„Euclidean minimum spanning trees and bichromatic closest pairs "

Seminar im SS 2003
ILKD Wagner
Fakultät für Informatik
Universität Karlsruhe

Betreuer :
Alexander Wolff
Marc Benkert

Bearbeitet von Khoder El-Zein

Euklidischer Minimaler Spannbaum

Gegeben:

- $V = \{v_1, v_2, .., v_N\}$ - Menge con Punkten (Knoten) im d-dimensionalen Euklidischen Raum E^d , N = card(V).
- L : $\{v_i, v_j\} \rightarrow [0, \infty)$ - Kantenbewertung (Euklidischer Abstand)

Gesucht:

Euklidischer Minimaler Spannbaum (EMST) (V, E‘):

- (V, E‘) ist ungerichteter Baum
- card(E‘) = N - 1
- Summe der Kantenlängen von (V, E‘) hat den kleinstmöglichen Wert.

Bichromatic closest pairs (BCP)

Gegeben

Eine Menge S von n roten und m grünen Punkten im d-dimensionalen Euklidischen Raum E^d .

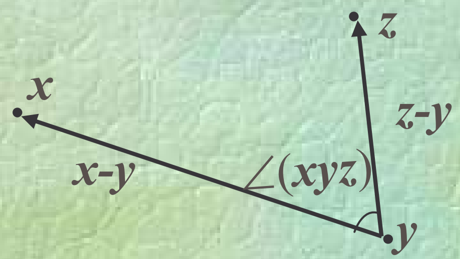
Gesucht

Ein bichromatisches nächstes Paar:

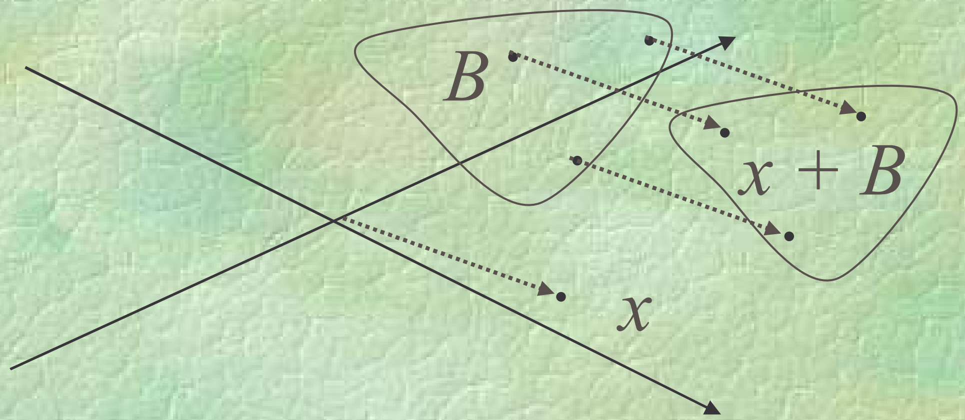
ein roter Punkt ***r*** und ein grüner ***g*** so, dass der Abstand zwischen ***r*** und ***g*** minimal unter allen rot-grün Paaren ist.

Notation

- Winkel zwischen $\boldsymbol{x}$ und $\boldsymbol{y}$ ist $\angle(\boldsymbol{x}, \boldsymbol{y}) = \arccos \boldsymbol{x}^T\boldsymbol{y}/(\|\boldsymbol{x}\|.\|\boldsymbol{y}\|)$.
- $\angle\,(\boldsymbol{xyz}) := \angle\,(\boldsymbol{z}\text{-}\boldsymbol{y}, \boldsymbol{x}\text{-}\boldsymbol{y})$.

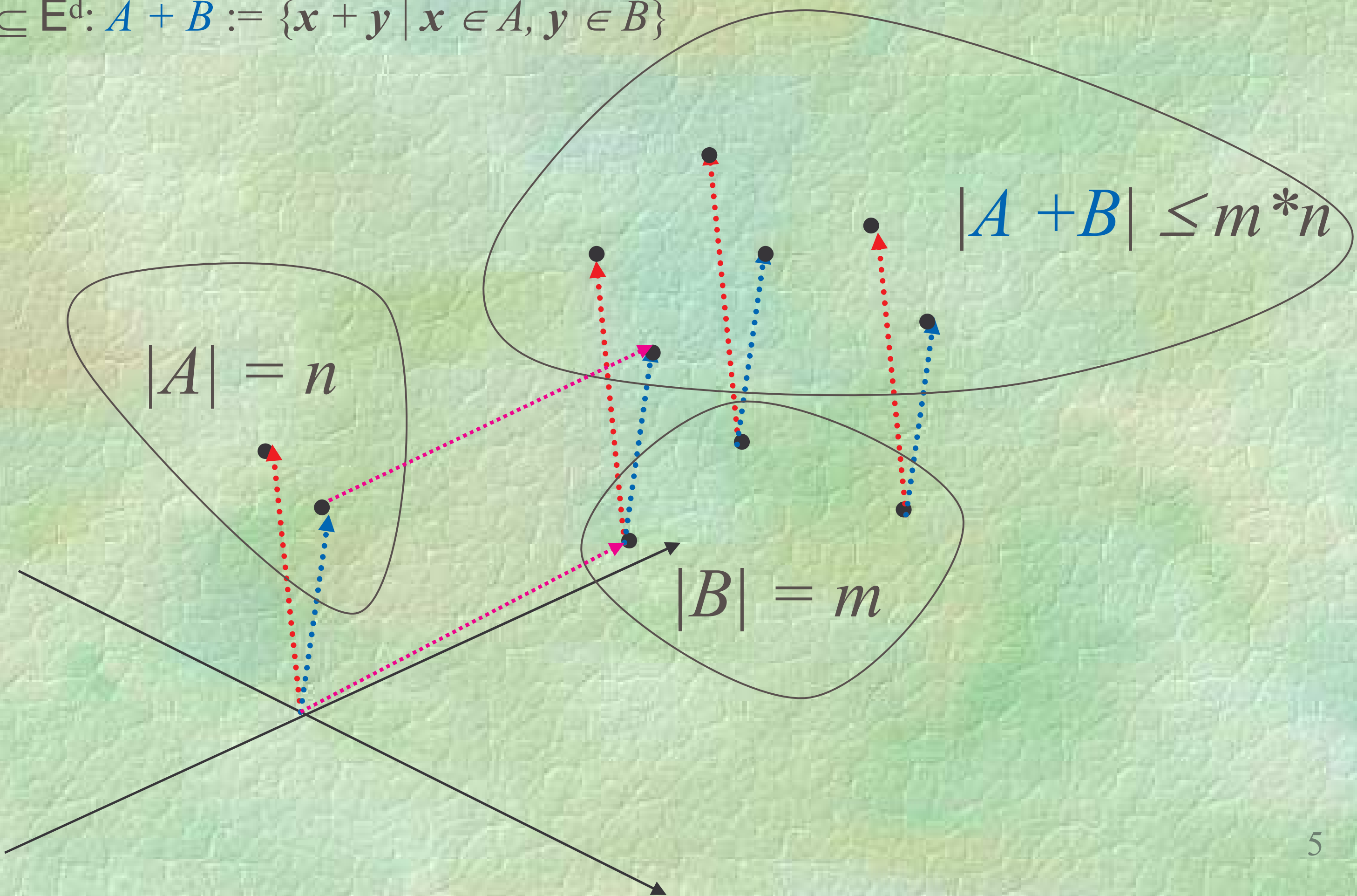


- $\boldsymbol{x} + B := \{\boldsymbol{x}\} + B$.



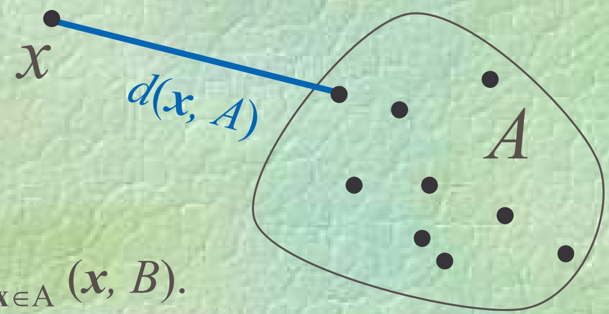
Notation

- $A, B \subseteq \mathsf{E}^d$: $A + B := \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in A, \boldsymbol{y} \in B\}$

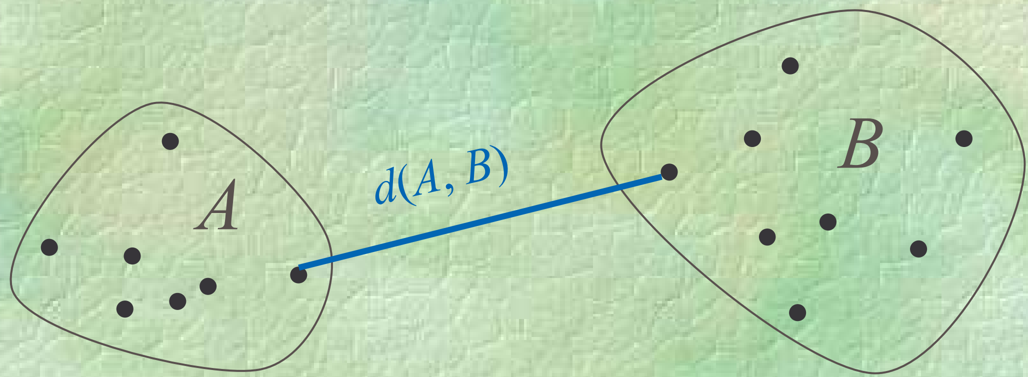


Notation

- A endlich $\subseteq E^d$: $d(\mathbf{x}, \mathrm{A}) := \min_{y \in A} d(\boldsymbol{x}, \boldsymbol{y})$.

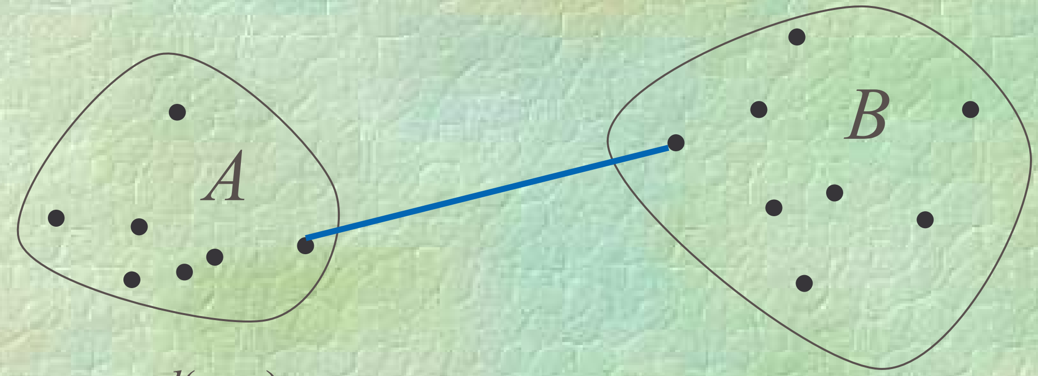


- $A, B \subseteq E^d$: $d(A, B) := \min_{\mathbf{x} \in \mathrm{A}} (\boldsymbol{x}, B)$.

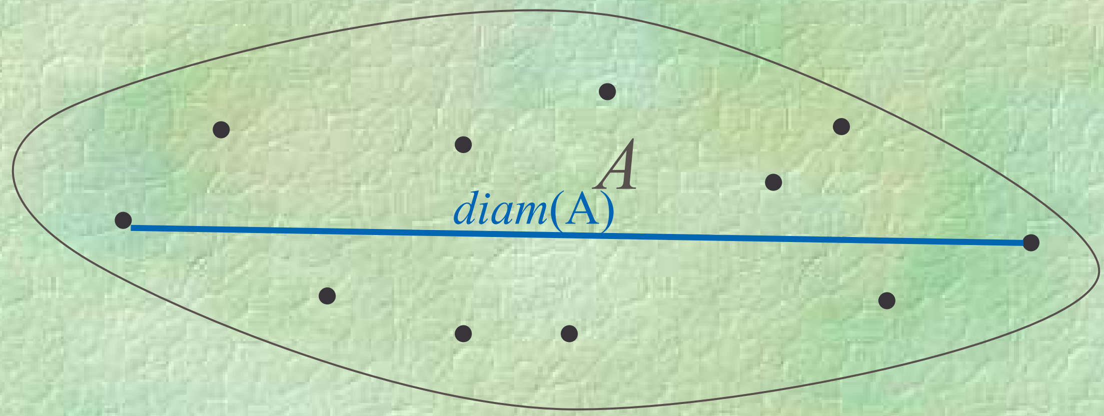


Notation

- nächstes (A, B)-Paar: $(\boldsymbol{x}, \boldsymbol{y})$ mit $\boldsymbol{x} \in A$, $\boldsymbol{y} \in B$ und $d(\boldsymbol{x}, \boldsymbol{y}) = d(A, B)$.

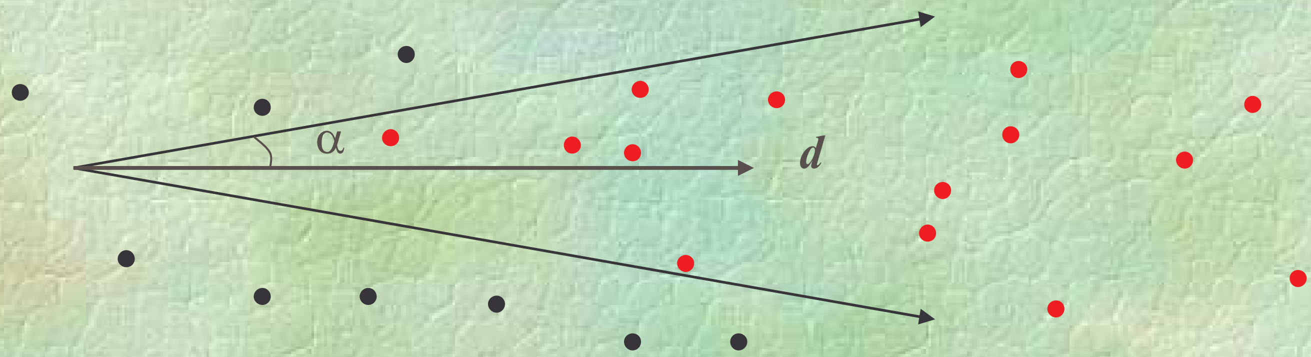


- $diam(\mathrm{A}) := \max_{x,\, y \in A} d(\boldsymbol{x}, \boldsymbol{y})$.



Notation

- $Cone(\boldsymbol{d}, \alpha)$:= $\{ \boldsymbol{x} \in E^d \,/\, \angle(\boldsymbol{x}, \boldsymbol{d}) \leq \alpha \}$ $\boldsymbol{d} \in E^d$ ein Einheitsvektor und α < 90° ein Winkel.

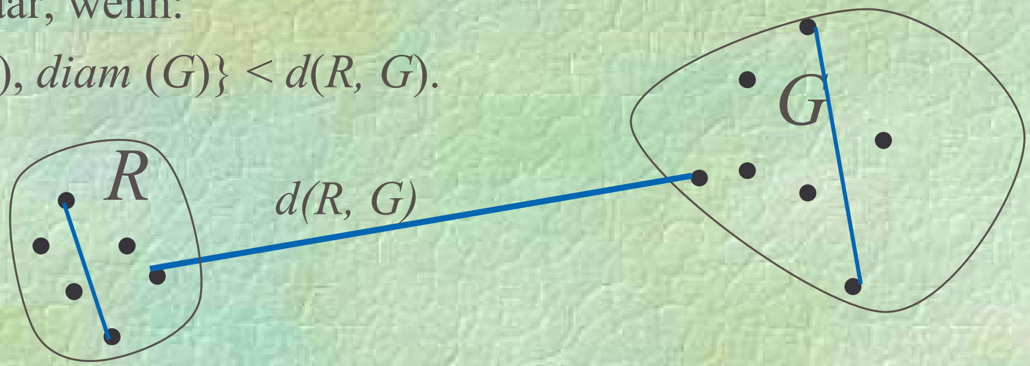


- Ab jetzt: α fest, $0 < \alpha < \alpha_0$, mit $\tan 2\alpha < \cos 2\alpha$.
  (α_0= (arcsin (√(5) - 1)/2)/2 ca. 19.08°).

Notation

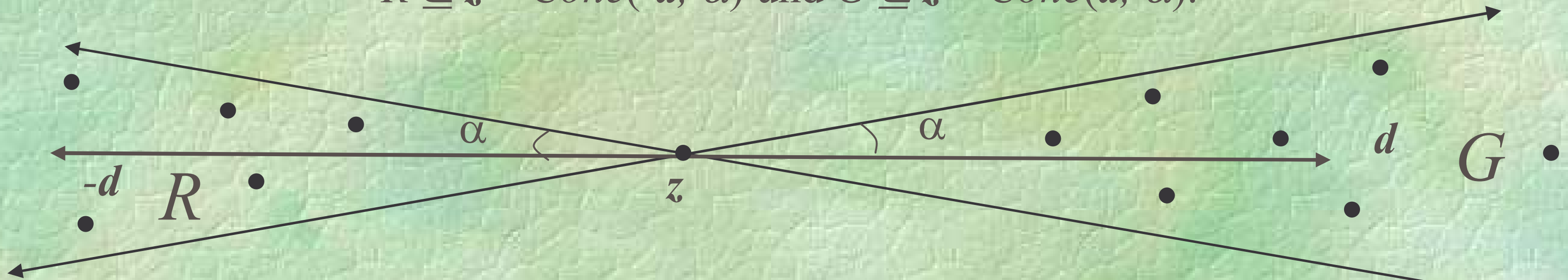
- (R, G) ist ein *stark separiertes* Paar, wenn:

$$\max \{diam(R), diam(G)\} < d(R, G).$$



- (R, G) ist ein *α-separiertes* Paar: $\exists\, \boldsymbol{z} \in \mathsf{E}^d$, $\exists\, \boldsymbol{d} \in \mathsf{E}^d$ Einheitsvektor, wenn:

$$R \subseteq \boldsymbol{z} + Cone(-\boldsymbol{d}, \alpha) \text{ und } G \subseteq \boldsymbol{z} + Cone(\boldsymbol{d}, \alpha).$$



Stark separierte Paare und EMST

Lemma 1.

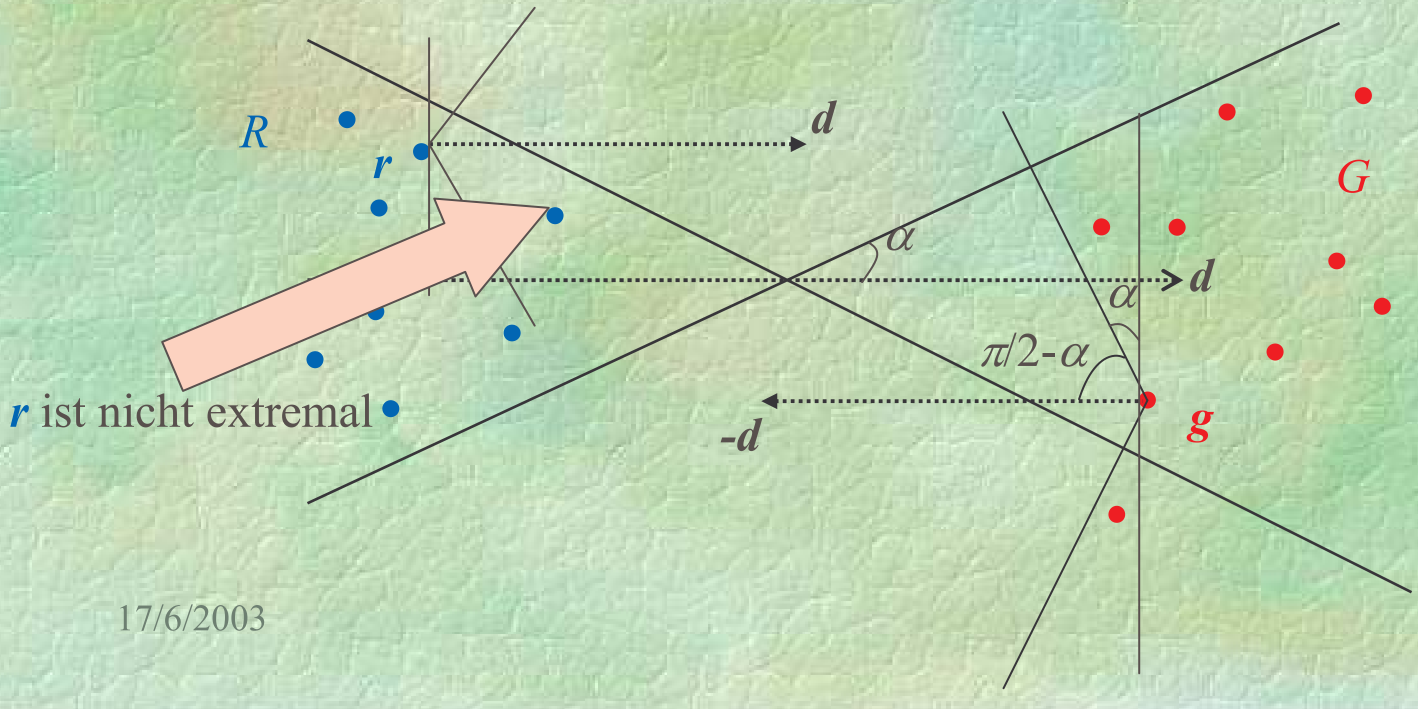
- *E eine Kantenmenge von S, die jeden EMST von S enthält*
- *$\mathcal{B}$ eine Menge von stark separierten Paaren (R, G):*
  $\forall$ *Kanten* $\{r, g\}$ $\exists\, (R, G) \in \mathcal{B}$: $r \in R$ *und* $g \in G$

Enthält $M \subseteq E$ *ein nächstes (R, G)- Paar für jedes* $(R, G) \in \mathcal{B}$, *dann enthält M einen EMST von S.*

Notation

Sei (R, G), R, $G \in S$, α-separiert in Richtung ***d*:**

- $\boldsymbol{r} \in R$ heißt *extremal*, wenn $\boldsymbol{r} + Cone(\boldsymbol{d}, \pi/2 - \alpha)$ kein Element aus R enthält.
- $\boldsymbol{g} \in$ G heißt *extremal*, wenn $\boldsymbol{g} + Cone(\boldsymbol{-d}, \pi/2 - \alpha)$ kein Element aus G enthält.
- Wir bezeichnen die Mengen der extremalen Elemente mit R' und G'.



EMST und extremale Elemente

Lemma 2.

- *(R, G), R, G $\subseteq$ S, ein α-separiertes Paar,*

*Ist {**r**, **g**}, **r** $\in$ R, **g** $\in$ G, eine Kante in einem EMST von S, so ist **r** $\in$ R′ und **g** $\in$ G′.*

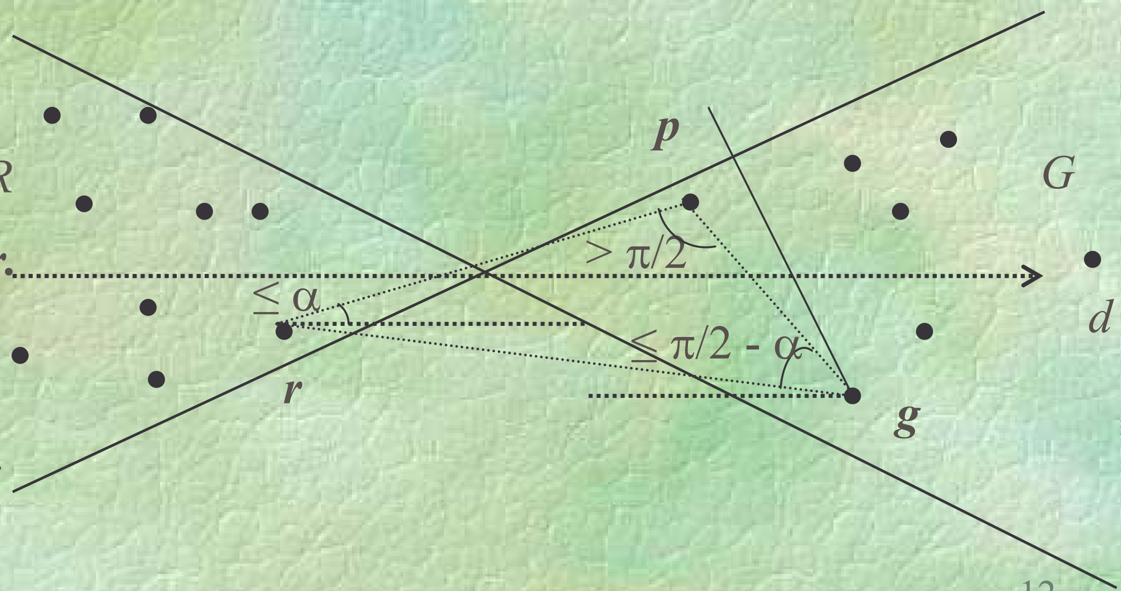
Beweis:

g ist nächster Nachbar von ***r***.

Annahme:

g ist nicht extremal:

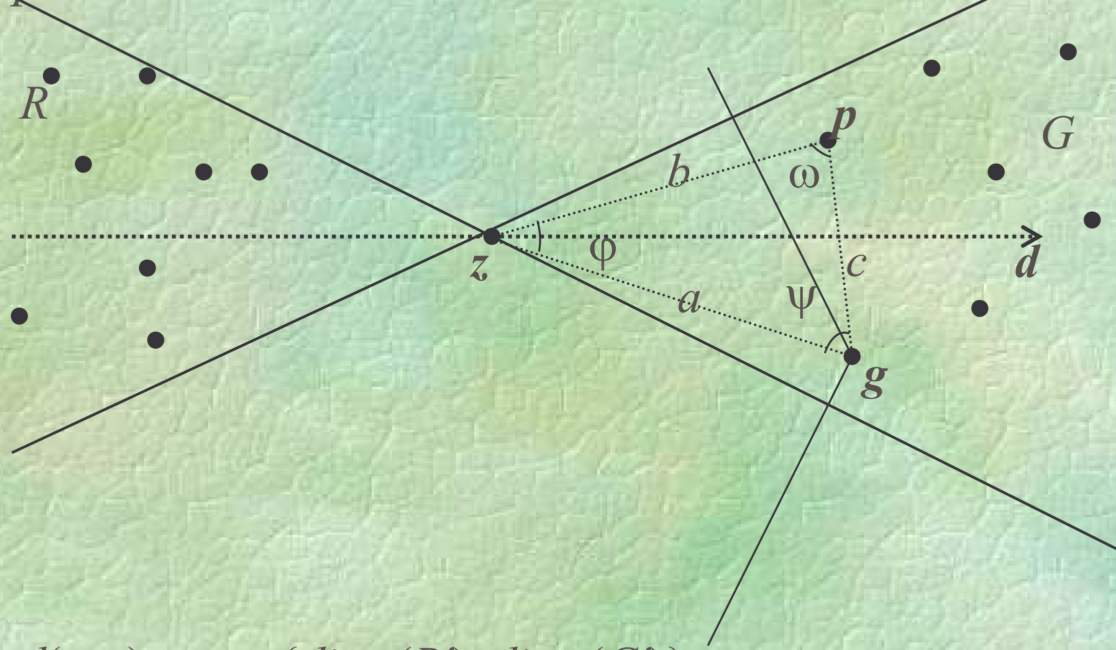
p $\in$ ***g*** + *Cone*(-***d***, $\pi/2 - \alpha$).



Extremale Elemente und separierte Paare

Lemma 3.

Ist (R′, G′) ein α-separiertes Paar von extremalen Elementen, dann ist (R′, G′) stark separiert.



Beweis:

$$\left.\begin{array}{l} \forall\, \mathbf{g}', \mathbf{g}'' \in G': \\ \quad d(\mathbf{g}', \mathbf{g}'') < d(\mathbf{z}, G'). \\ \forall\, \mathbf{r}', \mathbf{r}'' \in R': \\ \quad d(\mathbf{r}', \mathbf{r}'') < d(\mathbf{z}, R') \end{array}\right\} \Rightarrow d(\mathbf{r}, \mathbf{g}) > \max\{diam(R'), diam(G')\}.$$

α-separierte Paare und nächste Paare

Lemma 4.

- *B eine Menge von α-separierten Paaren mit:*

  für jedes Paar $\mathbf{r}, \mathbf{g} \in S$ *existiert ein Paar* $(R, G) \in B$ *mit* $r \in R$ *und* $g \in G$.

Existiert eine Kantenmenge M,
die ein nächstes (R, G) Paar für jedes $(R, G) \in B$ *enthält,*
dann enthält M einen EMST von S.

$\mathcal{B}$ Menge von α-separierten Paaren: $\forall \{r, g\}\ \exists (R,G) \in \mathcal{B}$.
M enthält ein nächstes (R, G)-Paar für jedes $(R, G) \in \mathcal{B}$.

$E = \{ \{\mathbf{r}', \mathbf{g}'\} : (R', G')\ \alpha\text{-separiert}\}$

$\mathcal{B}' = \{\alpha\text{-separierte } (R', G')\}$

L2

(R, G) α-separiert.
$\{\mathbf{r}, \mathbf{g}\}$ Kante in einem EMST $\Rightarrow \mathbf{r} \in R'$ und $\mathbf{g} \in G'$

L3

(R', G') α-separiert $\Rightarrow$
(R', G') stark separiert.

nächstes Paar von (R, G)
ist nächstes von (R', G')

E enthält jeden
EMST von S

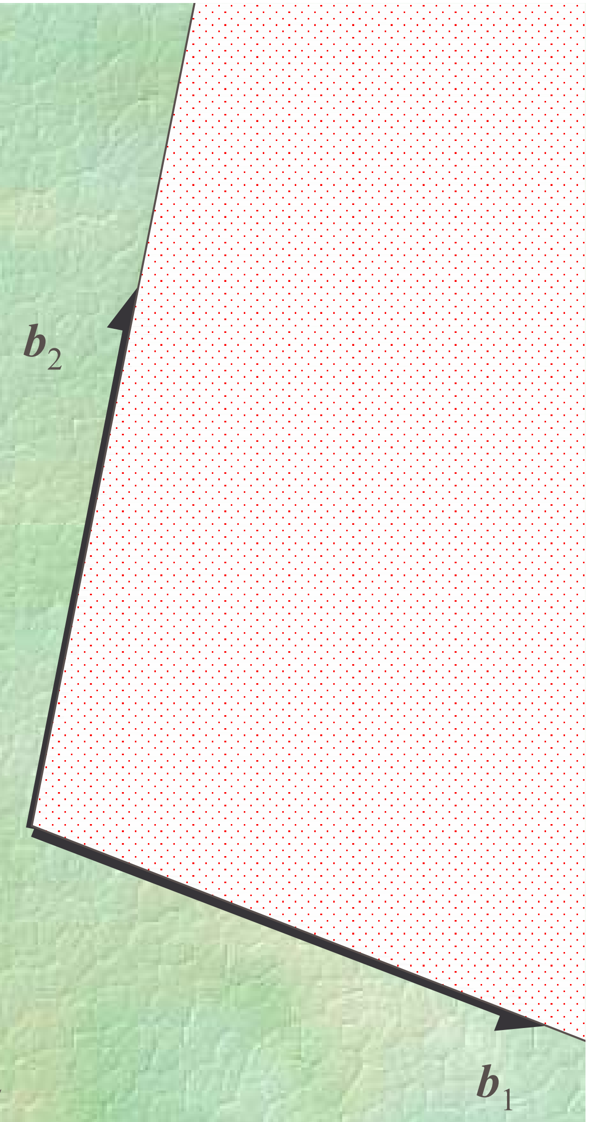
$\mathcal{B}' = \{(R', G') :$
$(R, G) \in \mathcal{B}$ stark separiert$\}$

L1

jeder EMST aus E, $\mathcal{B} = \{(R, G)$ stark separiert: $\forall \{r, g\} \in E\ \exists (R,G)\}$
Enthält $M \subseteq E$ ein nächstes (R, G)-Paar für jedes $(R, G) \Rightarrow M$ enthält einen EMST.

M enthält einen EMST von E

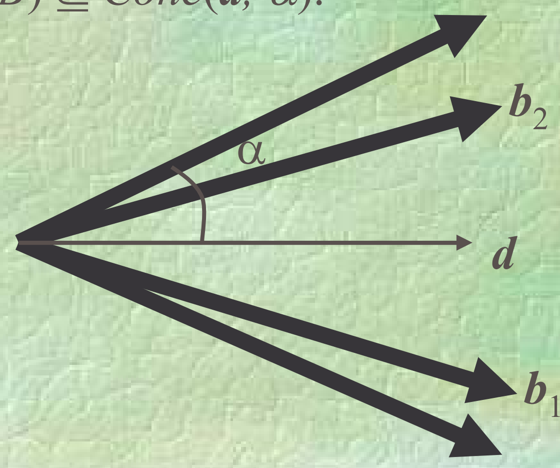
Notation



$B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, .., \boldsymbol{b}_d\}$ eine Basis von E^d ,

- Konvexer Kegel:
  $Conv(B) :=\{\sum_{i=1}^{d} \lambda_i b_i : \lambda_i \geq 0, \forall i\}$.

- $Conv(B)$ heißt eng: $Conv(B) \subseteq Cone(\boldsymbol{d}, \alpha)$.

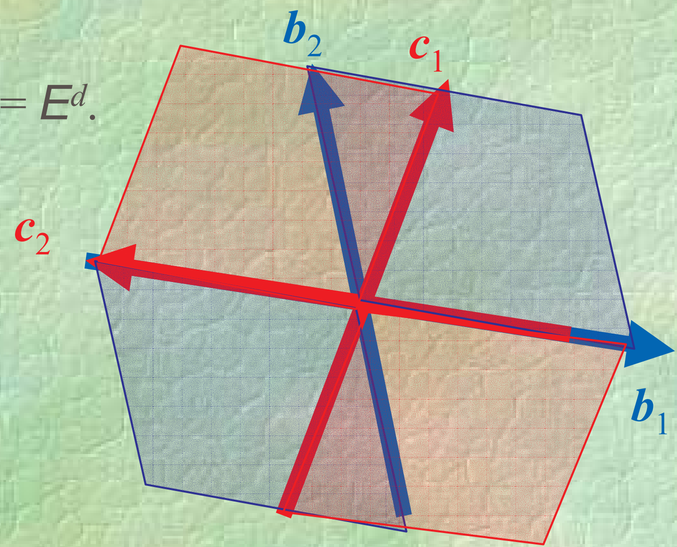


Notation

F eine endliche Familie von Basen.

- *F* heißt Fächer von E^d:

$$\bigcup_{B \in F} (Conv(B) \cup -Conv(B)) = E^d.$$



- *F* heißt eng, wenn jedes *Conv*(*B*) aus *F* eng ist.

Bemerkung: Wir können für jede Dimension einen engen Fächer in endlicher Anzahl von Schritten konstruieren.

Algorithmus für α-separierte Paare

$\forall B \in \mathsf{F}$ berechnen wir eine Menge B_B von α-separierten Paaren.
Keine 2 Punkte von S besitzen eine gemeinsame Koordinate bzgl. *B*.

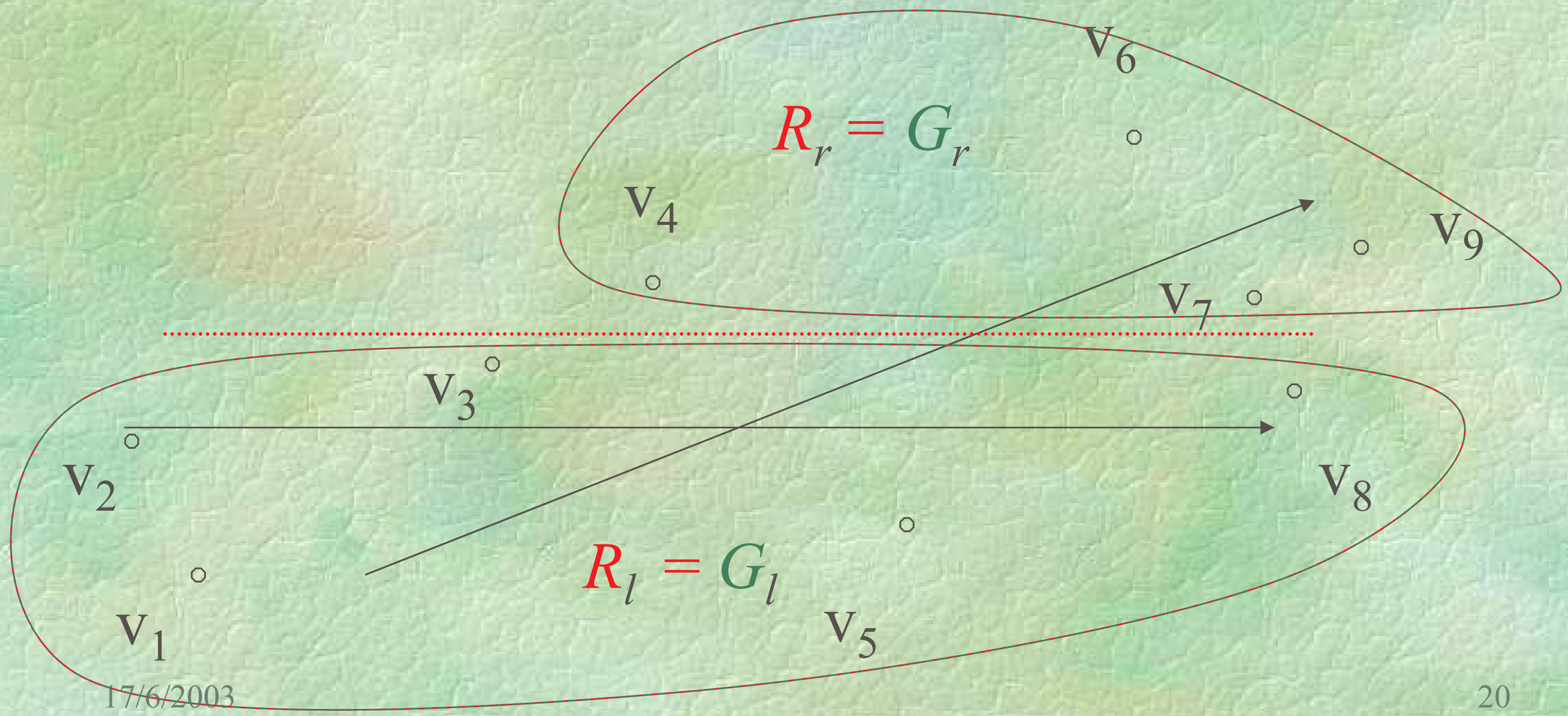
- Der Algorithmus ist rekursiv.
- Eingabe: Dimension $k <= d$,
  Menge von roten Punkten *R*,
  Menge von grünen Punkten *G* aus E^d.
- Am Anfang ist $k = d$ und $R = G = S$.
- Jeder rekursive Aufruf reduziert entweder die Anzahl der Punkte oder die Dimension.
- Wenn $k = 0$, werden die α-separierten Paare ausgegeben.

Algorithmus α-sep(k, R, G)

- if $k = 0$, then output (R, G) as an α-seperated pair.
- Otherwise :
  - x_k := median of the k-th coordinate of $R \cup G$.
  - $R_l := \{ \boldsymbol{r} \in R / r_k \leq x_k \}$,
    $R_r := \{ \boldsymbol{r} \in R / r_k > x_k \}$,
    $G_l := \{ \boldsymbol{g} \in G / g_k \leq x_k \}$,
    $G_r := \{ \boldsymbol{g} \in G / g_k > x_k \}$.
  - If $R_l \neq \emptyset$ and $G_r \neq \emptyset$, then α-sep(k-1, R_l, G_r).
  - If $R_l \neq \emptyset$ and $G_l \neq \emptyset$, then α-sep(k, R_l, G_l).
  - If $R_r \neq \emptyset$ and $G_r \neq \emptyset$, then α-sep(k, R_r, G_r).

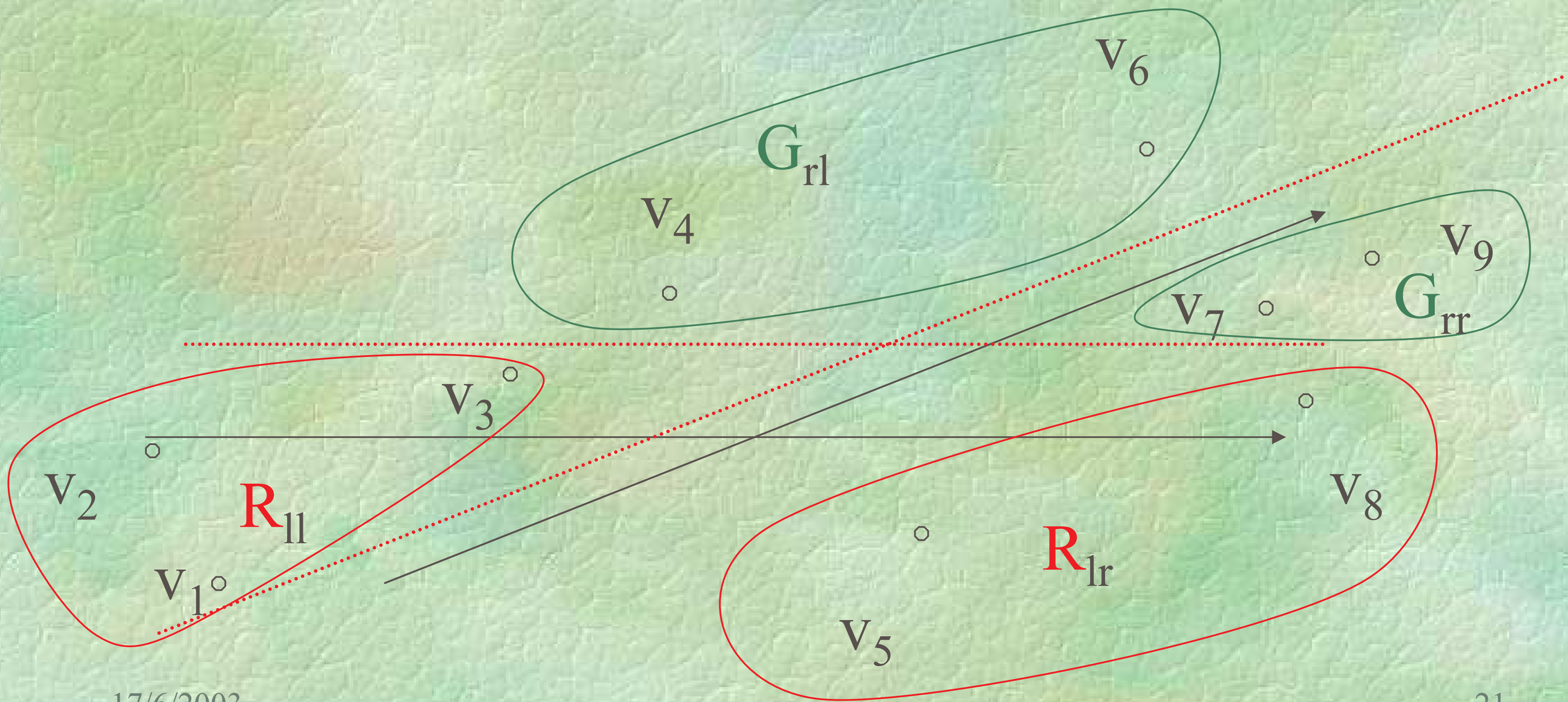
Beispiel für α-sep(k, R, G)

1. R = G = {v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9}; k=2



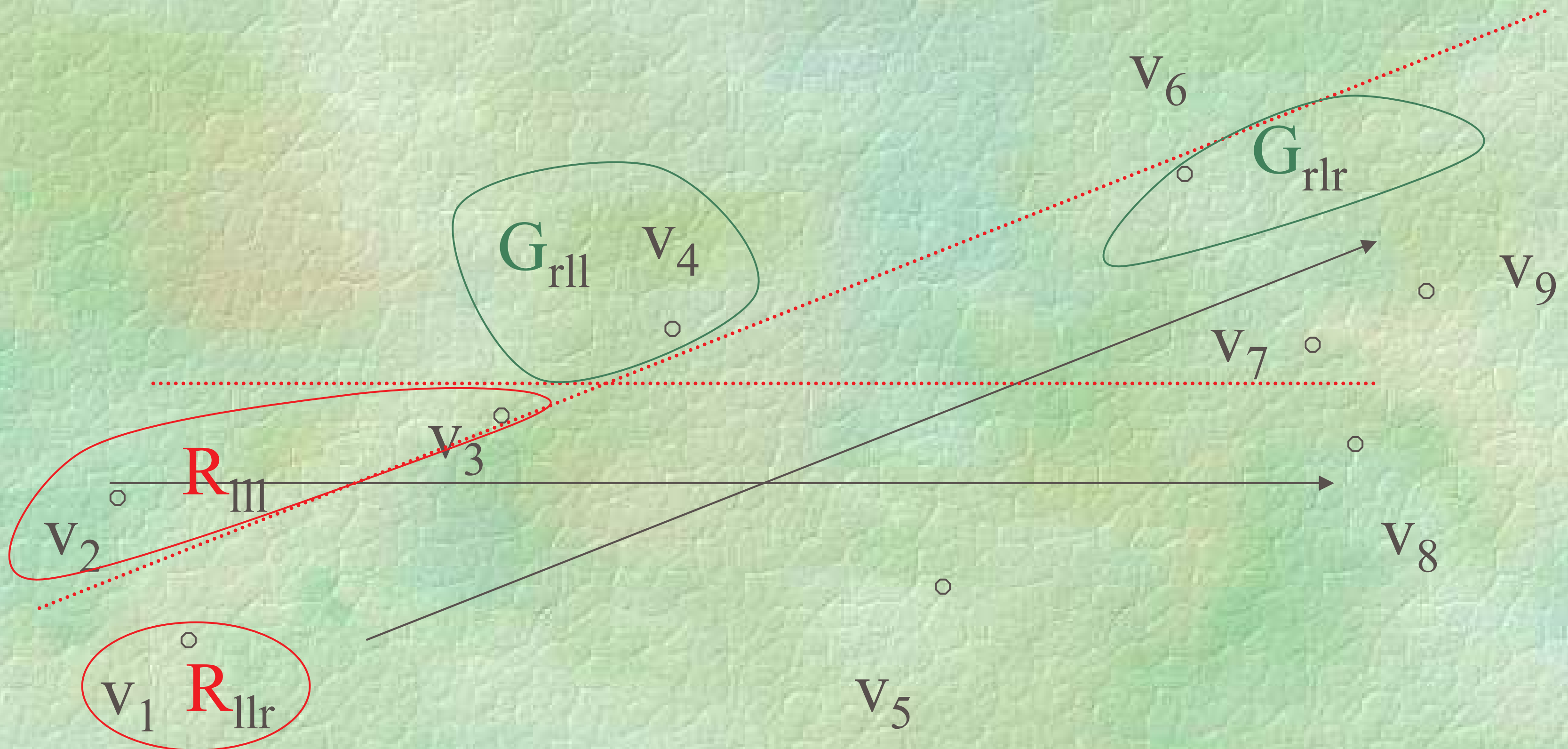
Beispiel für α-sep(k, R, G)

$R_l=\{v_1, v_2, v_3, v_5, v_8\}$; $G_r=\{v_4, v_6, v_7, v_9\}$; k=1



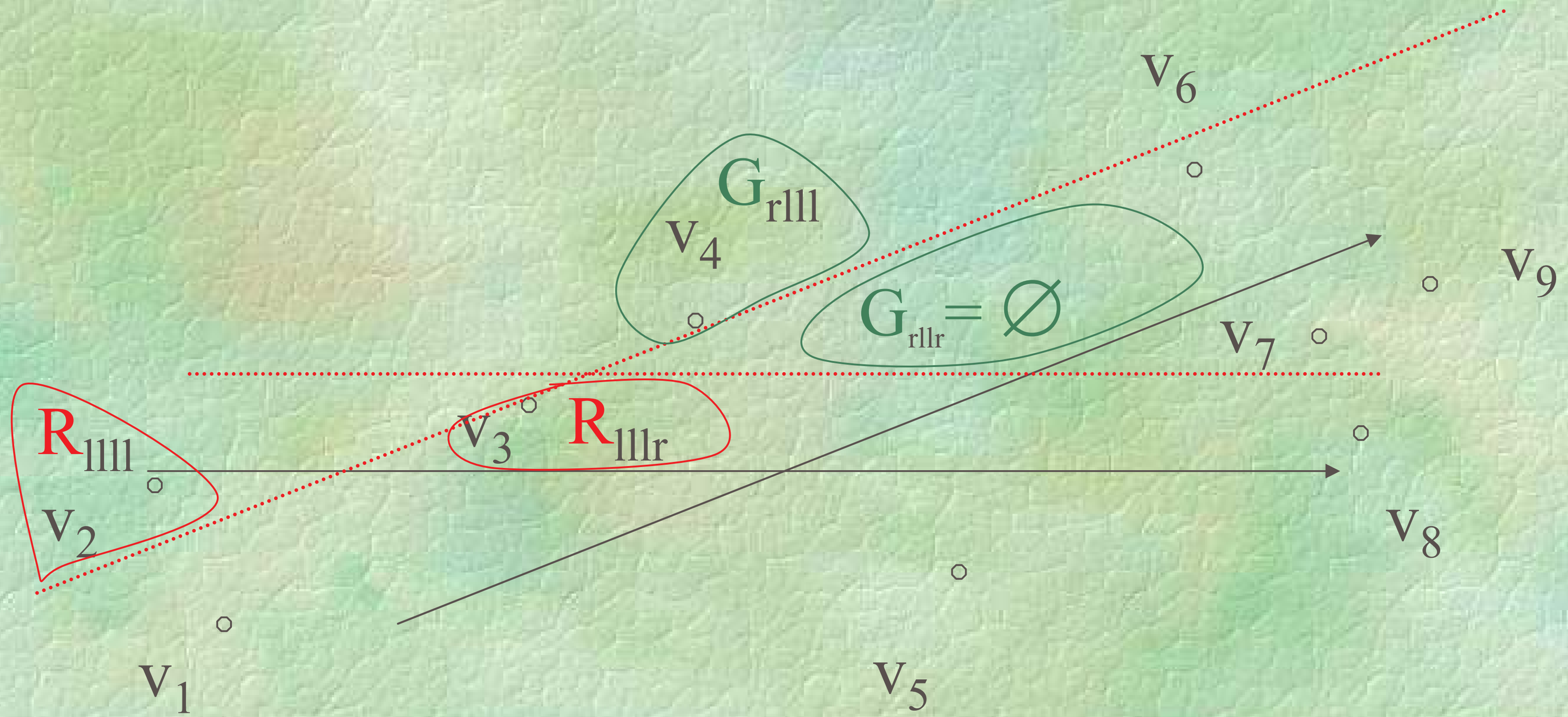
Beispiel für α-sep(k, R, G)

$R_{ll}=\{v_1, v_2, v_3\}$; $G_{rl}=\{v_4, v_6\}$; k=1



Beispiel für α-sep(k, R, G)

$R_{lll}=\{v_2, v_3\}$; $G_{rll}=\{v_4\}$; $k=1$



$R = G = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9\}$; $k=2$

$R_l=\{v_1, v_2, v_3, v_5, v_8\}$;$G_r=\{v_4, v_6, v_7, v_9\}$; $k=1$

$R_{ll}=\{v_1, v_2, v_3\}$;$G_{rr}=\{v_7, v_9\}$; $k=0$

$R_{ll}=\{v_1, v_2, v_3\}$;$G_{rl}=\{v_4, v_6\}$; $k=1$

$R_{lll}=\{v_2,v_3\}$;$G_{rlr}=\{v_6\}$; $k=0$

$R_{lll}=\{v_2, v_3\}$;$G_{rll}=\{v_4\}$; $k=1$

$R_{llll}=\{v_2\}$;$G_{rllr}=\{\}$; $k= 0$

$R_{llll}=\{v_2\}$;$G_{rlll}=\{v_4\}$; $k=1$

$R_{lllll}=\{v_2\}$;$G_{rlllr}=\{v_4\}$; $k=0$

$R_{lllr}=\{v_3\}$;$G_{rllr}=\{v_4\}$; $k=1$

$R_{lllrl}=\{v_3\}$;$G_{rllrr}=\{v_4\}$; $k=0$

$R_{llr}=\{v_1\}$;$G_{rlr}=\{v_6\}$; $k=1$

$R_{llrl}=\{v_1\}$;$G_{rlrr}=\{v_6\}$; $k=0$

$R_{lr}=\{v_5, v_8\}$;$G_{rr}=\{v_7, v_9\}$; $k=1$

$R_l=\{v_1, v_2, v_3, v_5, v_8\}$;$G_l=\{v_1, v_2, v_3, v_5, v_8\}$; $k=1$

k

Anzahl α-separierter Paare in E^2

$t_d(N) \le 2t_d(N/2) + t_{d-1}(N)$

$N / 2^s = 1$

$$
\begin{array}{cccccccc}
t_2(N) & \rightarrow & 2^1 t_2(N/2) & \rightarrow & 2^2 t_2(N/2^2) & \rightarrow 2^3 t_2(N/2^3) & \rightarrow 2^4 t_2(N/2^4) \;\ldots\ldots\ldots\ldots & 2^s t_2(N/2^s) \\
\downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow \\
t_1(N) & \rightarrow & 2^2 t_1(N/2) & \rightarrow & (2^2+2^3)t_1(N/2^2) & \rightarrow (2*2^3+2^4)t_1(N/2^3) \;\ldots\ldots & ((s-2)*2^{s-1}+2^s)t_1(N/2^{s-1}) & \rightarrow ((s-1)*2^s+2^{s+1})t_1(N/2^s) \\
\downarrow & & \downarrow & & \downarrow & \downarrow & \downarrow & \downarrow \\
t_0(N) & & 2^2 t_0(N/2) & & (2^2+2^3)t_0(N/2^2) & \ldots\ldots \; ((s-3)*2^{s-2}+2^{s-1})t_1(N/2^{s-2}) & ((s-2)*2^{s-1}+2^s)t_0(N/2^{s-1}) & ((s-1)*2^s+2^{s+1})t_0(N/2^s)
\end{array}
$$

$1 + 2^2 + (2^2+2^3) + (2*2^3+2^4) + ..+ ((s-1)*2^s + 2^{s+1})$

$1*2^2+ 2*2^3+ 3*2^4+ ..+ (s-1)*2^s \quad = \sum_{i=1}^{s-1}(i*2^{i+1}) = 2^2(\sum_{i=1}^{s-1}(i*2^{i-1}) = 2^2(\sum_{i=0}^{s-1}(i*2^{i-1})$

$= 2^2((s-1)*2^s+1)= 2^2(\log(N/2)*N +1) \in O\,(N\log N)$

N

Korrektheit und Vollständigkeit von α-sep(k, R, G)

Korrektheit:

- B ist eng $\Rightarrow$ jedes ausgegebene Paar ist α-separiert.

Vollständigkeit:

- ist $\boldsymbol{g} \in \boldsymbol{r} + Conv(B)$, dann gibt α-sep(k, R, G) (R, G) mit $\boldsymbol{r} \in R$ und $\boldsymbol{g} \in G$ aus.

Induktion:

$k = 0$: richtig

Annahme: die Behauptung ist richtig für $k = d$-1.

$\boldsymbol{g} \in \boldsymbol{r} + Conv(B)$

$\Rightarrow r_i < g_i \quad 1 \leq k \leq d$

$\Rightarrow r_d < g_d$

$\Rightarrow$ Aufruf α-sep (d, $\mathbf{r} \in R_l$, $\mathbf{g} \in G_r$)

$\Rightarrow$ Aufruf α-sep (d-1, $\mathbf{r} \in R_l$, $\mathbf{g} \in G_r$).

Anzahl α-separierter Paare

$t_k(n + m)$: max. Anzahl von ausgegebenen α-sep. Paaren bei $|R| = n$, $|G| = m$ und k.

- $t_0(n + m) = 1$
- $t_k(n+m) \leq 2t_k((n+m)/2\ )+ t_{k-1}(n+m)$.

d. h. $t_k(n+m) = O((n+m)\log^{k-1}(n+m)) \Rightarrow t_d\,(N) \in O(N\log^{d-1}N)$.

- Laufzeit von Prim´s Algorithmus für MST: $O(|E| + N\log N)$
- Von M können wir EMST in $O(N\log^{d-1}N)$ Zeit berechnen.

Berechnung nächster Paare

- Die von α-sep(k, R, G) ausgegebene α-separierten Paare erfüllen den Bed. von L4

- Für jedes ausgegebene (R, G) mit $|R| = n$ und $|G| = m$ berechnen wir ein nächstes Paar in $T_d^0(n + m)$ Zeit.

- $T_d^k(n + m)$: Laufzeit bei Dimension k und $|R| = n$ und $|G| = m$.

- $T_d^0(n + m) \in \Omega(n + m)$:

$$T_d^k(n + m) = 2T_d^k((n + m)/2) + T_d^{k-1}(n + m) \text{ für } k \geq 1.$$

d.h. $T_d^d(N) \in O(T_d^0(n + m) \log^d N)$.

Berechnung nächster Paare

Satz 5.

- $T_d^0(n + m)$ *die Laufzeit um ein BCP für n roten und m grünen Punkten in* E^d *zu berechnen.*

Ein EMST von N Punkten in E^d *kann in:*

$$O(\, T_d^0(n + m) \log^d N)$$

Zeit berechnet werden.

Berechnung nächster Paare

- Lösung des BCP:
  für jeden Punkt $\boldsymbol{g} \in G$ suche seinen nächsten Nachbarn $\boldsymbol{r} \in R$.

- Triviale Lösung in $O(mn)$.

- Schnellere Lösung:
  - Wähle für R eine Datenstruktur, welche folgendes unterstützt:
    Gegeben sei ein Punkt $\boldsymbol{y}$ in E^d. Welcher Punkt in R ist dann der Nächste zu $\boldsymbol{y}$?

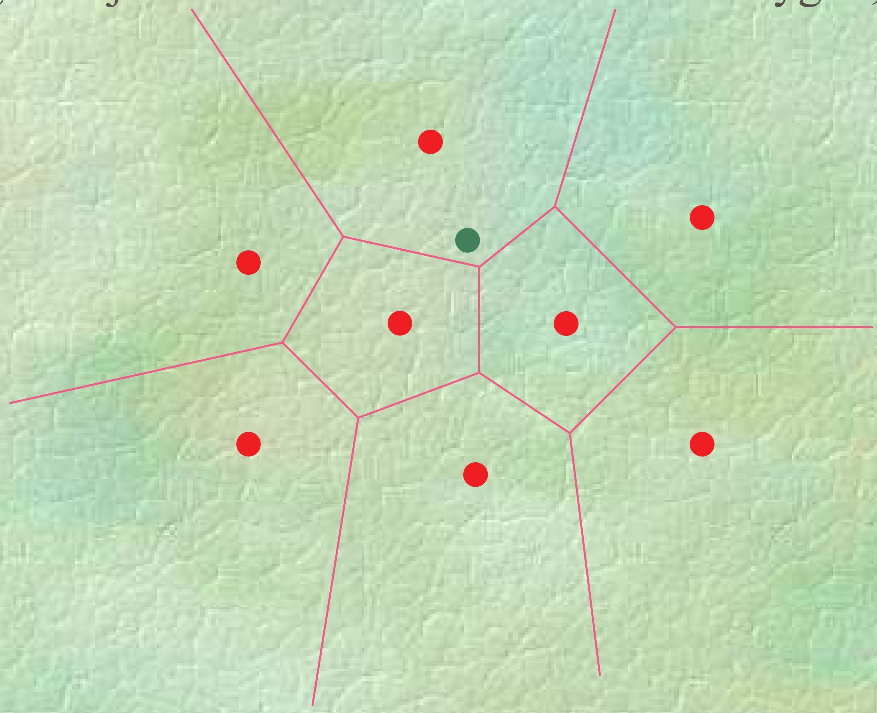
- Die Vorverarbeitung einer solchen Datenstruktur ist in der Literatur als das ***Postamt-Problem*** bekannt:

  Gegeben sind N Punkte P_i (Postämter), bestimme für jedes Postamt das Gebiet, in dem es „zuständig" (das nächste) ist.

Berechnung nächster Paare

In der 2-dim. Ebene mit Hilfe des Voronoi-Diagramms:

- Grenzen zwischen den Einzugsgebieten benachbarter Postämter sind Teile der Mittelsenkrechten.
- Das Einzugsgebiet jedes Postamts ist ein Voronoi-Polygon, eventuell nach einer Seite offen.



Berechnung nächster Paare

Annahme für das *post office problem*:

- Vorverarbeitung für n Punkte aus E^d in $n^{\lceil d/2 \rceil} p(n) \in \Omega(n^{\lceil d/2 \rceil})$ Zeit,
- Antwort auf nächstes-Paar-Abfragen in polylogarithmischer Zeit $q(n)$.

- $\Omega(n^{\lceil d/2 \rceil})$: kombinatorische Komplexität des Voronoi-Diagramms im *worst case*.

Berechnung nächster Paare

- BCP Lösung:

$$O(n^{\lceil d/2 \rceil} p(n) + mq(n)).$$

- Gut für m $\geq n^{\lceil d/2 \rceil} p(n)/\, q(n)$ (G >> R).

- $n^{\lceil d/2 \rceil} > mq(n)/\, p(n)$:
  - teile R in $t = \lceil n((p(n)/mq(n)))^{\lceil d/2 \rceil} \rceil > 1$ Untermengen $R_1, R_2, .., R_t$ jeweils der Größe höchstens $\lceil n/\mathrm{t} \rceil$.
  - für jedes (R_i, G) obige Methode anwenden

- Laufzeit:

$$O(t((n/t)^{\lceil d/2 \rceil} p(n/t) + mq(n/t)))$$
$$= O(nm^{1 - 1/\lceil d/2 \rceil} p(n)^{1/\lceil d/2 \rceil} q(n)^{1 - 1/\lceil d/2 \rceil})$$

Berechnung nächster Paare

Lemma 6.

- *post office problem hat:*
  - $n^{\lceil d/2 \rceil} p(n)$ *Vorverarbeitungszeit,*
  - $q(n)$ *Suchezeit.*

Das BCP Problem für eine Menge von n roten und m grünen Punkten in E^d *kann in*

$$O(nm^{1-1/\lceil d/2\rceil} p(n)^{1/\lceil d/2\rceil} q(n)^{1-1/\lceil d/2\rceil} + mq(n))$$

gelöst werden.

BCP vs. EMST

Unser Algorithmus:

- lösen das BCP-Problem → Berechnen einen EMST
- Wenn BCP $\in T_d^0(n+m) \rightarrow$ EMST $\in O(T_d^0(2N) \log^d N)$

Wichtig

- Ein EMST von der Vereinigung der grünen und roten Punkten enthält mindestens ein nächstes rot-grün Paar.
- Berechnen einen EMST → lösen das BCP-Problem

Euclidean minimum spanning trees and bichromatic closest pairs

EMST:

- Stark separierte Paare und EMST (Lemma 1)
- EMST und extremale Elemente (Lemma 2)
- Extremale Elemente und separierte Paare (Lemma 3)
- α-separierte Paare und nächste Paare (Lemma 4)
- Algorithmus um α-sep. Paare zu berechnen.
- Nächstes Paar (BCP) zwischen jedem α-sep. Paar finden.
- EMST aus den Nächsten Paaren bestimmen.

BCP:

- Postamt-Problem und Voronoi-Diagramm
- Algorithmus um BCP schnell zu bestimmen

EMST vs. BCP