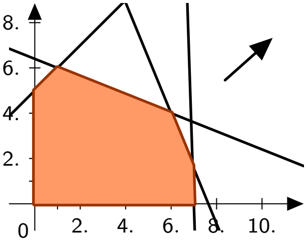
Smoothed Analysis of the Simplex Method

Sophie Huiberts (CWI)
joint work with Daniel Dadush (CWI)

Linear Programming

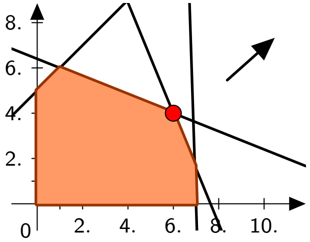
$$\text{maximize } c^{\mathsf{T}} x$$
$$\text{subject to } Ax \leq b$$



- d variables
- n constraints

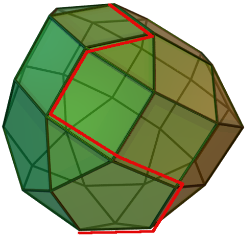
Linear Programming

$$\text{maximize } c^{\mathsf{T}}x$$
$$\text{subject to } Ax \leq b$$

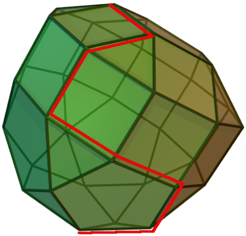


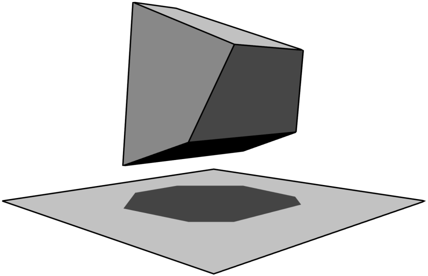
- d variables
- n constraints

Simplex Method



Simplex Method

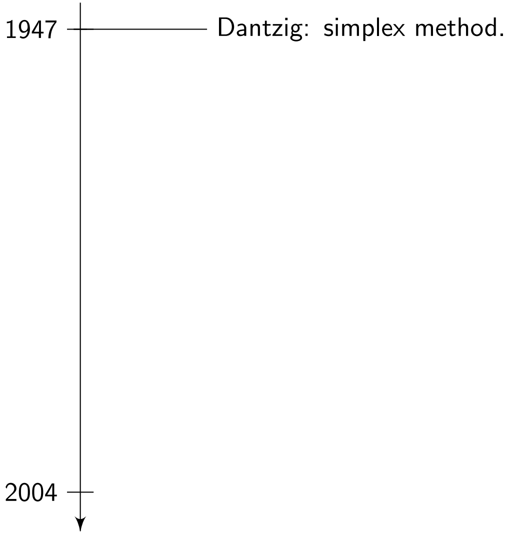




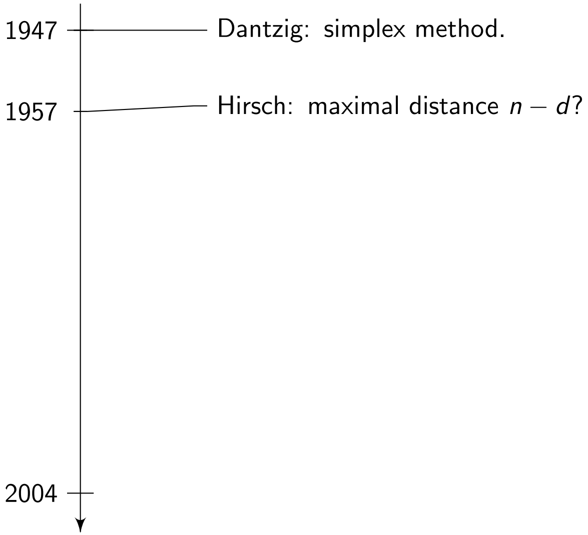
$\approx 3(n+d)$ steps in practice.

$\approx n^{\lfloor d/2 \rfloor}$ steps in theoretical worst case.

Simplex method: A short history



Simplex method: A short history



Simplex method: A short history

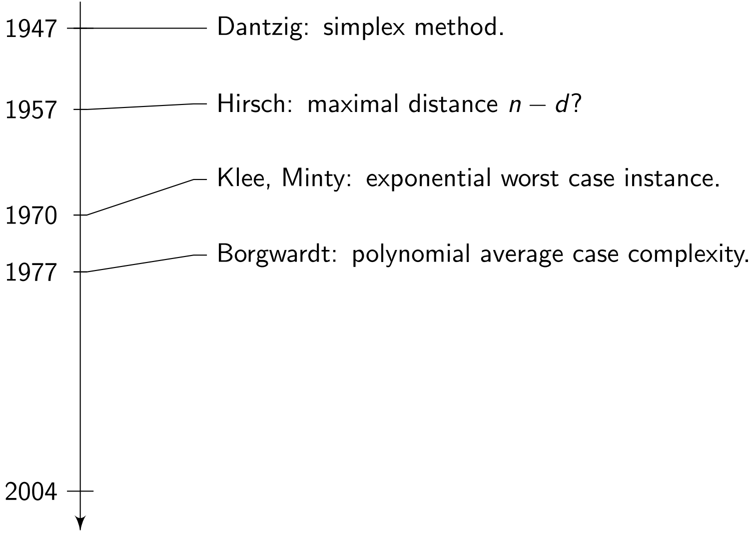
1947 — Dantzig: simplex method.

1957 — Hirsch: maximal distance $n - d$?

1970 — Klee, Minty: exponential worst case instance.

2004

Simplex method: A short history



Simplex method: A short history

- 1947 — Dantzig: simplex method.
- 1957 — Hirsch: maximal distance $n - d$?
- 1970 — Klee, Minty: exponential worst case instance.
- 1977 — Borgwardt: polynomial average case complexity.
- 1992 — Kalai, Kleitman: paths of length $n^{\log d+2}$ exist.
- 1992 — Kalai: sub-exponential pivot rule.
- 2004

Simplex method: A short history

- 1947 — Dantzig: simplex method.
- 1957 — Hirsch: maximal distance $n - d$?
- 1970 — Klee, Minty: exponential worst case instance.
- 1977 — Borgwardt: polynomial average case complexity.
- 1992 — Kalai, Kleitman: paths of length $n^{\log d+2}$ exist.
- 1992 — Kalai: sub-exponential pivot rule.
- 2004 — Spielman-Teng: polynomial smoothed complexity.

Average-case analysis

$$\begin{aligned} \text{maximize } & c^{\mathsf{T}} x \\ \text{subject to } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

- $b = 1$, Rows of A sampled from a rotationally symmetric distribution (RSD). $O(n^{1/d} d^3)$.
  [Borgwardt '77,'82,'87,'99]

Average-case analysis

$$\begin{aligned} \text{maximize } & c^T x \\ \text{subject to } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

- $b = 1$, Rows of A sampled from a rotationally symmetric distribution (RSD). $O(n^{1/d} d^3)$.
  [Borgwardt '77,'82,'87,'99]
- Rows of A,b,c sampled independently from an RSD.
  [Smale '83, Megiddo '86]

Average-case analysis

$$\begin{aligned} \text{maximize } & c^{\mathsf{T}} x \\ \text{subject to } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

- $b = 1$, Rows of A sampled from a rotationally symmetric distribution (RSD). $O(n^{1/d} d^3)$.
  [Borgwardt '77,'82,'87,'99]
- Rows of A,b,c sampled independently from an RSD.
  [Smale '83, Megiddo '86]
- Fixed data. Flip signs of constraints at random. $O(\min\{n^2, d^2\})$.
  [Adler '83, Haimovich '83 Adler, Megiddo '85, Todd '86, Adler,Karp,Shamir '87]

Random is Not Typical

Random is Not Typical

Smoothed Complexity (Spielman, Teng ' 01)

Worst case, $\sigma = 0$

Smoothed analysis, σ variable

Defining polynomial smoothed complexity

- $c \in \mathbb{R}^d, \bar{A} \in \mathbb{R}^{n \times d}, \bar{b} \in \mathbb{R}^n$. Rows of $(\bar{A}, \bar{b})$ norm at most 1.

Defining polynomial smoothed complexity

- $c \in \mathbb{R}^d, \bar{A} \in \mathbb{R}^{n \times d}, \bar{b} \in \mathbb{R}^n$. Rows of $(\bar{A}, \bar{b})$ norm at most 1.
- $\hat{A}, \hat{b}$: entries iid $N(0, \sigma^2)$.
- $A = \bar{A} + \hat{A}, \ b = \bar{b} + \hat{b}$.

Defining polynomial smoothed complexity

- $c \in \mathbb{R}^d, \bar{A} \in \mathbb{R}^{n \times d}, \bar{b} \in \mathbb{R}^n$. Rows of $(\bar{A}, \bar{b})$ norm at most 1.
- $\hat{A}, \hat{b}$: entries iid $N(0, \sigma^2)$.
- $A = \bar{A} + \hat{A}, \; b = \bar{b} + \hat{b}$.

- **Smoothed Linear Program:**

$$\begin{array}{ll} \text{maximize} & c^{\mathsf{T}} x \\ \text{subject to} & Ax \le b. \end{array}$$

Defining polynomial smoothed complexity

- $c \in \mathbb{R}^d, \bar{A} \in \mathbb{R}^{n \times d}, \bar{b} \in \mathbb{R}^n$. Rows of $(\bar{A}, \bar{b})$ norm at most 1.
- $\hat{A}, \hat{b}$: entries iid $N(0, \sigma^2)$.
- $A = \bar{A} + \hat{A},\ b = \bar{b} + \hat{b}$.

- **Smoothed Linear Program:**

$$\begin{aligned} \text{maximize} \quad & c^\mathsf{T} x \\ \text{subject to} \quad & Ax \leq b. \end{aligned}$$

Polynomial smoothed complexity: expected poly(n, d, σ^{-1}) pivots.

Why smoothed analysis?

- Only looks at large-scale structure.
- Models measurement error, rounding error, ...
- Hard instances are a vanishingly small subset of every neighborhood.

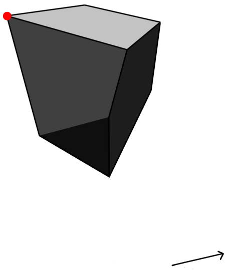
Results: smoothed complexity bounds

- d variables.
- n constraints.
- $N(0, \sigma^2)$ Gaussian noise.

Works	Expected Number of Pivots
Spielman, Teng '04	$\widetilde{O}(n^{86}d^{55}\sigma^{-30} + n^{86}d^{70})$
Vershynin '09	$O(d^3 \ln^3 n\sigma^{-4} + d^9 \ln^7 n)$
Dadush, H. '18	$O(d^2\sqrt{\ln n}\sigma^{-2} + d^3 \ln^{3/2} n)$

9 out of 10 Theoreticians recommend: Shadow Vertex rule

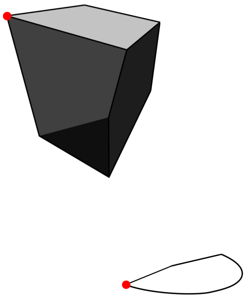
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

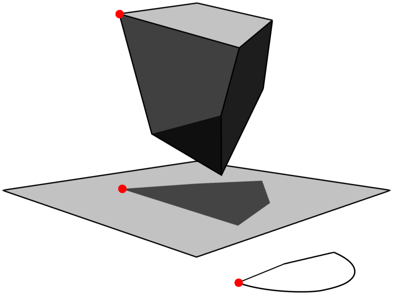
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

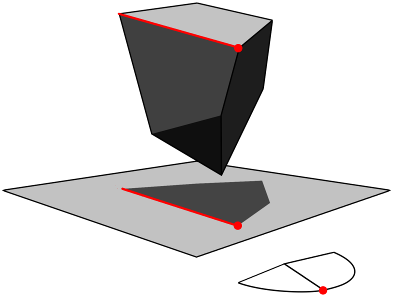
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

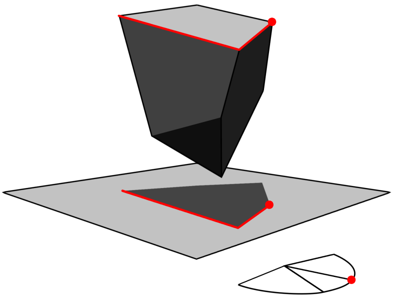
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

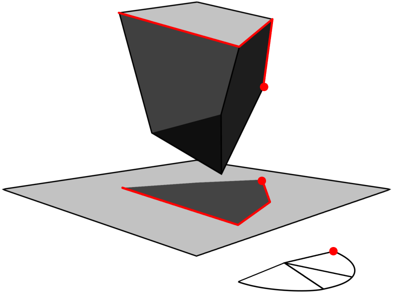
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

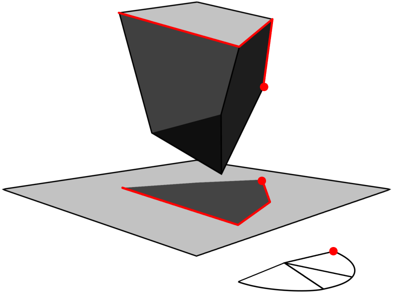
1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1 - \lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



Gass, Saaty '55: shadow vertex rule.

9 out of 10 Theoreticians recommend: Shadow Vertex rule

1. Start at vertex x optimizing an objective $c' \in \mathbb{R}^d$.
2. $c_\lambda := \lambda c + (1-\lambda)c'$.
3. Increase λ from 0 to 1, tracking optimal vertex for c_λ.



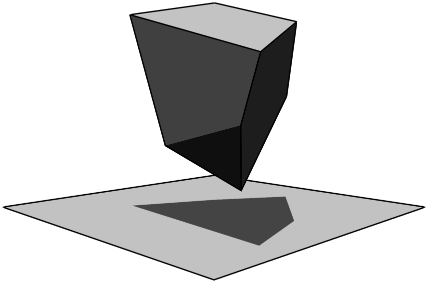
Gass, Saaty '55: shadow vertex rule.

Why work with shadow vertex rule?

Can **locally** determine if a vertex is on the path.

Fundamental estimate: number of shadow edges

- $P = \{x : Ax \leq 1\}$.
- A smoothed.
- RHS 1 fixed.
- W fixed 2D plane.



Shadow bound := Expected # vertices in projection of P onto W.

Results: shadow size

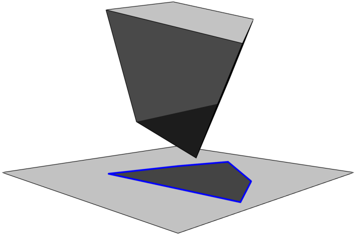
- Shadow of $P = \{x : Ax \leq 1\}$ on fixed 2D plane W.
- d variables.
- n constraints.
- σ standard deviation.

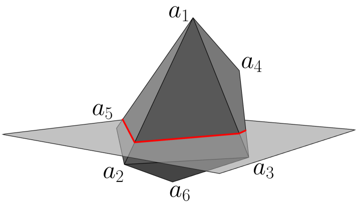
Works	Expected Number of Vertices
Spielman, Teng '04	$O(d^3 n\sigma^{-6} + d^6 n \ln^3 n)$
Deshpande, Spielman '05	$O(dn^2 \ln n \; \sigma^{-2} + d^2 n^2 \ln^2 n)$
Vershynin '09	$O(d^3 \sigma^{-4} + d^5 \ln^2 n)$
Dadush, H. '18	$O(d^2 \sqrt{\ln n} \; \sigma^{-2} + d^{2.5} \ln^{3/2} n \; (1 + \sigma^{-1}))$

Polyhedral duality

$P := \{x : a_i^T x \leq 1 \ \forall i \leq n\}$.

$Q := \text{ConvexHull}(a_1, \ldots, a_n)$



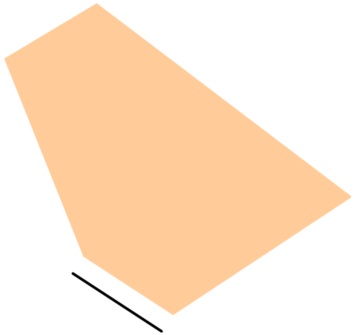


number of vertices in $\pi_W(P) \leq$ number of edges in $Q \cap W$.

Counting polyhedron edges: comparing lengths

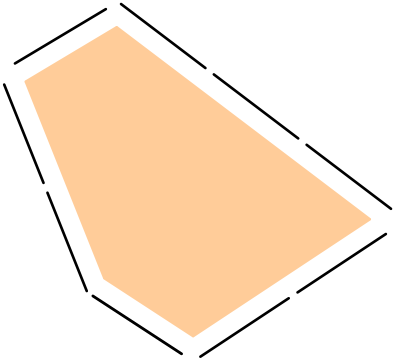
$$\#\text{edges} \leq \frac{\text{perimeter}}{\text{minimum edge length}}$$

Counting polyhedron edges: comparing lengths



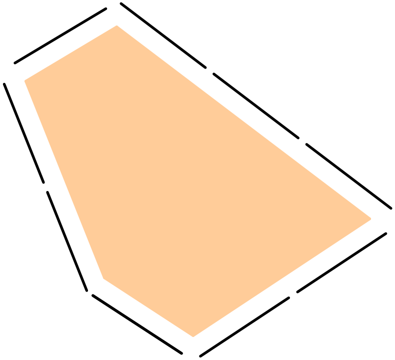
$$\#\text{edges} \leq \frac{\text{perimeter}}{\text{minimum edge length}}$$

Counting polyhedron edges: comparing lengths



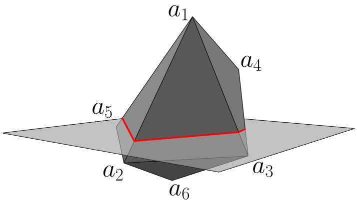
$$\#\text{edges} \leq \frac{\text{perimeter}}{\text{minimum edge length}}$$

Counting polyhedron edges: comparing lengths



$$\mathbb{E}[\#\text{edges}] \leq \frac{\mathbb{E}[\text{perimeter}]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$

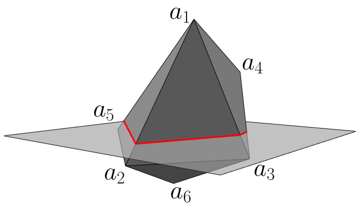
Counting shadow edges: proof of lemma



Let $E(B)$ be the event that $\mathrm{conv}(B) \cap W$ forms an edge of $Q \cap W$.

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] = \sum_{\substack{B \subset \{a_1, \ldots, a_n\} \\ |B| = d}} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \Pr[E(B)]$$

Counting shadow edges: proof of lemma

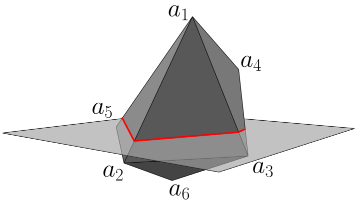


Let $E(B)$ be the event that $\mathrm{conv}(B) \cap W$ forms an edge of $Q \cap W$.

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] = \sum_{\substack{B \subset \{a_1, \ldots, a_n\} \\ |B| = d}} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \Pr[E(B)]$$

$$\geq \min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \sum_{|B'|=d} \Pr[E(B')]$$

Counting shadow edges: proof of lemma



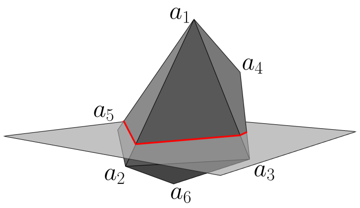
Let $E(B)$ be the event that $\mathrm{conv}(B) \cap W$ forms an edge of $Q \cap W$.

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] = \sum_{\substack{B \subset \{a_1,\dots,a_n\} \\ |B|=d}} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \Pr[E(B)]$$

$$\geq \min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \sum_{|B'|=d} \Pr[E(B')]$$

$$= \min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \mathbb{E}[\#\mathrm{edges}]$$

Counting shadow edges: proof of lemma



Let $E(B)$ be the event that $\mathrm{conv}(B) \cap W$ forms an edge of $Q \cap W$.

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] = \sum_{\substack{B \subset \{a_1, \ldots, a_n\} \\ |B| = d}} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \Pr[E(B)]$$

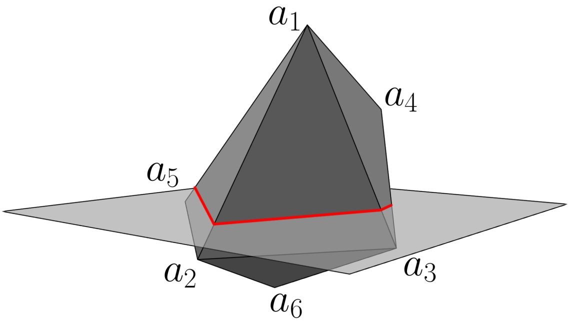
$$\geq \min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \sum_{|B'|=d} \Pr[E(B')]$$

$$= \min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) \mid E(B)] \mathbb{E}[\#\mathrm{edges}]$$

So $\mathbb{E}[\#\mathrm{edges}] \leq \frac{\mathbb{E}[\mathrm{perimeter}(Q \cap W)]}{\min_{|B|=d} \mathbb{E}[\mathrm{length}(\mathrm{conv}(B) \cap W) | E(B)]}$.

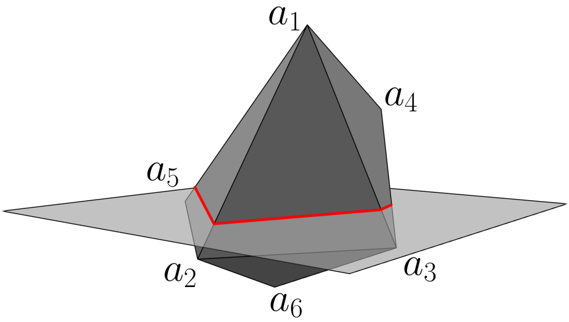
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



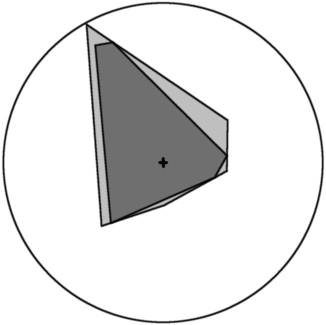
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



High-level ideas

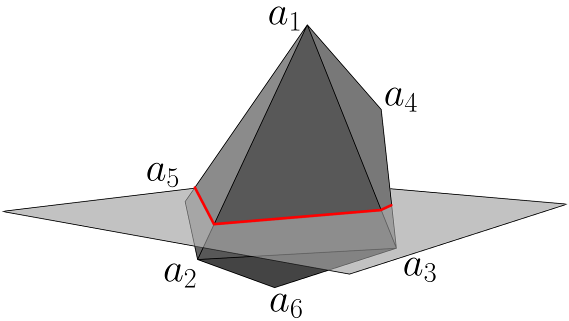
$$\mathbb{E}[\#\text{edges}(Q \cap W)] \le \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



$$\begin{aligned}\mathbb{E}[\text{perimeter}(Q \cap W)] \\ &\le 2\pi \mathbb{E}[\max_{x \in Q \cap W} \|x\|] \\ &\le 2\pi \mathbb{E}[\max \|\pi_W(a_i)\|] \\ &\le O(1 + \sigma\sqrt{\ln n})\end{aligned}$$

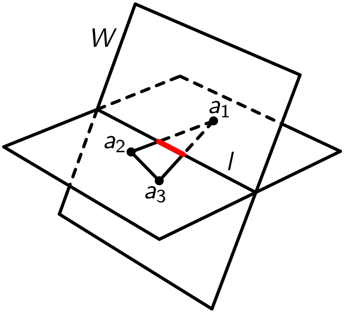
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



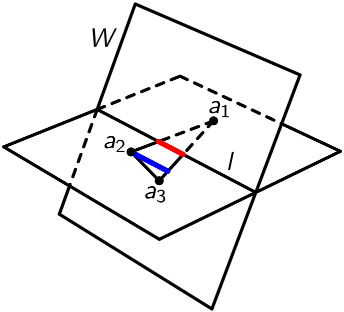
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



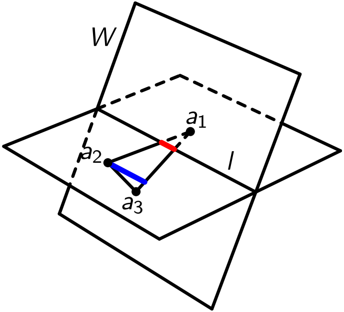
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



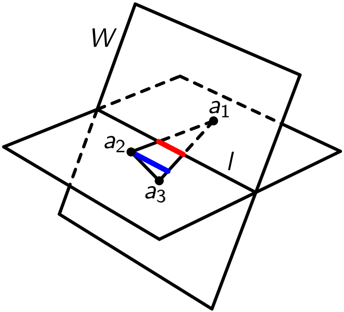
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



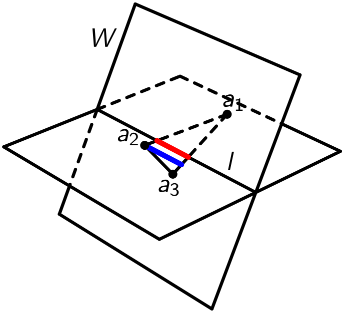
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



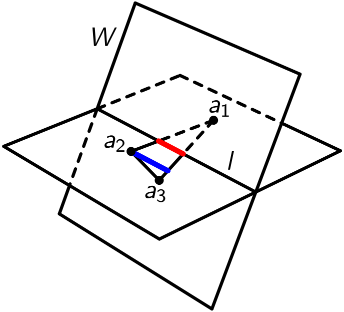
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



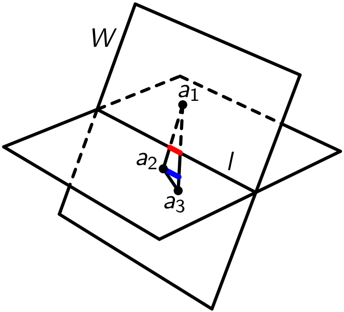
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



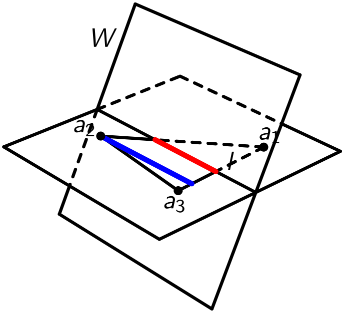
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



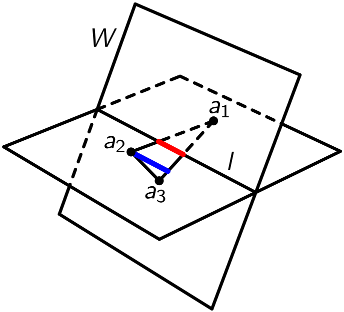
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



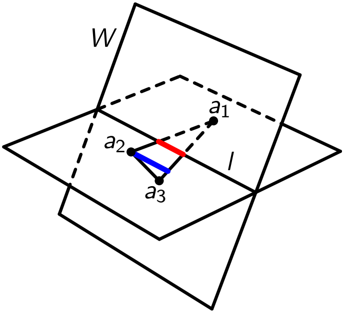
High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



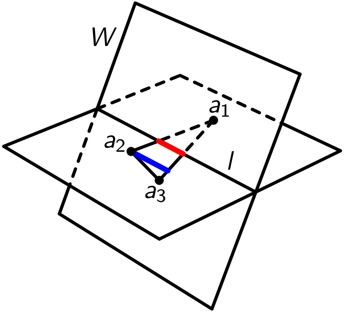
$\mathbb{E}$[height of simplex along l]

$$\geq \Omega(\tau/\sqrt{d})$$

	Parameter Description	Gaussian
τ	a_i restricted to line has variance at least τ^2	σ

High-level ideas

$$\mathbb{E}[\#\text{edges}(Q \cap W)] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\text{minimum } \mathbb{E}[\text{ edge length }]}$$



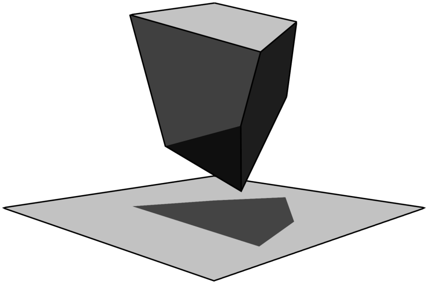
$$\mathbb{E}[\text{height of simplex along } l] \geq \Omega(\tau/\sqrt{d})$$

$$\mathbb{E}\left[\frac{\text{intersection length}}{\text{height of simplex along } l}\right] \geq \Omega\left(\frac{1}{dL(1+R)}\right)$$

	Parameter Description	Gaussian
τ	noise variance when restricted to a line	σ
R	High-probability noise norm bound	$O(\sigma\sqrt{d \ln n})$
L	log-Lipschitzness of prob. density within radius R	$O(\sigma^{-1}\sqrt{d \ln n})$

Fundamental estimate: number of shadow edges

- $P = \{x : Ax \leq 1\}$.
- A smoothed.
- RHS 1 fixed.
- W fixed 2D plane.



Shadow bound $:= O(d^2\sigma^{-2}\sqrt{\ln n} + d^{2.5}\sigma^{-1}\ln n + d^{2.5}\ln^{1.5} n)$

Open Problems

1. Improve upper/lower bounds on expected shadow size.
2. Sparse noise? Bounded noise?
3. Diameter bounds for random polyhedra?