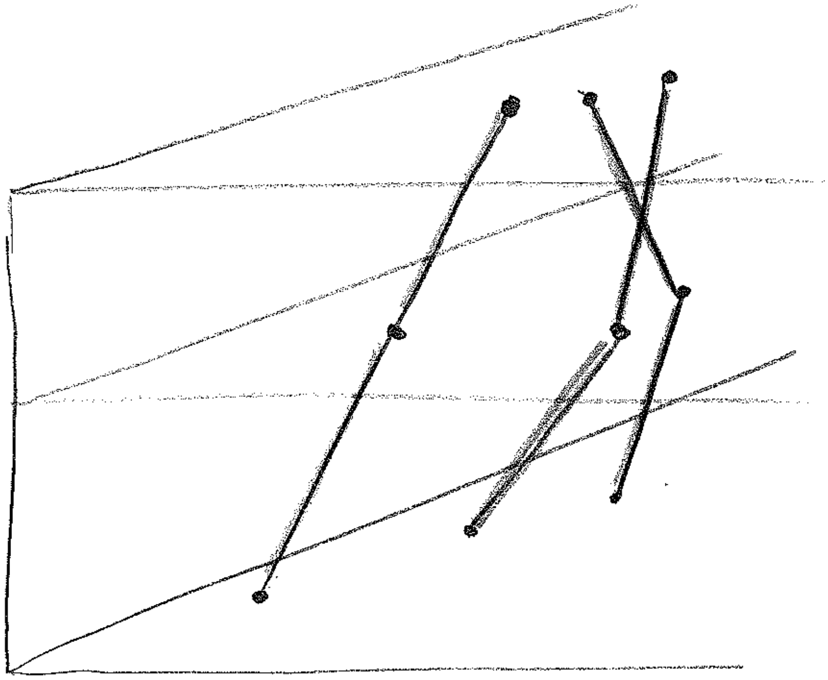
⑤ Beweis des Stabilitätssatzes mit Reben im Weingarten

Wir betrachten wieder

$$f_t = (1-t) f + t g$$

und stellen die Persistenzdiagramme für die f_t; $t \in [0,1]$ in $\mathbb{R}^2 \times [0,1]$ dar.

Bsp.



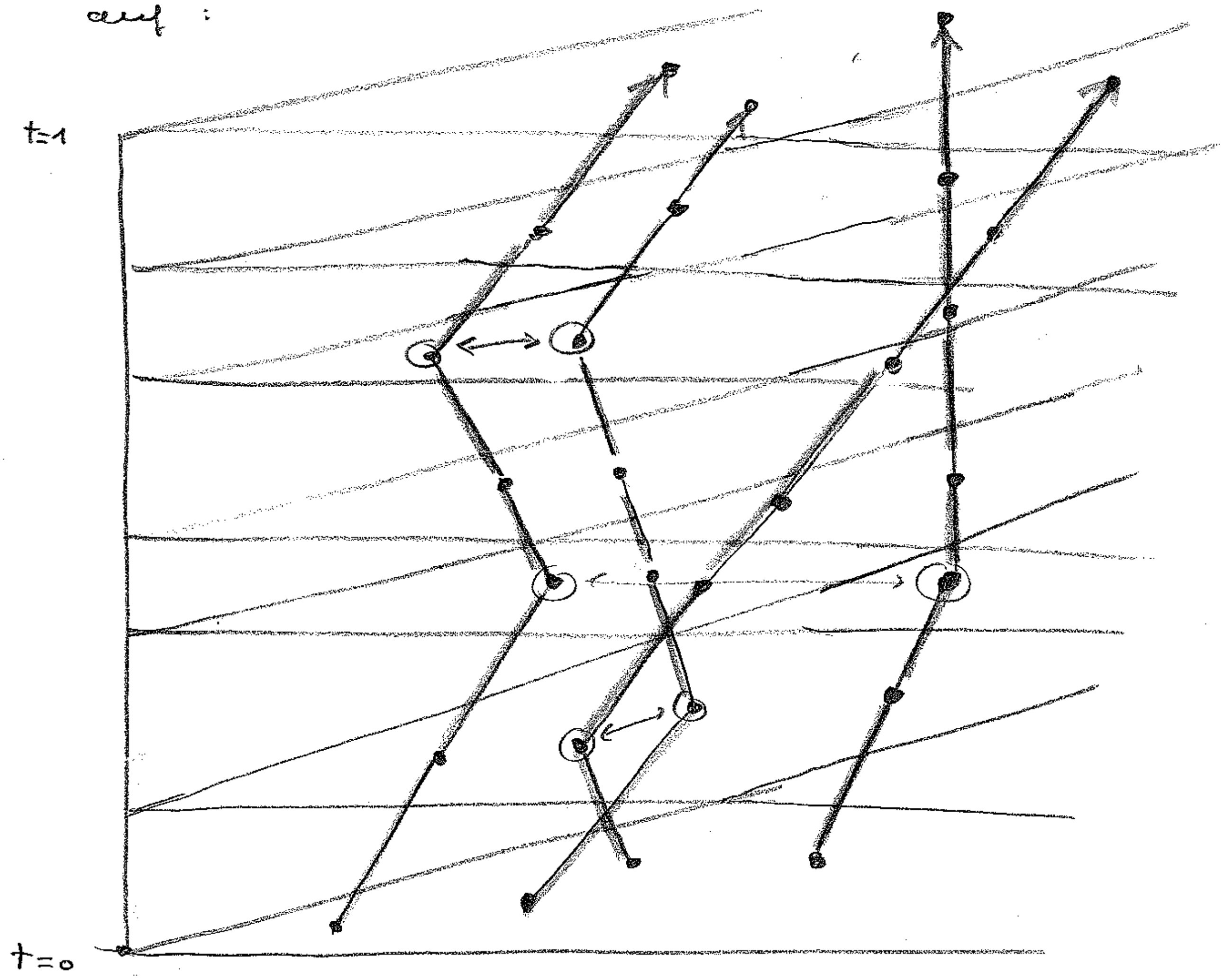
Ein Knick tritt auf, wenn sich durch die Transposition zwei Paare ändern; in diesem Fall sagt man die Transposition liefert ein switch.

Bem. Bei einem switch ändern sich genau zwei Paare, und zwar wie folgt:

- entweder $$\begin{matrix} (\sigma_{i_1}, \sigma_{j_1}) \\ (\sigma_{i_2}, \sigma_{j_2}) \end{matrix} \rightsquigarrow \begin{matrix} (\sigma_{i_2}, \sigma_{j_1}) \\ \updownarrow \\ (\sigma_{i_1}, \sigma_{j_2}) \end{matrix}$$

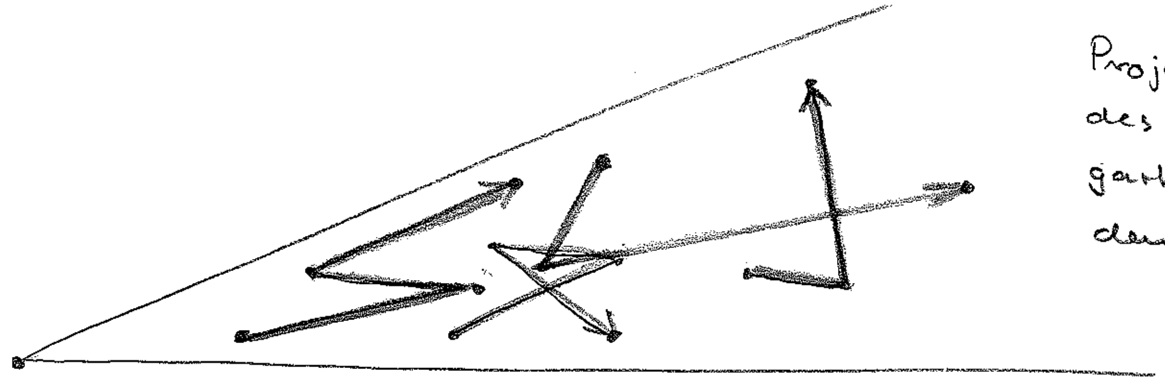
- oder $$\begin{matrix} (\sigma_{i_1}, \sigma_{j_1}) \\ (\sigma_{i_2}, \sigma_{j_2}) \end{matrix} \rightsquigarrow \begin{matrix} (\sigma_{i_1}, \sigma_{j_2}) \\ \updownarrow \\ (\sigma_{i_2}, \sigma_{j_1}) \end{matrix}$$

Im allgemeinen treten mehrere Transpositionen auf:



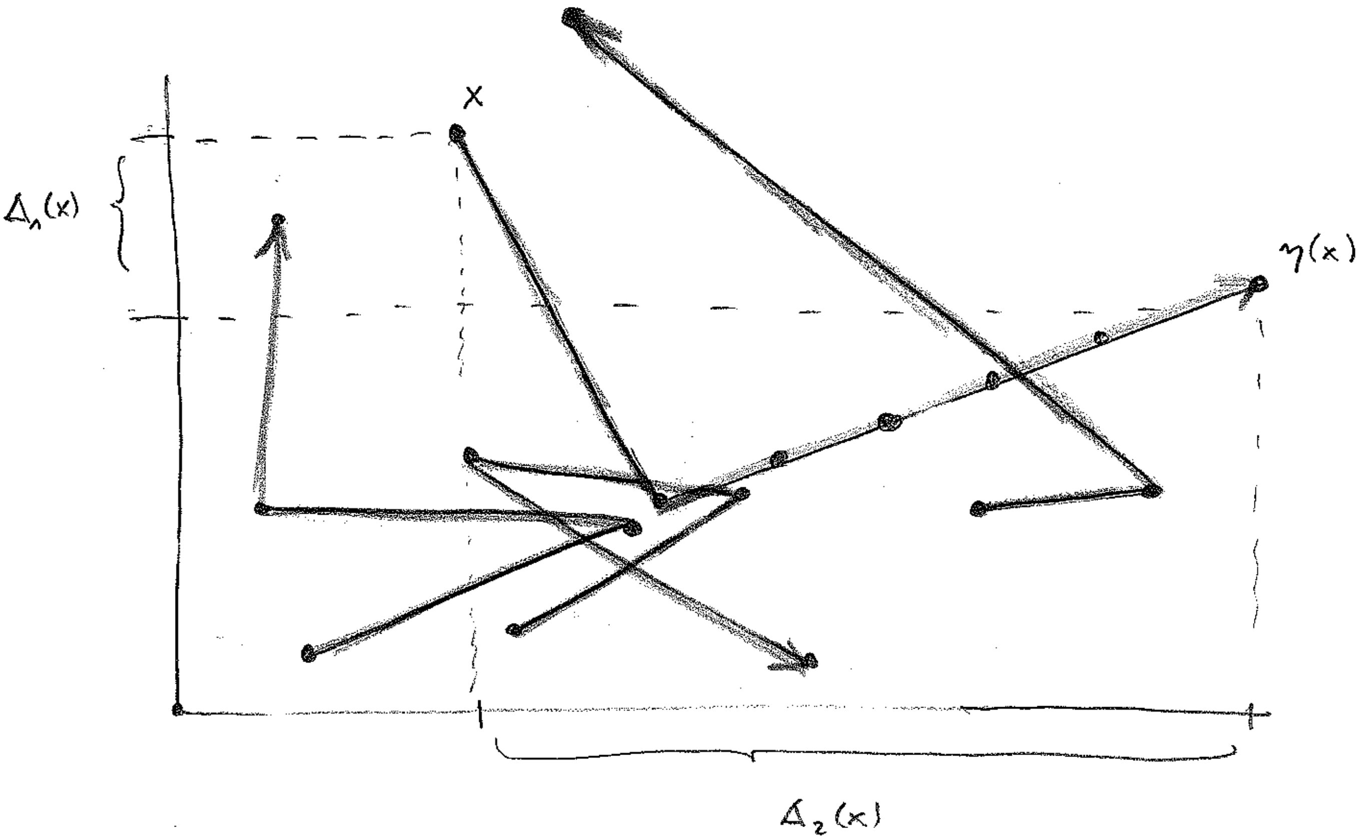
Beispiel mit 5 Transpositionen und 3 Switches.

Notation: Die einzelnen Stränge nennt man Reben (vines), und das ganze Ensemble einen Weingarten (vineyard).



Projektion des Weingartens auf den $\mathbb{R}^2$

Durch die Projektion auf den $\mathbb{R}^2$ kann man eine Abschätzung für den Flaschenhalsabstand zwischen $\mathrm{Dgm}'(f)$ und $\mathrm{Dgm}'(g)$ erhalten



Ein Weingärtner liefert über die Reben eine Bijektion

$$\mathrm{Dgm}'(f) \xrightarrow{\eta_0} \mathrm{Dgm}'(g)$$

Für Punkte $x \in \mathrm{Dgm}'(f)$ gilt:

$$\| x - \eta_0(x) \|_\infty = \max\{|\Delta_1(x)|, |\Delta_2(x)|\}$$

Es sei $x = (f(\sigma), f(\tau))$, i.e. für $t=0$ ist (σ, τ) ein Paar.

Weiterhin seien für $i = 1, .., n$ die t_i mit

$t_0 = 0 < t_1 < t_2 < \dots < t_{n+1} = 1$ die krit. Werte.

Der polygonale Weg von x nach $\eta(x)$ ist dann gegeben durch

$$\omega : [0,1] \to \mathbb{R}^2$$

$$\omega(t) = \big((1-t) f(\sigma_i) + t\, g(\sigma_i),\ (1-t) f(\tau_i) + t\, g(\tau_i)\big)$$

für $t \in [t_i, t_{i+1}]$ und ein Paar (σ_i, τ_i) bzgl. f_t.

$$\Rightarrow \| \omega(t_{i+1}) - \omega(t_i) \|_\infty \leq (t_{i+1} - t_i) \cdot \max \{ \| f(\sigma_i) - g(\sigma_i) \|_\infty, \| f(\tau_i) - g(\tau_i) \|_\infty \}$$

$$\leq (t_{i+1} - t_i) \cdot \| f - g \|_\infty$$

$$\Rightarrow \| \omega(1) - \omega(0) \|_\infty \leq \sum_{i=0}^{n} (t_{i+1} - t_i) \cdot \| f - g \|_\infty = \| f - g \|_\infty$$

$$\rightsquigarrow W_\infty(\mathrm{Dgm}'(f), \mathrm{Dgm}'(g)) = \inf_{\substack{x \xrightarrow{\eta} y \\ \text{bij.}}} \ \sup_{x \in \mathrm{Dgm}'(f)} \| x - \eta(x) \|_\infty$$

$$\leq \sup_{x \in \mathrm{Dgm}'(f)} \| x - \eta_0(x) \|_\infty$$

$$\leq \| f - g \|_\infty$$

q.e.d.

6 Der Wassersteinabstand

Bem: der Flaschenhalsabstand wird durch den Abstand von einem Punktepaar realisiert; und er kontrolliert nicht mehr das Verhalten von Punktepaaren, die enger zusammenliegen.

$\leadsto$

Def. Für bijektive endl. Teilmengen X und Y in $\mathbb{R}^2$ setze für $q \in (0,\infty)$

$$W_q(X,Y) = \Big(\inf_{\substack{\eta: X \to Y \\ \text{bij.}}} \sum_{x \in X} \| x - \eta(x) \|_\infty^q \Big)^{1/q}$$

Der Wassersteinabstand zweier Persistenzdiagramme $Dgm'(f)$ und $Dgm'(g)$ ist dann

$$W_q(Dgm'(f), Dgm'(g))$$

Prop Es gilt

$$\lim_{q\to\infty} w_q(X,Y) = w_\infty(X,Y)$$

sowie:
- $w_q(X,Y) = 0 \iff X = Y$
- $w_q(X,Y) = w_q(Y,X)$
- $w_q(X,Z) \leq w_q(X,Y) + w_q(Y,Z)$

für drei bijektive, endliche Teilmengen X, Y, Z in $\mathbb{R}^2$.

Bem: auch für w_q kann man ein Stabilitätssatz herleiten