



Entscheidungstheorie Teil 2

Thomas Kämpke

Inhalt

- Präferenzrelation
- Referenzpunktansatz
- Referenzpunktmethode (Zusammenfassung)
- Distanzfunktion
- Design von PCR Primerpaaren
- Vorwärtsprimer p
- Rückwärtsprimer q
- Primerwechselwirkung



Präferenzrelation (1/12)

Definition

Eine (strikte) **Präferenzrelation** \prec ist eine binäre Relation,
die asymmetrisch und negativ transitiv ist.

für deterministische und
stochastische Situationen!

Asymmetrie

$$a \prec b \Rightarrow \neg (b \prec a)$$

Negative Transitivität

$$\neg (a \prec b) \quad \text{und} \quad \neg (b \prec c) \Rightarrow \neg (a \prec c)$$

Präferenzrelation (2/12)

Beispiel

Eine asymmetrische und transitive, aber nicht negativ transitive Relation ist die komponentenweise Ordnung auf \mathbb{R}^n mit Verschärfung in mind. einer Koordinate.

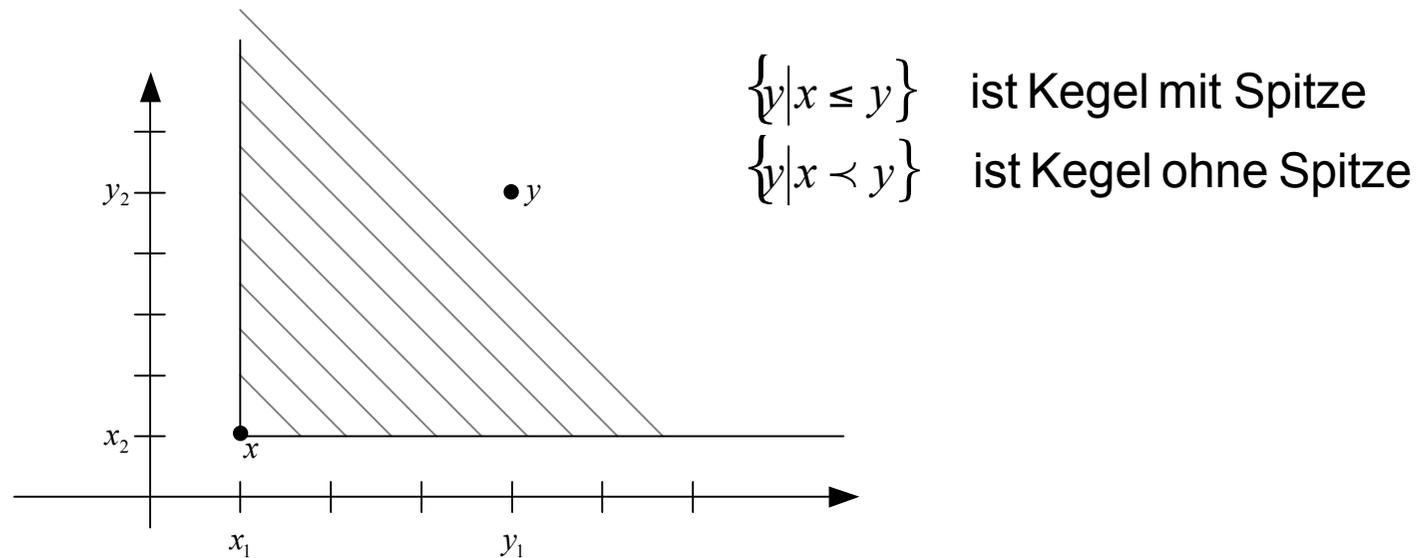
Komponentenweise Ordnung

Für $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ und $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ ist $x \leq y \Leftrightarrow x_i \leq y_i \quad i = 1, \dots, n$

Komponentenweise Ordnung mit Verschärfung

$x \prec y \Leftrightarrow x_i \leq y_i \quad i = 1, \dots, n \quad \text{und} \quad x_j < y_j \quad \text{für mindestens einen Index } j$

Präferenzrelation (3/12)



$<$ ist **asymmetrisch**, denn

$$x \prec y \Leftrightarrow x_i \leq y_i \quad i = 1, \dots, n \quad \text{und} \quad x_j < y_j \quad \text{für ein } j$$

Wäre $y \prec x$ so $y_i \leq x_i \quad i = 1, \dots, n$, d.h. insbesondere

$$y_j \leq x_j < y_j \quad \textbf{Widerspruch}$$

$$\Rightarrow \neg(y \prec x)$$

Präferenzrelation (4/12)

\prec ist **transitiv**, denn

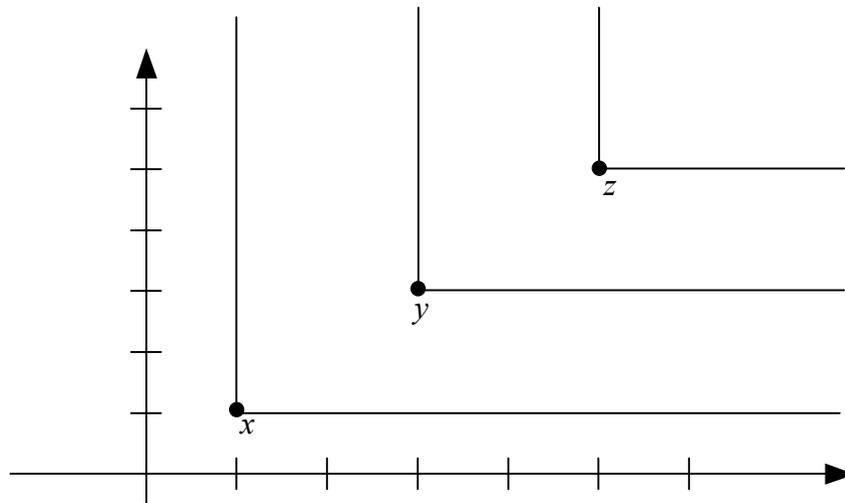
$$\begin{array}{l}
 x \prec y \quad \text{und} \quad y \prec z \Rightarrow x_i \leq y_i \leq z_i \quad i = 1, \dots, n \\
 \text{und} \quad x_j < y_j \leq z_j \quad \text{also} \quad x_j < z_j \\
 \Rightarrow x \prec z
 \end{array}$$

\prec ist **nicht negativ transitiv**, denn

$$\begin{array}{l}
 x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 5 \end{pmatrix} \quad z = \begin{pmatrix} 3 \\ 4 \end{pmatrix} \\
 \neg(x \prec y) \quad \text{und} \quad \neg(y \prec z) \quad \text{aber} \quad x \prec z
 \end{array}$$

Präferenzrelation (5/12)

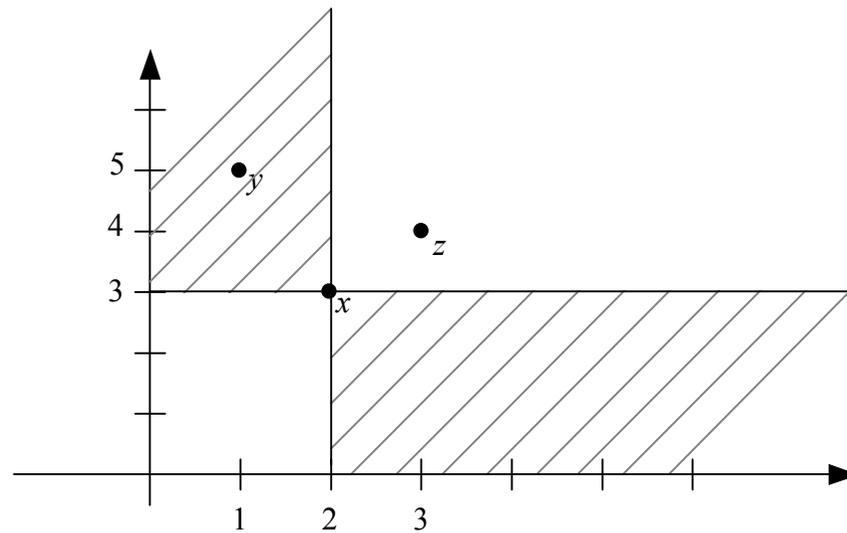
Veranschaulichung Transitivität



$$\left. \begin{array}{l} \text{Kegel von } y \subseteq \text{Kegel von } x \\ \text{Kegel von } z \subseteq \text{Kegel von } y \end{array} \right\} \Rightarrow \text{Kegel von } z \subseteq \text{Kegel von } x$$

Präferenzrelation (6/12)

Veranschaulichung Fehlen negativer Transitivität



Bereich des Fehlens einer
< Relation

Präferenzrelation (7/12)

Eine asymmetrische und negativ transitive Relation \prec ist transitiv, denn

$$a \prec b \quad \text{und} \quad b \prec c \Rightarrow a \prec c$$

$$\text{Annahme: } \left. \begin{array}{l} \neg(a \prec c) \\ b \prec c \Rightarrow \neg(c \prec b) \end{array} \right\} \begin{array}{l} \text{negativ transitiv} \\ \Rightarrow \end{array} \neg(a \prec b) \quad \text{Widerspruch zu } a \prec b$$

Präferenzrelation (8/12)

Definition

Die zu einer (strikten) Präferenzrelation \prec gehörende **Indifferenzrelation** \sim ist definiert als das Fehlen von Präferenzen d.h.

$$a \sim b : \Leftrightarrow \neg (a \prec b) \quad \text{und} \quad \neg (b \prec a)$$

Von der Intuition fast die Indifferenzrelation ein bewusstes Gleichwertigkeitsurteil und ein „sich nicht entscheiden können“ zusammen!

Präferenzrelation (10/12)

Definition

Präferenz-Indifferenzrelation oder schwache Präferenzordnung

$$a \leq b \Leftrightarrow a \prec b \quad \text{oder} \quad a \sim b$$

Alternative Formulierung

$$\begin{aligned} a \leq b &\Leftrightarrow a \prec b \quad \text{oder} \quad (\neg(a \prec b) \quad \text{und} \quad \neg(b \prec a)) \\ &\Leftrightarrow a \prec b \quad \text{und} \quad \neg(b \prec a) \end{aligned}$$

\leq ist **transitiv!**

Präferenzrelation (11/12)

Für eine strikte Präferenzrelation \prec über einer Alternativenmenge \mathcal{A} (abzählbar) gibt es reellwertige, ordnungserhaltende Funktionen, d.h. $f: \mathcal{A} \rightarrow \mathbb{R}$ mit

$$a \prec b \Leftrightarrow f(a) < f(b).$$

Die „Schwierigkeit“ einer strikten Präferenzordnung liegt in Funktionen, die Ordnung ist dann die gewöhnliche Ordnung reeller Zahlen!

$$\begin{aligned} a \sim b \quad \text{ist} \quad & \neg(a \prec b) \quad \text{und} \quad \neg(b \prec a) \quad , \text{d.h.} \\ & \neg(f(a) < f(b)) \quad \text{und} \quad \neg(f(b) < f(a)) \quad , \text{d.h.} \\ & f(a) = f(b) \end{aligned}$$

Die Indifferenz wird als Gleichheit dargestellt.

Präferenzrelation (12/12)

Gleichheit reeller Zahlen ist transitiv $f(a) = f(b)$ und $f(b) = f(c) \Rightarrow f(a) = f(c)$, d.h. Indifferenzrelation ist transitiv, sofern strikte Ordnung mit reellwertigen Funktionen darstellbar.

dies im nachhinein Erklärung für entsprechenden Aufwand!

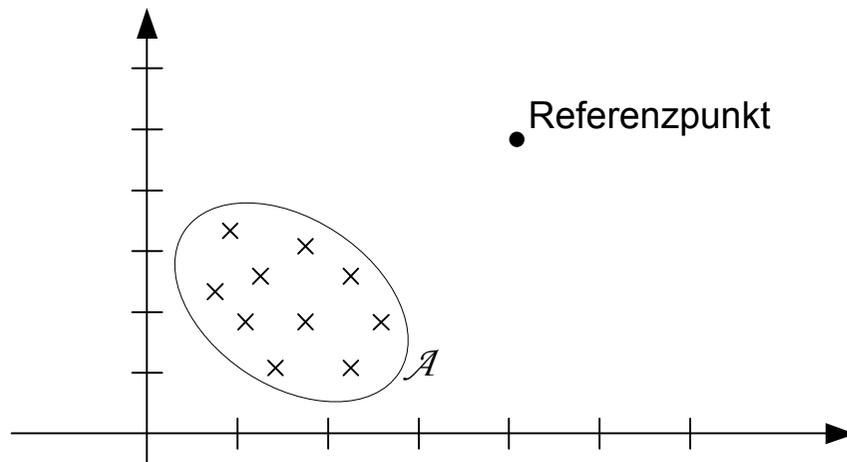
Funktionen, die Präferenzrelationen darstellen, nennt man **Präferenzfunktionen**.

deterministischer und stochastischer Fall gleich behandelt!

Referenzpunktansatz (1/5)

Wie konstruiert man Präferenzfunktionen?

Für deterministische Situationen gibt es den **Referenzpunktansatz**



Idee

1. Als Referenzpunkt wird ein ideal guter oder utopischer Punkt gewählt
2. Es wird eine Alternative gesucht, die dem Referenzpunkt am nächsten liegt. → „beste Alternative“

Referenzpunktansatz (2/5)

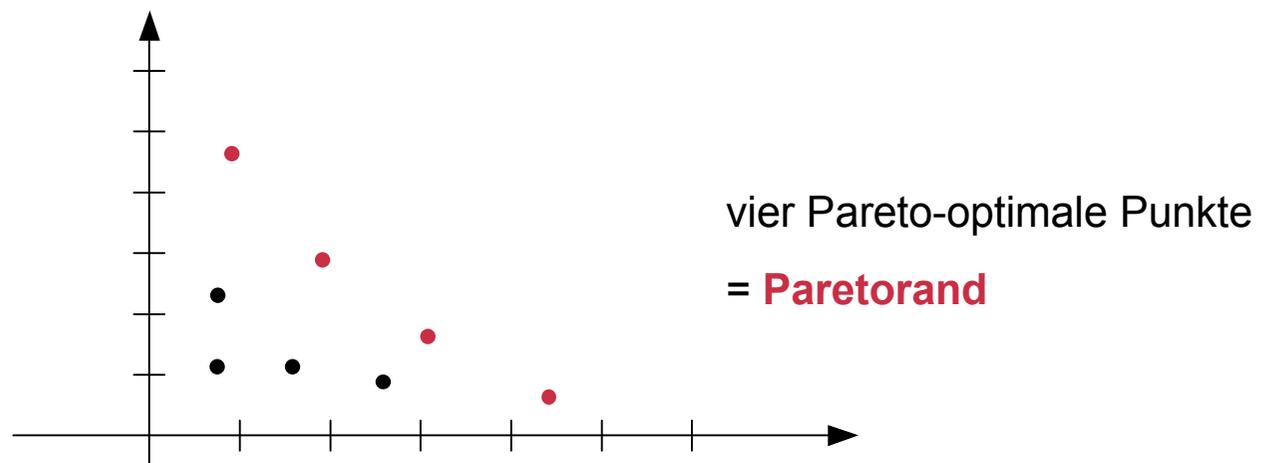
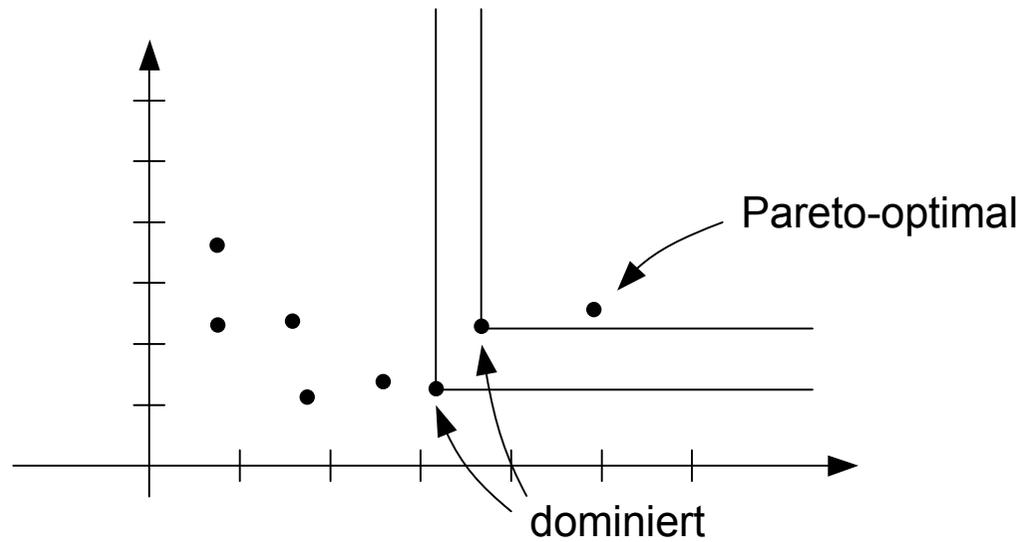
Vorteil

Man kann auch die komponentenweise Ordnung darstellen, wenn die Indifferenzen speziell beachtet werden. Dazu Begriff der **Dominanz oder Pareto-Optimalität**

Alternative $a \in \mathcal{A}$ ist **Pareto-optimal**, wenn es kein $b \in \mathcal{A}$ gibt mit $a < b$;

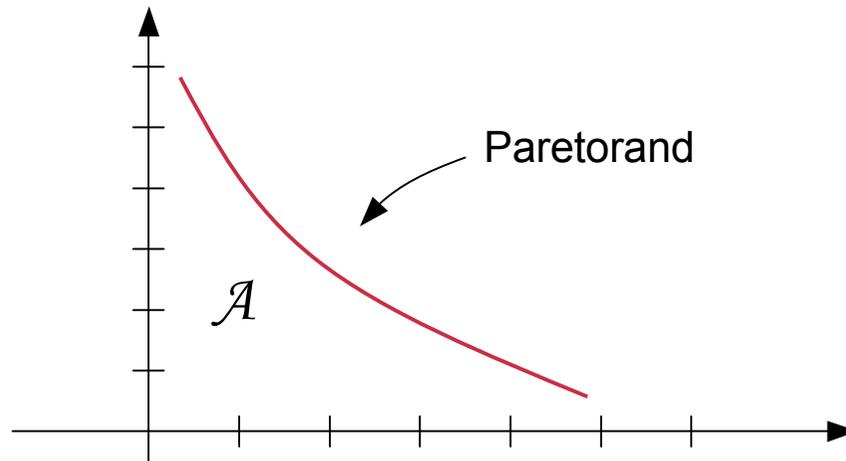
$<$ komponentenweise Ordnung. a heisst **dominiert** wenn es $b \in \mathcal{A}$, $b \neq a$ gibt mit $a < b$

Referenzpunktansatz (3/5)



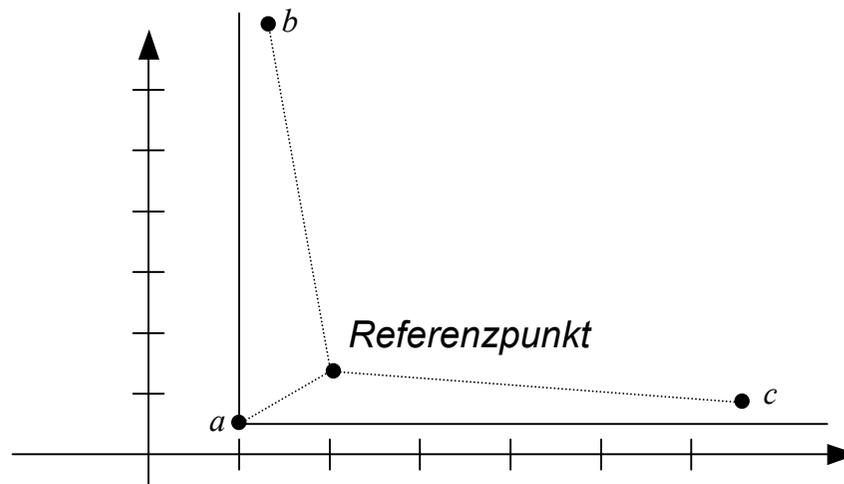
Referenzpunktansatz (4/5)

Im kontinuierlichen Fall



Referenzpunktansatz (5/5)

Beim Referenzpunktansatz kann eine dominierte Alternative gewählt werden, wenn Referenzpunkt ungeschickt gewählt:



$$\mathcal{A} = \{a, b, c\} \text{ aber } \mathcal{A}_{\text{pareto}} = \{b, c\}$$

Referenzpunktmethode (Zusammenfassung) (1/1)

1. Für Alternativemenge $\mathcal{A} \subseteq \mathbb{R}^n$ Bestimmung des Paretorands $\mathcal{A}_{\text{pareto}}$
2. Wahl Referenzpunkt $Ref \in \mathbb{R}^n$
3. Bestimmung argmin $a \in \mathcal{A}_{\text{pareto}} \text{ Dist}(a, Ref)$

Zu beachten

- $a \leq Ref \quad \forall a \in \mathcal{A}_{\text{pareto}}$
- $Dist$ kann Euklidischer Abstand o.a. sein.
Diese Distanz ist, bis auf das Vorzeichen, eine ordnungserhaltende Funktion

$$a \prec b \Leftrightarrow Dist(a, Ref) > Dist(b, Ref)$$

$$\Leftrightarrow -Dist(a, Ref) < -Dist(b, Ref)$$

$$\Leftrightarrow \underset{\parallel}{f(a)} < \underset{\parallel}{f(b)}$$

Distanzfunktion (1/7)

$Dist(a, Ref)$ kann Euklidischer Abstand, d.h.

$$\begin{aligned} Dist(x, x_{Ref}) &= \|x - x_{Ref}\| = \|x - x_{Ref}\|_2 \\ &= \sqrt{(x_1 - x_{Ref,1})^2 + \dots + (x_n - x_{Ref,n})^2} \end{aligned}$$

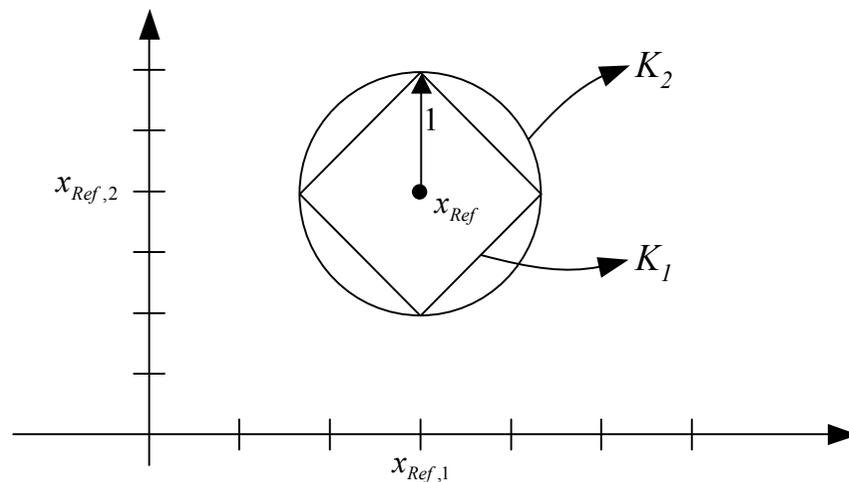
$$\text{mit } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad x_{Ref} = \begin{pmatrix} x_{Ref,1} \\ \vdots \\ x_{Ref,n} \end{pmatrix}$$

Aber auch andere (Lipschitz-)Normen, vor allem die 1-Norm

$$Dist(x, x_{Ref}) = \|x - x_{Ref}\|_1 = |x_1 - x_{Ref,1}| + \dots + |x_n - x_{Ref,n}|$$

Für $n=2$ sind Punkte gleichen Abstands Kreise für die Euklidische Norm und Rauten für die 1-Norm

Distanzfunktion (2/7)

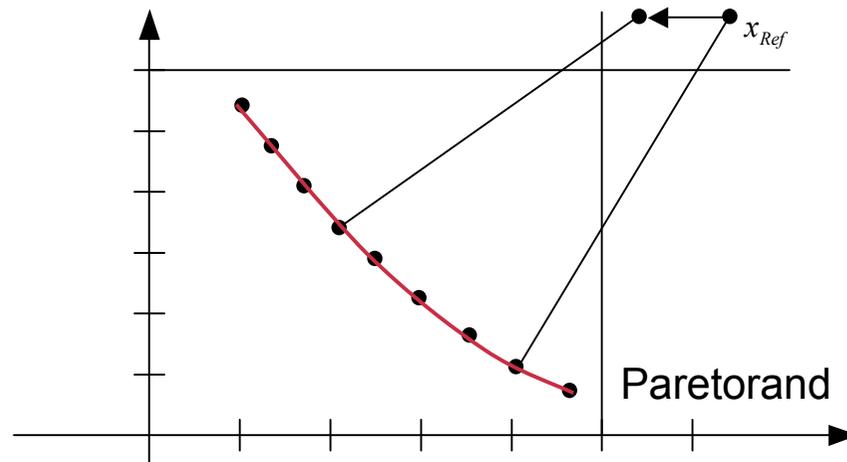


$$K_2 = \left\{ x \mid \|x - x_{Ref}\|_2 = 1 \right\}$$

$$K_1 = \left\{ x \mid \|x - x_{Ref}\|_1 = 1 \right\}$$

Der Unterschied zwischen 2-Norm und 1-Norm in Bezug auf den Referenzpunkt ist nicht so sehr die „Kugel“, sondern die Robustheit der ausgewählten Alternative.

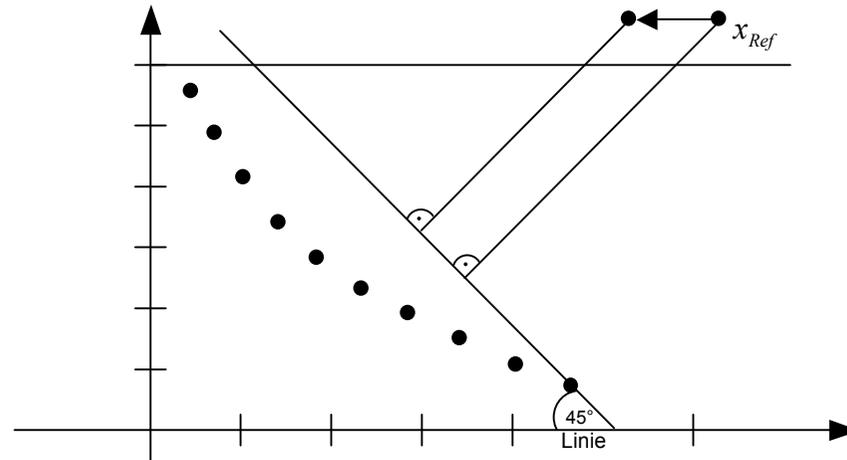
Distanzfunktion (3/7)



Wird der Referenzpunkt verschoben, so dass er auch in neuer Lage den gesamten Paretorand dominiert, kann sich die beste Alternative im Sinne der $\|\dots\|_2$ -Norm ändern.

Bei der $\|\dots\|_1$ -Norm bleibt die beste Alternative bestehen!

Distanzfunktion (4/7)



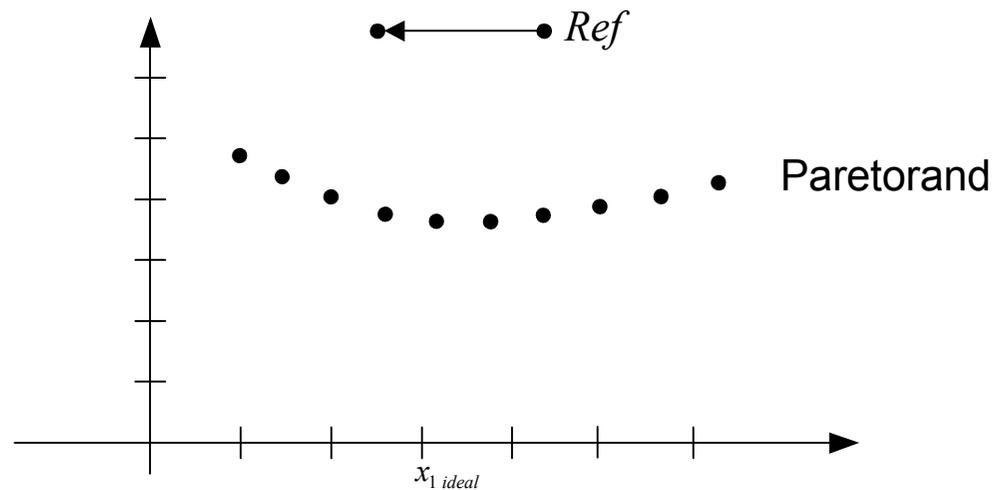
Von der Raute kommt nur die südwestliche 45° Linie zum Tragen. Die beste Alternative ist für alle sinnvollen Referenzpunkte identisch.

Die Distanzwerte ändern sich, sofern sich der Referenzpunkt ändert, nicht aber die beste Alternative.

Warum also 1-Norm beim Referenzpunktansatz?

Distanzfunktion (5/7)

1. Nicht alle Kriterien sind monoton, z.B. Raumtemperatur (ideal $18^\circ, \dots, 20^\circ\text{C}$) d.h. Paretorand sieht anders aus



1. Kriterium ist nicht monoton
2. Kriterium ist monoton

Die Änderung des Referenzpunktes beschreibt tatsächlich eine andere Alternativenwahl

Distanzfunktion (6/7)

2. Die Distanzfunktion kann mit Gewichten versetzt werden

$$\begin{aligned} \|x - x_{Ref}\|_1 &\rightarrow \|x - x_{Ref}\|_{1,w} \\ &= w_1 * |x_1 - x_{Ref,1}| + w_2 * |x_2 - x_{Ref,2}| + \dots + w_n * |x_n - x_{Ref,n}| \end{aligned}$$

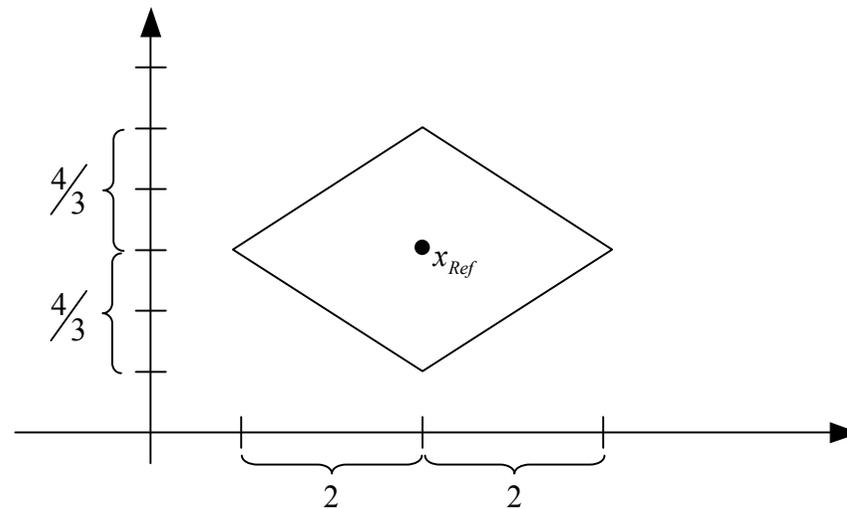
Beispiel

$$\begin{aligned} \|x - x_{Ref}\|_{1,w} \\ &= 2 * |x_1 - x_{Ref,1}| + 3 * |x_2 - x_{Ref,2}| \end{aligned}$$

Distanzfunktion (7/7)

Beispiel

$$\begin{aligned} & \|x - x_{Ref}\|_{1,w} \\ &= 2 * |x_1 - x_{Ref,1}| + 3 * |x_2 - x_{Ref,2}| \end{aligned}$$



Raute wird gestaucht

Design von PCR Primerpaaren* (1/3)

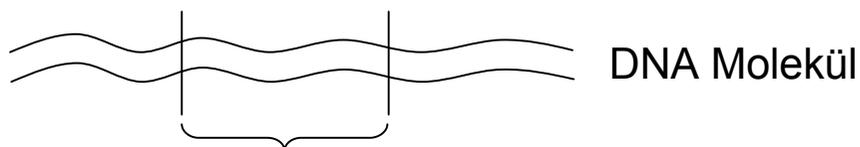
PCR = polymerase chain reaction

Ziel

Bestimmung eines Abschnitts „eines“ DNA Moleküls bzw. eines in geringer Konzentration vorliegenden DNA Moleküls.

Vorgehen

- Abschnitt wird festgelegt
- + Abschnitt wird vervielfacht (→ Konzentrationserhöhung)
- Abschnitt wird analysiert (erst aufgrund Konzentrationserhöhung möglich)



DNA Molekül

interessierender Abschnitt
z.B. 1400bp lang

Exon eines Gens

Teil eines Exons

Teil eines Introns, d.h.

„nicht kodierenden Bereichs“

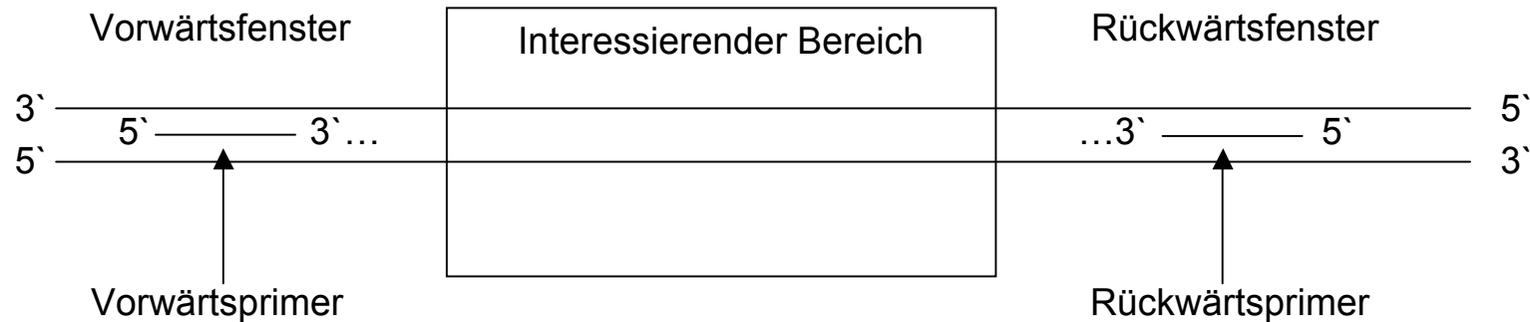
*Quelle: T. Kämpke, „The reference point method in primer design“, in: PCR Primer Design, Humana Press, Totowa, NJ, 2007.

Design von PCR Primerpaaren (Video)



**Polymerase Chain
Reaction**

Design von PCR Primerpaaren (2/3)



DNA Stränge **gerichtet**

Primer immer paarweise, jeder Primer ist **einsträngig**

Größenverhältnisse

Primerlänge z.B. 19

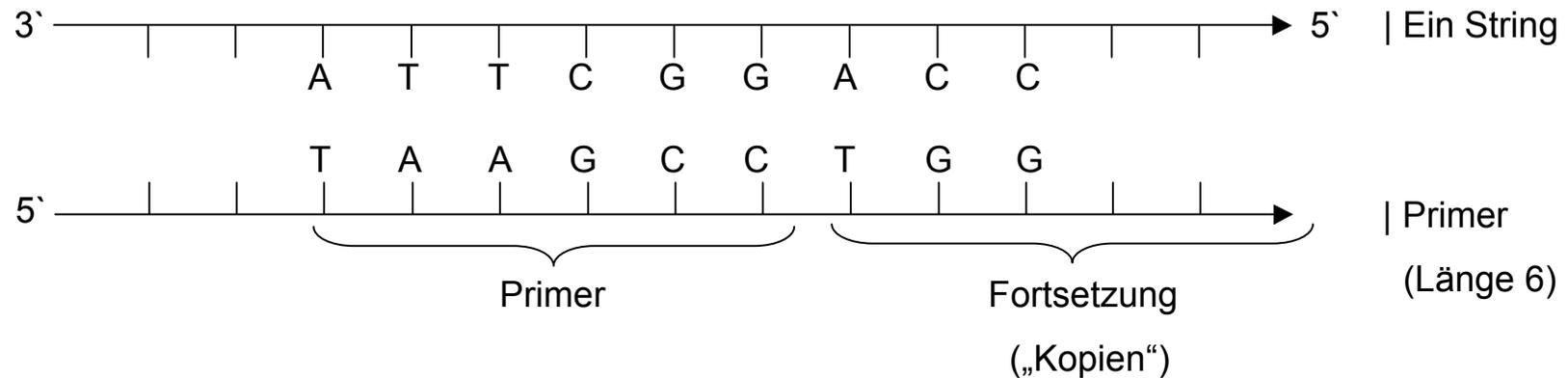
Fenster z.B. 39

interessierender

Bereiche z.B. 1400

Primer heißen Primer, weil sie „prime“ (der Anfang) für einen naturnahen Kopierprozess sind.

Design von PCR Primerpaaren (3/3)



Fortsetzen wird durch ein Enzym, die Taq Polymerase, ermöglicht. Viele chemische Tricks und Edukte notwendig, Temperaturzyklen.

Kandidaten für Primerpaare durch Abschnitte in Fenster erzeugt.

12 Kriterien für gute Primerpaare (p, q)

fünf für jeden Primer

zwei für das Zusammenwirken

Vorwärtsprimer p (1/1)

GGATTGATAATGTAATAGG

Länge $ p $	19
GC(p) [in%]	32
$T_m(p)$ [in°C]	50
sa(p)	12
sea(p)	0

$T_m(p)$:
kritische Temperatur

$$T_m(p) = 2 * \left(\begin{array}{l} 2\# G(p) + 2\# C(p) \\ +1\# T(p) + 1\# A(p) \end{array} \right)$$

sa	self anreal	}	sollen verhindern, dass sich Primer an sich selbst anlagern („primer dimer“)
sea	self end anreal		

Rückwärtsprimer q (1/1)

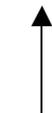
C A T T A T G G G T G G T A T G T T G G

Länge $ q $	20
GC(q) [in%]	45
Tm(q) [in°C]	58
sa(q)	20
sea(q)	4

sea(q):

5` - C A T T ...
3` - G G ...

... G G - 3`
... A C - 5`

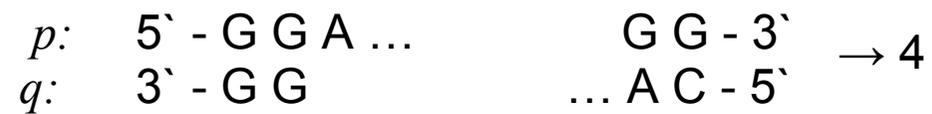


wird im end anreal nicht gezählt, da schon „am Anfang“ erfasst.

Primerwechselwirkung (1/4)

$pa(p,q)$	16
$pea(p,q)$	4

$pea(p,q)$



Primerwechselwirkung (2/4)

Bewertung jedes Paares (p, q) im \mathbb{R}^{12} mittels Vektoren

$$x(p, q) = x = \begin{pmatrix} |p| \\ |q| \\ GC(p) \\ GC(q) \\ Tm(p) \\ Tm(q) \\ sa(p) \\ sa(q) \\ sea(p) \\ sea(q) \\ pa(p, q) \\ pea(p, q) \end{pmatrix}$$

$$Ref = \begin{pmatrix} I_{\text{vorw}} \\ I_{\text{rück}} \\ GC_{\text{vorw}} \\ GC_{\text{rück}} \\ Tm_{\text{vorw}} \\ Tm_{\text{rück}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Die ersten sechs
Koordinaten werden manuell
vergeben;

$$Tm_{\text{vorw}} = Tm_{\text{rück}}$$

Primerwechselwirkung (3/4)

Nicht alle Einzelkriterien sind monoton z.B. gilt bei Primerlänge **nicht**:
„je länger desto besser“ oder „je kürzer desto besser“

$$\text{Mit } \text{Dist}(x, \text{Ref}) = \sum_{i=1}^{12} w_i * |x_i - \text{Ref}_i|$$

Anstelle von Euklidscher Distanz

$$w_1 = w_2 = 0.5$$

$$w_3 = w_4 = 1$$

$$w_5 = w_6 = 1$$

$$w_7 = w_8 = 0.1$$

$$w_9 = w_{10} = 0.2$$

$$w_{11} = 0.1$$

$$w_{12} = 0.2$$

Bestimme $\text{argmin}(p, q) \text{Dist}(x(p, q), \text{Ref})$

Primerwechselwirkung (4/4)

Verfahren wird durch Zusatzkriterien beschleunigt, z.B.

$$\lambda_l \leq |p|, \quad |q| \leq \lambda_u$$

Keine Garantie auf Ergebnisqualität:

$$\text{„}\frac{1}{3} - \frac{1}{3} - \frac{1}{3}\text{“ in Bioinformatik}$$

⇒ PCR Primerdesign:

eine der häufigsten Anwendungen von Entscheidungstheorie Verfahren,
wahrscheinlich die häufigste.

Viele Varianten

- Multiplex PCR (mehr als 2 Primerpaare)
- Biochipdesign