Constrained Minkowski Sums

A framework for solving subsequence problems efficiently

Friedrich Eisenbrand

Joint work with Thorsten Bernholt and Thomas Hofmeister

Subsequence Problems

Subsequence problem:

Given n numbers $a_1, \ldots, a_n \in \mathbb{R}$, find an interval $[i, j] \subseteq \{1, \ldots, n\}$ satisfying certain constraints such that a certain weight function is maximized.

Examples:

- Length-constrained heaviest segment: Given a length bound L, find interval $[i, j]$ with length at least L such that $\sum_{k=i}^{j} a_k/(j - i + 1)$ is maximal.

  (Lin et al., JCSS 2002)

- DNA copy number data analysis: Find an interval $[i, j]$ such that $|\sum_{k=i}^{j} a_k|/\sqrt{j - i + 1}$ is as large as possible.

  (Lipson et al., RECOMB 2005)

Geometric Interpretation

- Interval $I = [i, j]$ has length $\ell(I) = j - i + 1$ and weight $s(I) = \sum_{k=i}^{j} a_k$
- Interpret Interval I as point $p_I = (\ell(I), s(I))$ in the plane

Task:

- Length-constrained heaviest segment: Find point, maximizing function $f(\ell, s) = s/\ell$
- DNA copy number data analysis: Find point, maximizing $f(\ell, s) = s^2/\ell$

In both cases, f is quasi-convex

Geometric Interpretation

- Interval $I = [i, j]$ has length $\ell(I) = j - i + 1$ and weight $s(I) = \sum_{k=i}^{j} a_k$
- Interpret Interval I as point $p_I = (\ell(I), s(I))$ in the plane

Task:

- Length-constrained heaviest segment: Find point, maximizing function $f(\ell, s) = s/\ell$
- DNA copy number data analysis: Find point, maximizing $f(\ell, s) = s^2/\ell$

In both cases, f is quasi-convex

Recall: $\phi : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex if for all x, x' in domain of ϕ:

$$\text{for all } 0 \leqslant \lambda \leqslant 1 : \quad \phi(\lambda x + (1 - \lambda)x') \leqslant \max\{\phi(x), \phi(x')\}$$

Maximum is attained at extreme point of domain of ϕ

The link to Minkowski sums

Recall: Minkowski sum of P and Q is

$$P \oplus Q = \{p + q \mid p \in P,\, q \in Q\}$$

The link to Minkowski sums

Recall: Minkowski sum of P and Q is

$$P \oplus Q = \{p + q \mid p \in P,\ q \in Q\}$$

- Given numbers $a_1, \ldots, a_n$ consider points $P = \{p_i = (i, a_1 + \cdots + a_i) \mid i = 1, \ldots, n\}$ and $Q = \{q_i = (-i, -a_1 - \cdots - a_i) \mid i = 1, \ldots, n\}$
- If $j > i$, then $q_i + p_j = (j - i, \sum_{k=i+1}^{j} a_k)$
- This is p_I for interval $I = [i + 1, j]$
- Not all points $q_i + p_j$, $p \in P$, $q \in Q$ are points of intervals.
- Intervals are points in Minkowski sum of P and Q which satisfy constraint $x_1 \geqslant 1$.

The geometric approach to subsequence problems

- A quasi-convex function attains its maximum at extreme point
- Compute superset of vertices of

$$\operatorname{conv}\left(\left(P \oplus Q\right) \cap \left(x_1 \geqslant 1\right)\right)$$

- Return optimal vertex

Constrained Minkowski Sums

- $Ax \geqslant b$ a set of k constraints
- $P, Q \subseteq \mathbb{R}^2$ finite point-sets
- Constrained Minkowski sum of P and Q and $Ax \geqslant b$:

$$(P \oplus Q)_{Ax \geqslant b} = \{p + q \mid p \in P,\, q \in Q,\, A(p+q) \geqslant b\}$$

Main result of this talk:

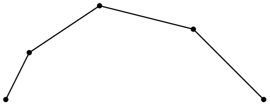
- Given $P \subseteq \mathbb{R}^2$ and $Q \subseteq \mathbb{R}^2$ with $|P| + |Q| = N$, one can compute a set $R \subseteq (P \oplus Q)_{Ax \geqslant b}$ containing all vertices of $(P \oplus Q)_{Ax \geqslant b}$ in time $O(N \log N)$ if k is fixed.
- A quasi-convex function can be maximized over $(P \oplus Q)_{Ax \geqslant b}$ in time $O(N \log N)$ if k is fixed.

A lower bound

- Set disjointness: Given sets of real numbers $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$, determine whether $A \cap B = \emptyset$.
- Lower bound of $\Omega(n \log n)$ in algebraic decision-tree model (Ben-Or 1983)
- Let $P = \{(0, -a) \mid a \in A\}$ and $Q = \{(0, b) \mid b \in B\}$
- A and B are not disjoint if and only if $(0, 0) \in P \oplus Q$
- if and only if $\min\{x_2 \mid x \in (P \oplus Q)_{x_2 \geqslant 0}\} = 0$

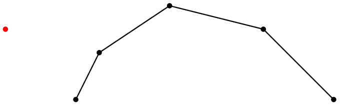
Maximizing a linear function over $(P \oplus Q)_{a^T x \geqslant b}$ requires $\Omega(N \log N)$ operations.

Computing convex hulls



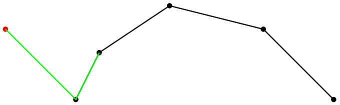
- Sort points of $P \subseteq \mathbb{R}^2$ w.r.t. some linear function
- Sweep from "right to left", thereby updating the upper convex hull
- Running time to compute upper hull is $O(N)$ after sorting and $O(N \log N)$ together with sorting

Computing convex hulls



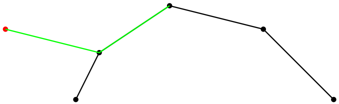
- Sort points of $P \subseteq \mathbb{R}^2$ w.r.t. some linear function
- Sweep from “right to left”, thereby updating the upper convex hull
- Running time to compute upper hull is $O(N)$ after sorting and $O(N \log N)$ together with sorting

Computing convex hulls



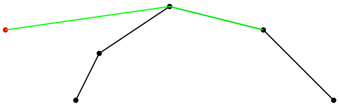
- Sort points of $P \subseteq \mathbb{R}^2$ w.r.t. some linear function
- Sweep from "right to left", thereby updating the upper convex hull
- Running time to compute upper hull is $O(N)$ after sorting and $O(N \log N)$ together with sorting

Computing convex hulls



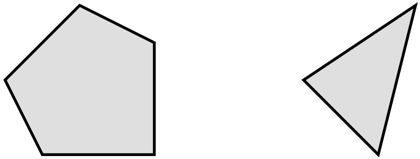
- Sort points of $P \subseteq \mathbb{R}^2$ w.r.t. some linear function
- Sweep from "right to left", thereby updating the upper convex hull
- Running time to compute upper hull is $O(N)$ after sorting and $O(N \log N)$ together with sorting

Computing convex hulls

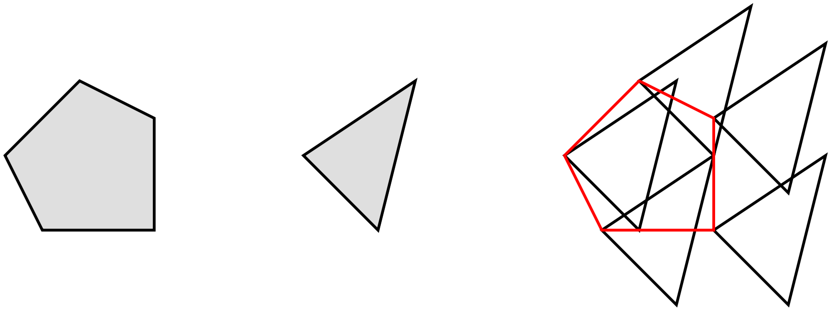


- Sort points of $P \subseteq \mathbb{R}^2$ w.r.t. some linear function
- Sweep from "right to left", thereby updating the upper convex hull
- Running time to compute upper hull is $O(N)$ after sorting and $O(N \log N)$ together with sorting

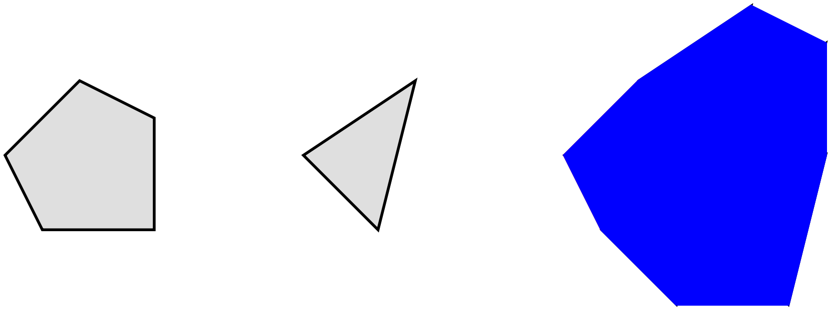
Computing convex hulls of Minkowski sums



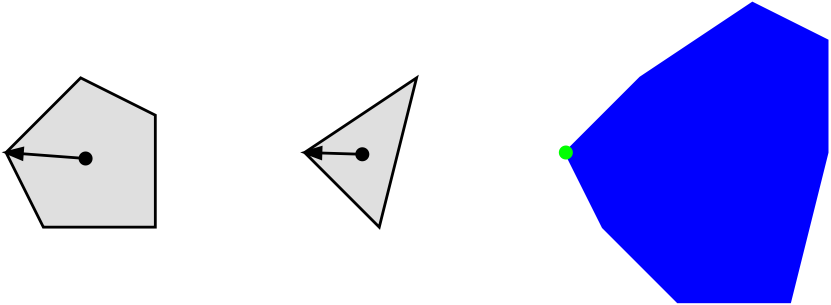
Computing convex hulls of Minkowski sums



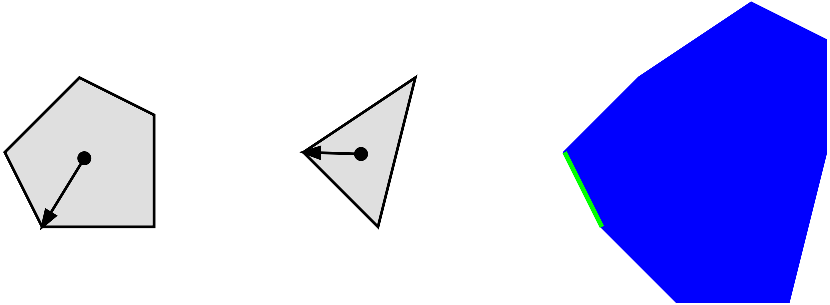
Computing convex hulls of Minkowski sums



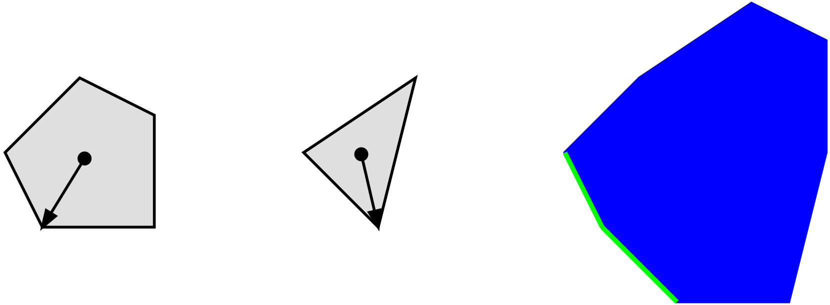
Computing convex hulls of Minkowski sums



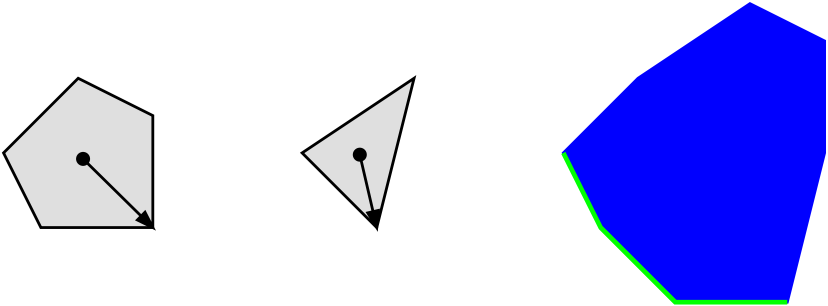
Computing convex hulls of Minkowski sums



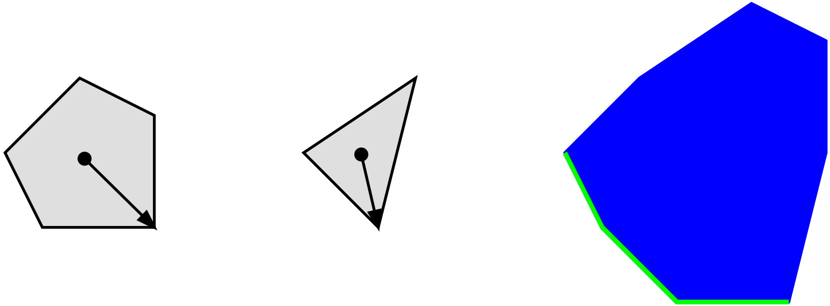
Computing convex hulls of Minkowski sums



Computing convex hulls of Minkowski sums



Computing convex hulls of Minkowski sums

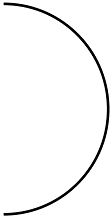


The appearance of $p \in P$ (or $q \in Q$) in a vertex $p + q$ of $\mathrm{conv}(P \oplus Q)$ is consecutive

Computing convex hulls of Minkowski Sums

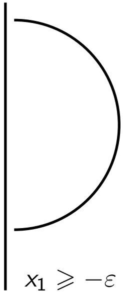
- $\text{conv}(P \oplus Q)$ has at most $|P| + |Q|$ vertices
- If P and Q are sorted w.r.t. some linear function, then $\text{conv}(P \oplus Q)$ can be computed in linear time

One constraint: Number of vertices, lower bound



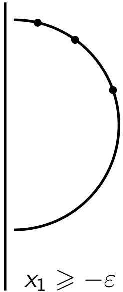
- Start with half circle with positive x_1-values

One constraint: Number of vertices, lower bound



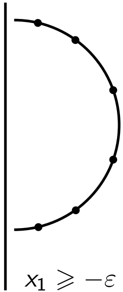
- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$

One constraint: Number of vertices, lower bound



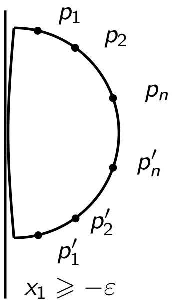
- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$
- Distribute n points on upper half of half-circle

One constraint: Number of vertices, lower bound



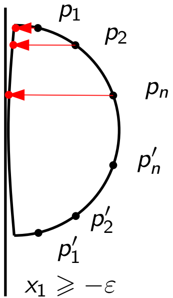
- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$
- Distribute n points on upper half of half-circle
- Mirror points on the lower part

One constraint: Number of vertices, lower bound



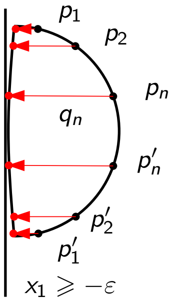
- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$
- Distribute n points on upper half of half-circle
- Mirror points on the lower part
- Bend line-segment which is closing circle slightly to the left

One constraint: Number of vertices, lower bound



- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$
- Distribute n points on upper half of half-circle
- Mirror points on the lower part
- Bend line-segment which is closing circle slightly to the left
- q_i is vector orthogonal to x_2-axis such that $q_i + p_i$ and $q_i + p_i'$ are on curve.

One constraint: Number of vertices, lower bound



- Start with half circle with positive x_1-values
- Constraint is $x_1 \geqslant -\varepsilon$ for small $\varepsilon > 0$
- Distribute n points on upper half of half-circle
- Mirror points on the lower part
- Bend line-segment which is closing circle slightly to the left
- q_i is vector orthogonal to x_2-axis such that $q_i + p_i$ and $q_i + p'_i$ are on curve.
- $(\{p_1, \ldots, p_n, p'_1, \ldots, p'_n\} \oplus \{0, q_1, \ldots, q_n\})_{x_1 \geqslant -\varepsilon}$ has $4 \cdot n$ vertices $= |P| + 2 \cdot |Q| - 2$

One constraint: Number of vertices, tight upper bound

- Given: $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_m\}$ sorted w.r.t. first coordinate
- Constraint: $x_1 \geqslant 0$
- $J(i)$: smallest index j such that $p_i + q_j$ satisfies $x_1 \geqslant 0$.

Set $(P \oplus Q)_{x_1 \geqslant 0}$ is union of

$$\begin{array}{rcl} \{p_{\mathbf{1}}, \ldots, p_n\} & \oplus & \{q_{J(\mathbf{1})}, \ldots, q_m\} \\ \{p_{\mathbf{2}}, \ldots, p_n\} & \oplus & \{q_{J(\mathbf{2})}, \ldots, q_m\} \\ & \ldots & \\ \{p_{\mathbf{n}}\} & \oplus & \{q_{J(\mathbf{n})}, \ldots, q_m\} \end{array}$$

One constraint: Number of vertices, tight upper bound

- Given: $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_m\}$ sorted w.r.t. first coordinate
- Constraint: $x_1 \geqslant 0$
- $J(i)$: smallest index j such that $p_i + q_j$ satisfies $x_1 \geqslant 0$.

Set $(P \oplus Q)_{x_1 \geqslant 0}$ is union of

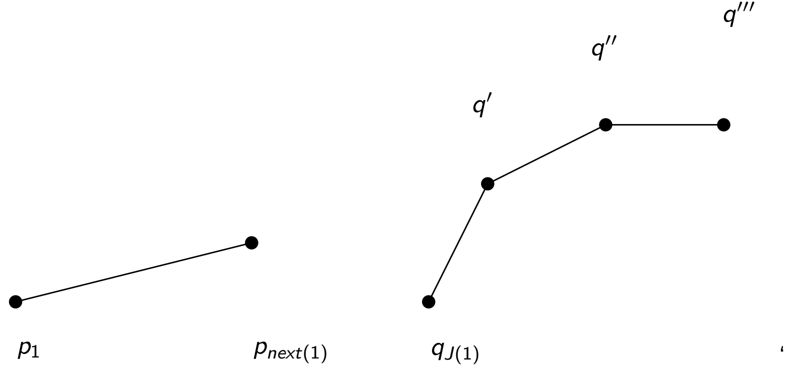
$$
\begin{array}{rcl}
\{p_{\mathbf{1}}, \ldots, p_n\} & \oplus & \{q_{J(\mathbf{1})}, \ldots, q_m\} \\
\{p_{\mathbf{2}}, \ldots, p_n\} & \oplus & \{q_{J(\mathbf{2})}, \ldots, q_m\} \\
\ldots & & \\
\{p_{\mathbf{n}}\} & \oplus & \{q_{J(\mathbf{n})}, \ldots, q_m\}
\end{array}
$$

$J(i)$ is decreasing sequence

One constraint: Number of vertices, tight upper bound

Consider convex hull of

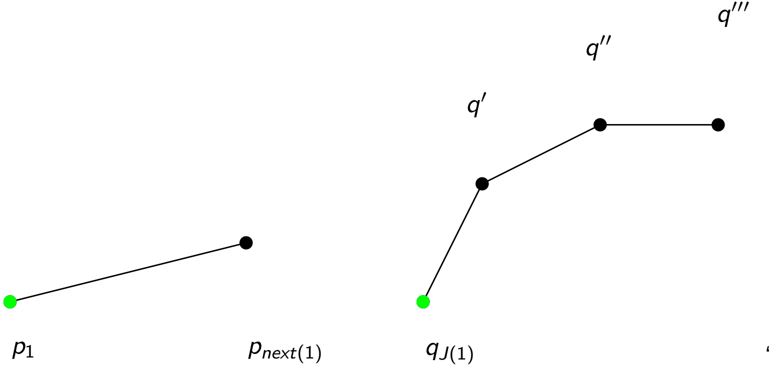
$$\{p_1, \ldots, p_n\} \oplus \{q_{J(1)}, \ldots, q_m\}$$



One constraint: Number of vertices, tight upper bound

Consider convex hull of

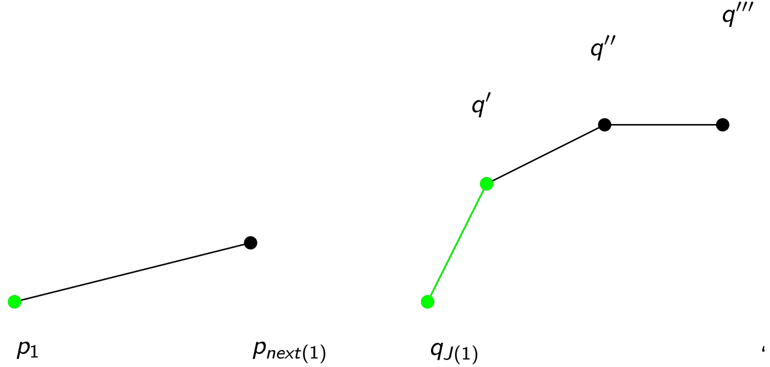
$$\{p_1, \ldots, p_n\} \oplus \{q_{J(1)}, \ldots, q_m\}$$



One constraint: Number of vertices, tight upper bound

Consider convex hull of

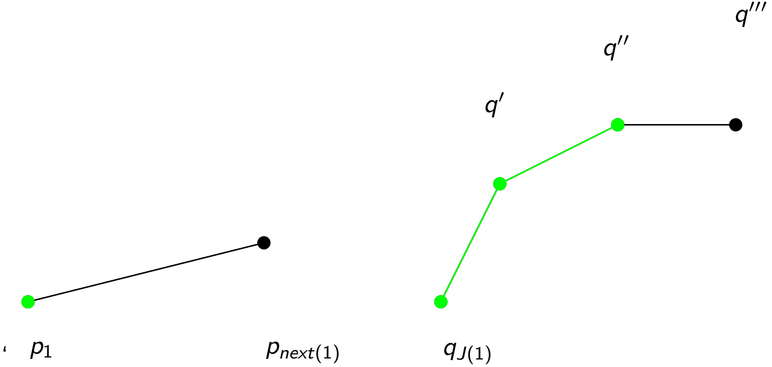
$$\{p_1, \ldots, p_n\} \oplus \{q_{J(1)}, \ldots, q_m\}$$



One constraint: Number of vertices, tight upper bound

Consider convex hull of

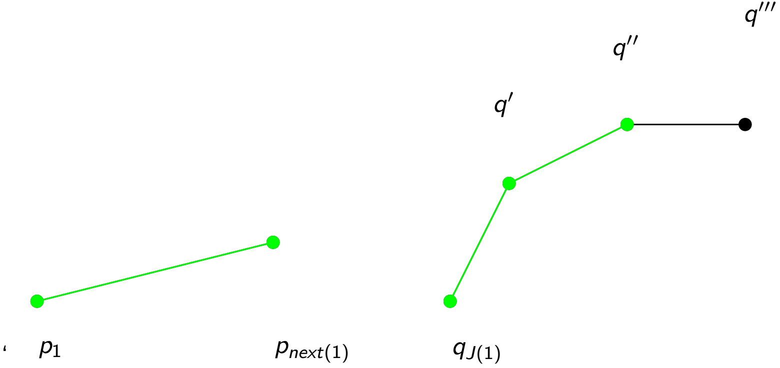
$$\{p_1, \ldots, p_n\} \oplus \{q_{J(1)}, \ldots, q_m\}$$



One constraint: Number of vertices, tight upper bound

Consider convex hull of

$$\{p_1, \ldots, p_n\} \oplus \{q_{J(1)}, \ldots, q_m\}$$



There is no vertex of the form $p + q'$ with $p \neq p_1$

One constraint: Number of vertices, tight upper bound

- If i-th block $\{p_{\mathbf{i}}, \ldots, p_n\} \oplus \{q_{J(\mathbf{i})}, \ldots, q_m\}$ creates $k_i + 2$ vertices involving p_i, then k_i points can be removed from Q.
- Yields bound $\min\{2 \cdot n + m, n + 2 \cdot m\} - 2$

One constraint: Number of vertices, tight upper bound

- If i-th block $\{p_{\mathbf{i}}, \ldots, p_n\} \oplus \{q_{J(\mathbf{i})}, \ldots, q_m\}$ creates $k_i + 2$ vertices involving p_i, then k_i points can be removed from Q.
- Yields bound $\min\{2 \cdot n + m, n + 2 \cdot m\} - 2$

Number of vertices of $\text{conv}(U \oplus V)_{a^T x \geqslant \beta}$ is bounded by

$$\min\{2 \cdot |P| + |Q|, |P| + 2 \cdot |Q|\} - 2.$$

There are $P \subseteq \mathbb{R}^2$ and $Q \subseteq \mathbb{R}^2$ such that $\text{conv}(P \oplus Q)_{x_1 \geqslant 0}$ has $|P| + 2 \cdot |Q| - 2$ points.

One constraint: A linear algorithm

Given $P, Q \subseteq \mathbb{R}^2$ and a constraint $a^T x \geqslant \beta$ where P and Q are sorted w.r.t. $a^T x$, one can compute a $R \subseteq (P \oplus Q)_{a^T x \geqslant \beta}$ containing all vertices of $\operatorname{conv}\left((P \oplus Q)_{a^T x \geqslant \beta}\right)$ in linear time.

Two constraints

- Consider $(P \oplus Q)_{Ax \geqslant b}$ where $Ax \geqslant b$ consists of $a_1^T x \geqslant 0$ and $a_2^T x \geqslant b_2$
- Suppose that P and Q are sorted w.r.t. $a_1^T x$ and $a_2^T x$
- Notice that each subset of P and Q can be sorted w.r.t. $a_1^T x$ and $a_2^T x$ in time $O(N)$

Two constraints

- Consider $(P \oplus Q)_{Ax \geqslant b}$ where $Ax \geqslant b$ consists of $a_1^T x \geqslant 0$ and $a_2^T x \geqslant b_2$
- Suppose that P and Q are sorted w.r.t. $a_1^T x$ and $a_2^T x$
- Notice that each subset of P and Q can be sorted w.r.t. $a_1^T x$ and $a_2^T x$ in time $O(N)$

Divide and conquer algorithm:

- For $\gamma \in \mathbb{R}$ define

$$P_L = \{p \in P \mid a_1^T p < -\gamma\} \quad \text{and} \quad P_R = \{p \in P \mid a_1^T p > -\gamma\}$$
$$Q_L = \{q \in Q \mid a_1^T q < \gamma\} \quad \text{and} \quad Q_R = \{q \in Q \mid a_1^T Q > \gamma\}$$

- If $p \in P_L$ and $q \in Q_L$ then $a_1^T (p + q) < 0$
- Thus $P_L \oplus Q_L$ can be discarded
- $P_R \oplus Q_R$ satisfies $a_1^T x \geqslant 0$.

Two constraints: The algorithm

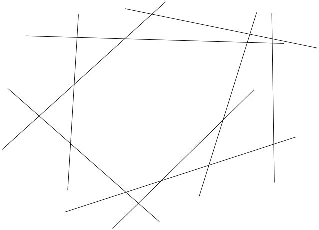
- Determine $\gamma \in \mathbb{R}$ such that $|P_L| + |Q_R| = N/2$
- Recursively compute sets containing vertices of $(P_L \oplus Q_R)_{Ax \geqslant b}$ and $(P_R \oplus Q_L)_{Ax \geqslant b}$
- Compute set containing vertices of $(P_R \oplus Q_R)_{a_2^T x \geqslant \beta}$

Running time:

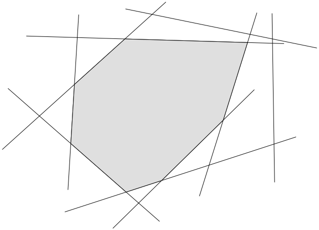
$$\begin{aligned} T(N) &= 2T(n/2) + O(n) \\ &= O(n \log n) \end{aligned}$$

Given $P, Q \subseteq \mathbb{R}^2$ and two constraints $Ax \geqslant b$ one can compute a subset $R \subseteq (P \oplus Q)_{Ax \geqslant b}$ containing all vertices of $\operatorname{conv}((P \oplus Q)_{Ax \geqslant b})$ in time $O(n \log n)$.

A fixed nummber of constraints

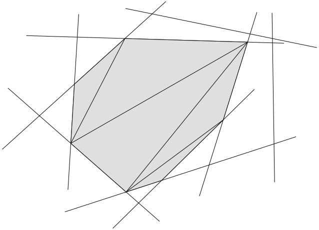


A fixed nummber of constraints



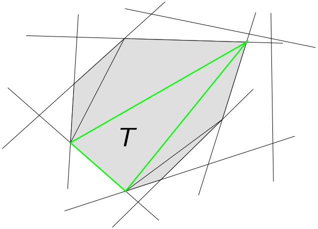
- Compute polygon defined by k constraints $O(k \log k)$

A fixed nummber of constraints



- Compute polygon defined by k constraints $O(k \log k)$
- Triangulate polygon into k triangles

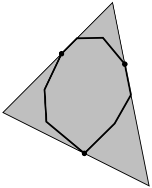
A fixed nummber of constraints



- Compute polygon defined by k constraints $O(k \log k)$
- Triangulate polygon into k triangles

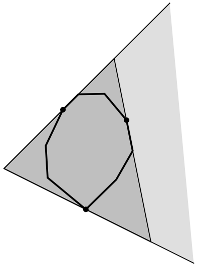
Remains to be shown:
A superset of vertices of $(P \oplus Q) \cap T$ for triangle T can be computed in time $O(N \log N)$

Triangles



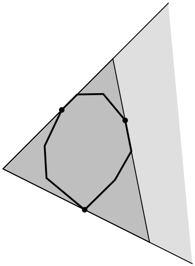
- If each edge of T contains point of $P \oplus Q$, then vertices of three cones contain vertices of triangle

Triangles



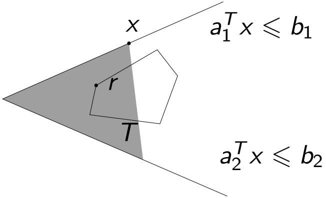
- If each edge of T contains point of $P \oplus Q$, then vertices of three cones contain vertices of triangle

Triangles



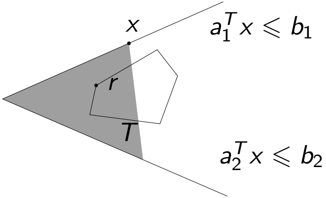
- If each edge of T contains point of $P \oplus Q$, then vertices of three cones contain vertices of triangle
- Goal: Transform T into triangle T' such that:
  - i) $(P \oplus Q) \cap T' = (P \oplus Q) \cap T$.
  - ii) Each edge of T' contains a point of $(P \oplus Q)$.

Triangle: Reducing to two constraints



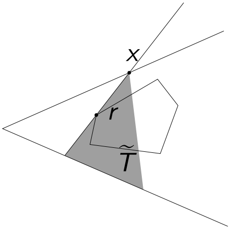
- Consider cone C

Triangle: Reducing to two constraints



- Consider cone C
- Compute a set $R \subseteq (P \oplus Q)$ containing all vertices of $(P \oplus Q) \cap C$

Triangle: Reducing to two constraints



- Consider cone C
- Compute a set $R \subseteq (P \oplus Q)$ containing all vertices of $(P \oplus Q) \cap C$
- Rotate "upper" constraint $a^T x \leqslant \beta$ inwards

Main result

- Given $P \subseteq \mathbb{R}^2$ and $Q \subseteq \mathbb{R}^2$ with $|P| + |Q| = N$, and $Ax \geqslant b$ with $A \in \mathbb{R}^{k \times 2}$ one can compute a set $R \subseteq (P \oplus Q)_{Ax \geqslant b}$ containing all vertices of $(P \oplus Q)_{Ax \geqslant b}$ in time $O(k \log k + k \cdot N \log N)$
- A quasi-convex function can be maximized over $(P \oplus Q)_{Ax \geqslant b}$ in time $O(k \log k + k \cdot N \log N)$

Subsequence problems

- Points corresponding to intervals have first component in $\{1, \ldots, n\}$
- Constraint $x_1 \geqslant 0$ can be "sorted in linear time"
- Length-constrained heaviest segment, DNA copy number data analysis and many other subsequence problems from computational biology can be computed in linear time

Open problems

- How many vertices has constrained Minkowski sum with two constraints.

  Our bound is $O(N \log N)$

- More generally: Let $S \subseteq P \oplus Q$ be arbitrary subset. How many vertices has $\text{conv}(S)$?

  $O(N^{3/2})$ is best upper bound.