IBM Research

Hypercube locality-sensitive hashing for approximate near neighbors

Thijs Laarhoven

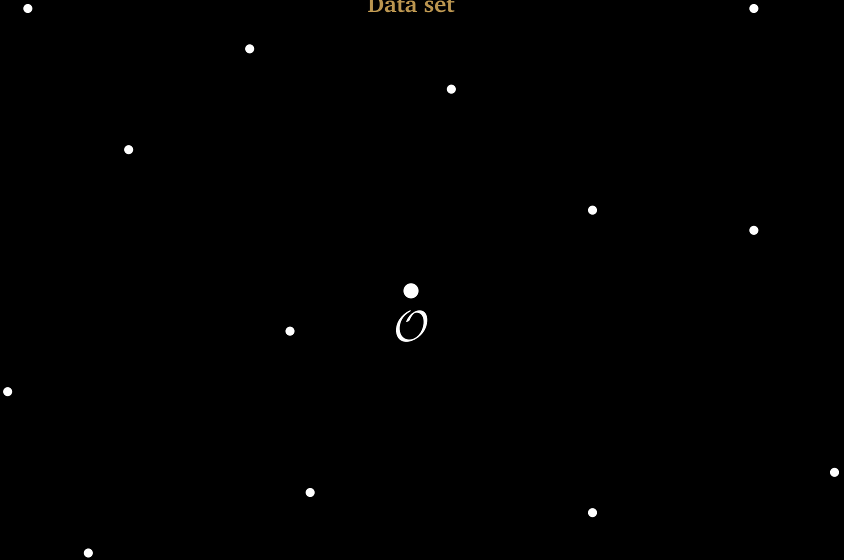
mail@thijs.com
http://www.thijs.com/

MFCS 2017, Aalborg, Denmark
(August 23, 2017)

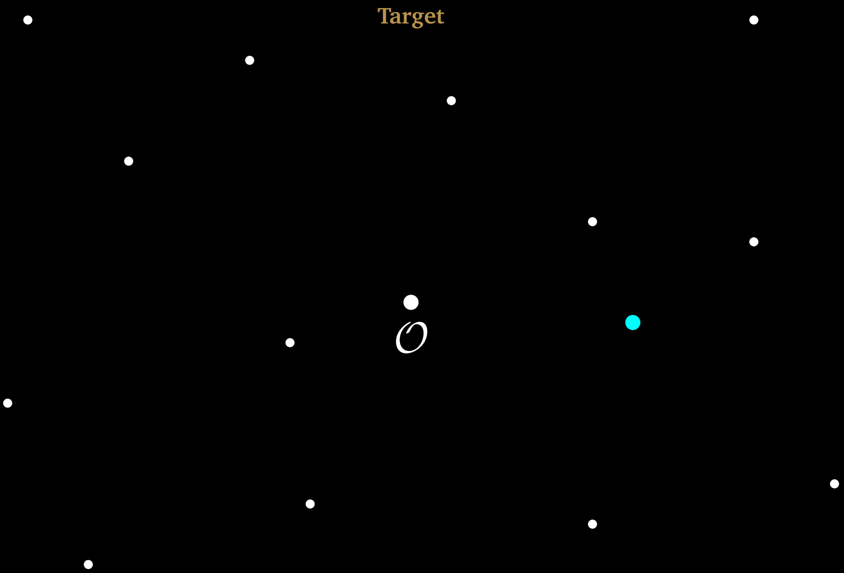
Nearest neighbor searching

Nearest neighbor searching

Data set

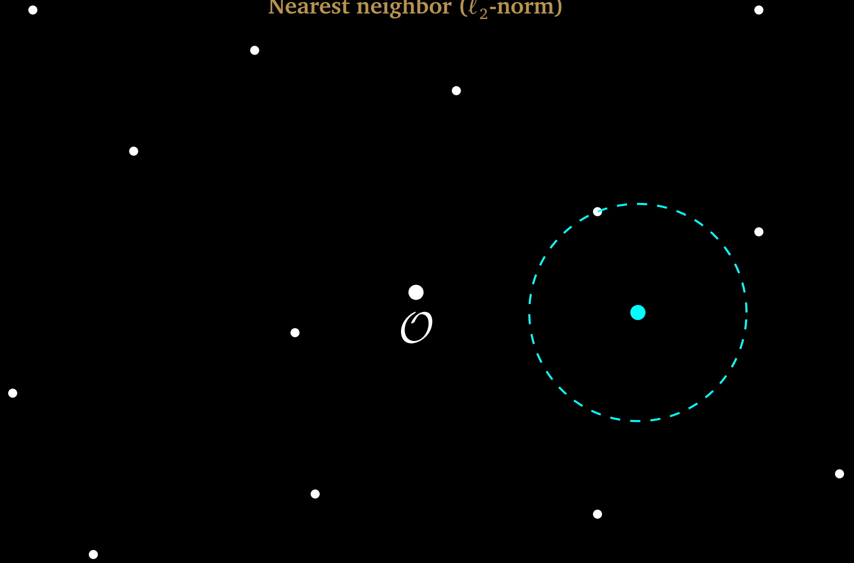


Nearest neighbor searching



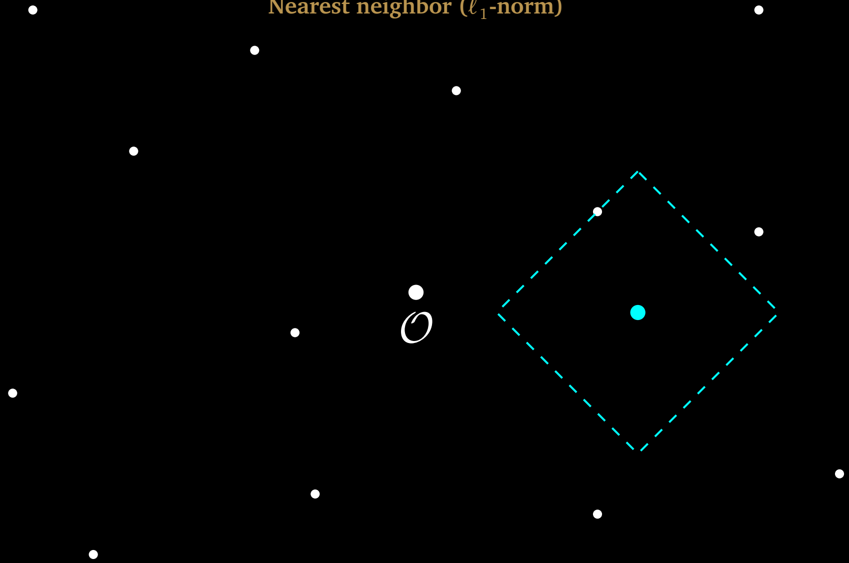
Nearest neighbor searching

Nearest neighbor (ℓ_2-norm)



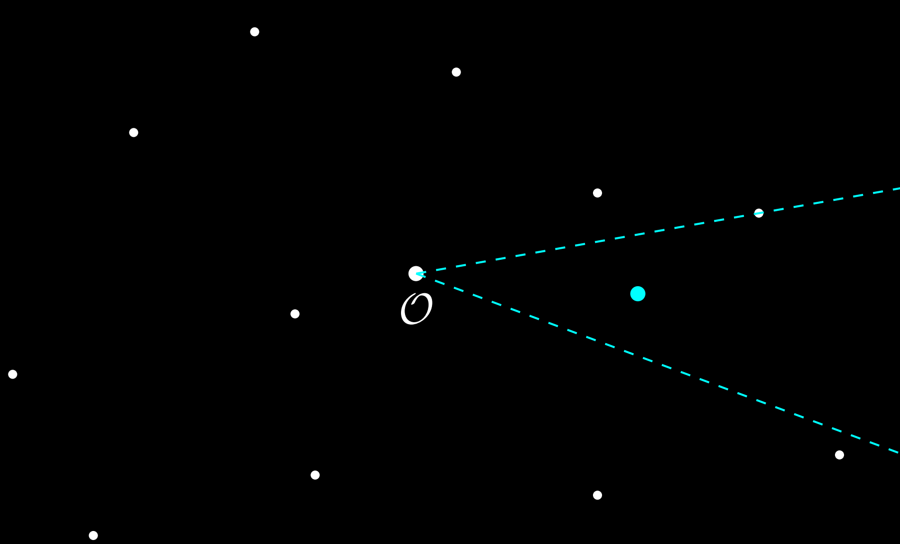
Nearest neighbor searching

Nearest neighbor (ℓ_1-norm)



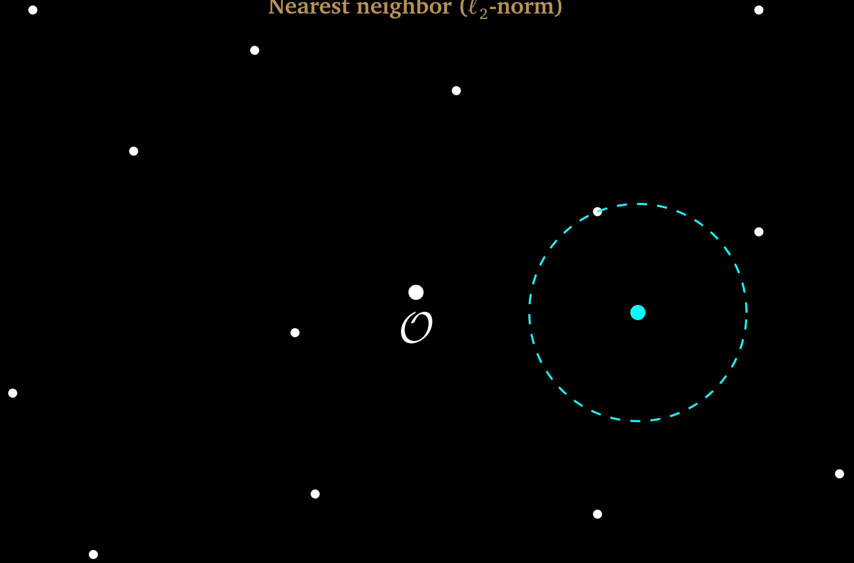
Nearest neighbor searching

Nearest neighbor (angular distance)



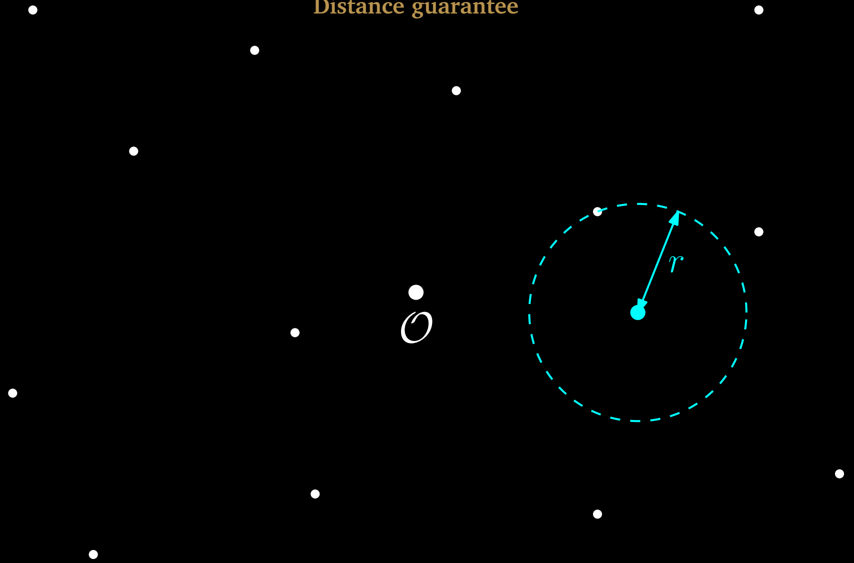
Nearest neighbor searching

Nearest neighbor (ℓ_2-norm)



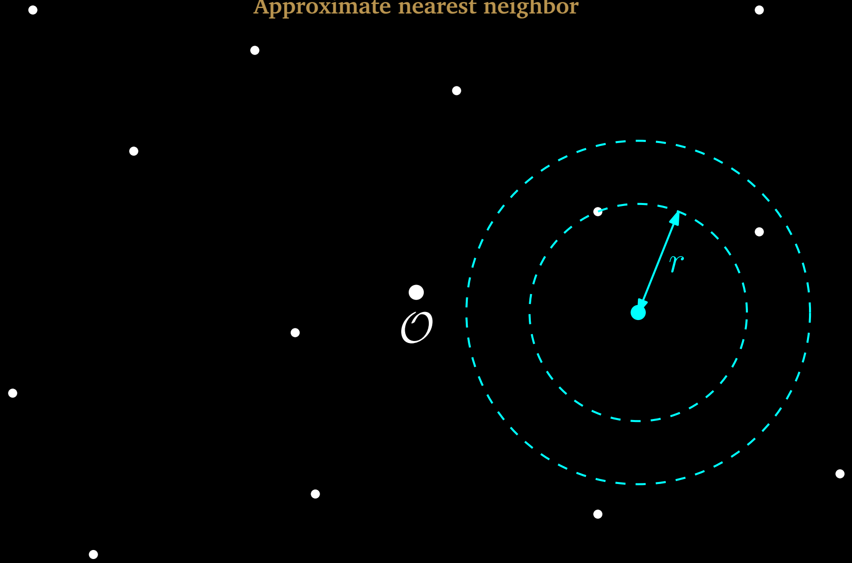
Nearest neighbor searching

Distance guarantee



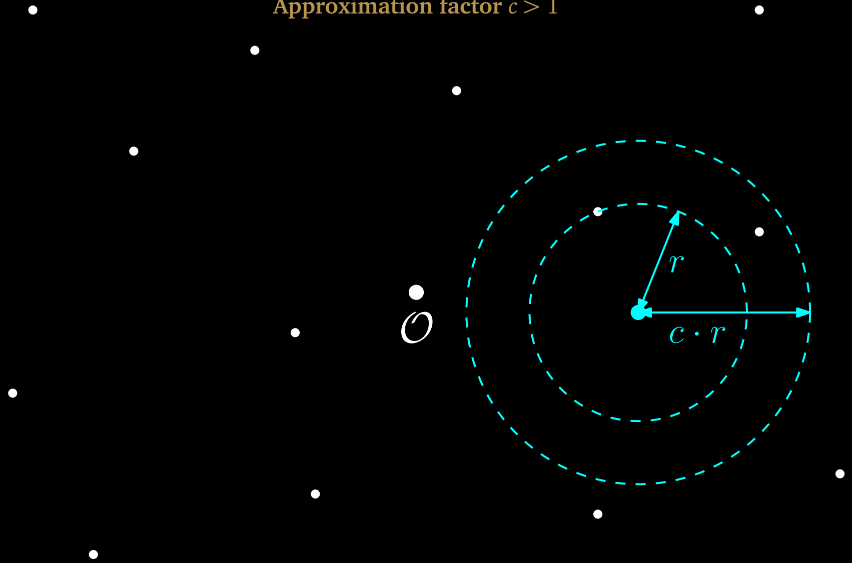
Nearest neighbor searching

Approximate nearest neighbor



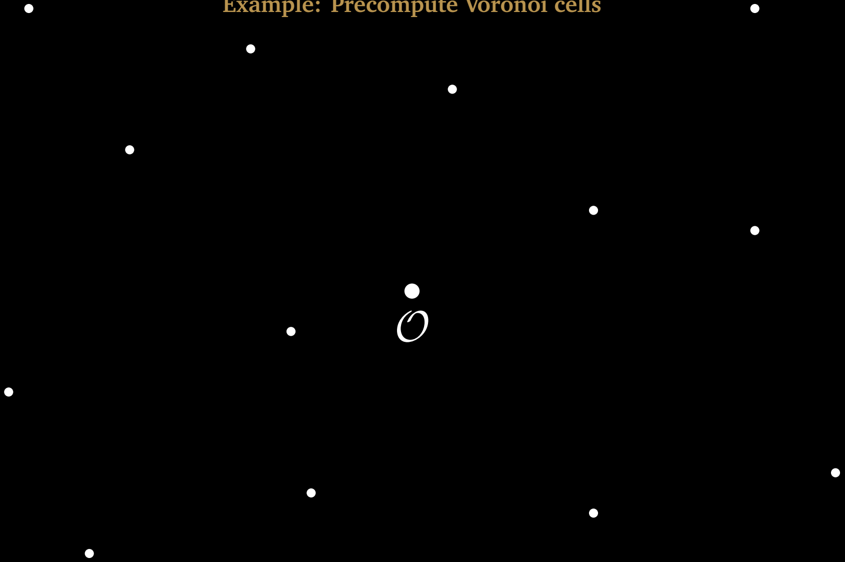
Nearest neighbor searching

Approximation factor $c > 1$



Nearest neighbor searching

Example: Precompute Voronoi cells

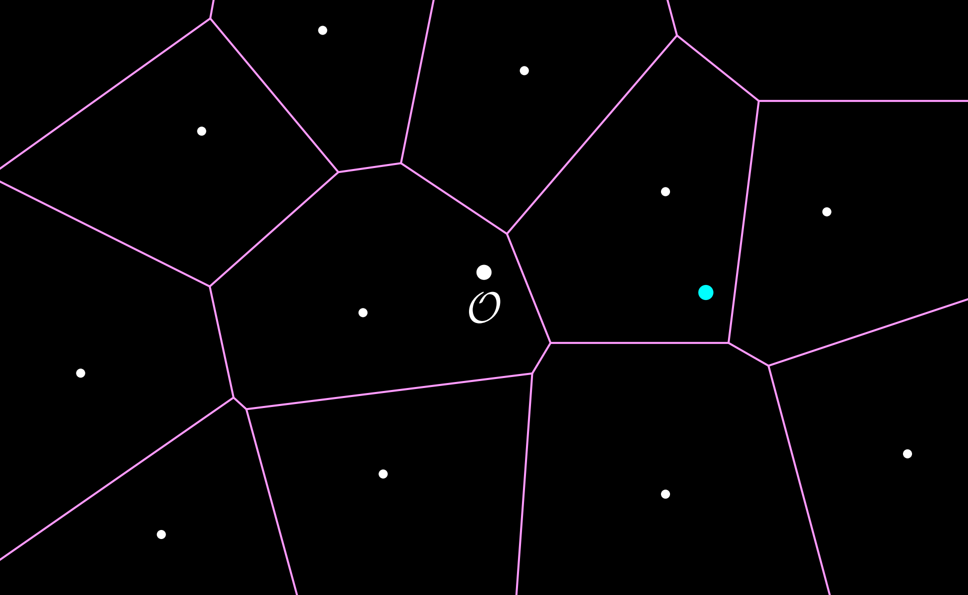


Nearest neighbor searching

Example: Precompute Voronoi cells

Nearest neighbor searching

Given a target...



Nearest neighbor searching

...quickly find the right cell

Nearest neighbor searching

Works well in low dimensions

Nearest neighbor searching

Problem setting

- High dimensions d

Nearest neighbor searching

Problem setting

- High dimensions d
- Large data set of size $n = 2^{\Omega(d/\log d)}$
  - Smaller n? $\implies$ Use JLT to reduce d

Nearest neighbor searching

Problem setting

- High dimensions d
- Large data set of size $n = 2^{\Omega(d/\log d)}$
  - Smaller n? $\implies$ Use JLT to reduce d
- Assumption: Data set lies on the sphere
  - Equivalent to angular distance/cosine similarity in all of $\mathbb{R}^d$
  - Reduction from Eucl. NNS in $\mathbb{R}^d$ to Eucl. NNS on the sphere [AR'15]

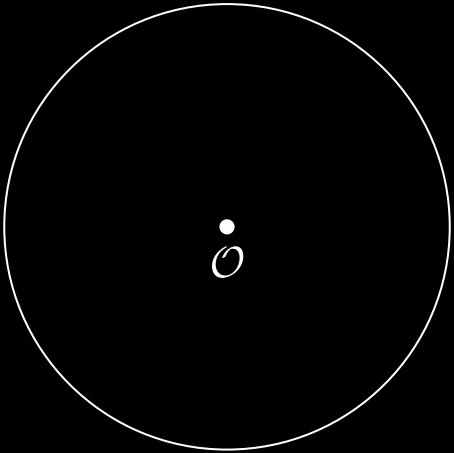
Nearest neighbor searching

Problem setting

- High dimensions d
- Large data set of size $n = 2^{\Omega(d/\log d)}$
  - Smaller n? $\implies$ Use JLT to reduce d
- Assumption: Data set lies on the sphere
  - Equivalent to angular distance/cosine similarity in all of $\mathbb{R}^d$
  - Reduction from Eucl. NNS in $\mathbb{R}^d$ to Eucl. NNS on the sphere [AR'15]
- Goal: Query time $O(n^\rho)$ with $\rho < 1$

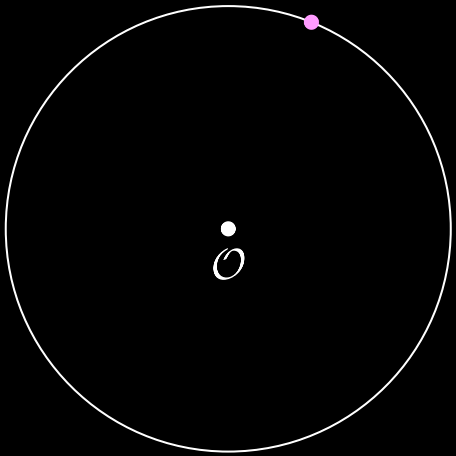
Hyperplane LSH

[Charikar, STOC'02]



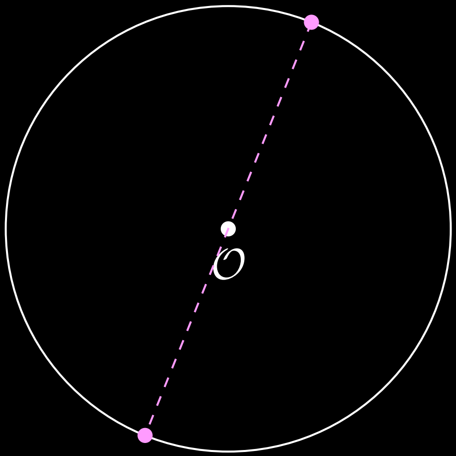
Hyperplane LSH

Random point



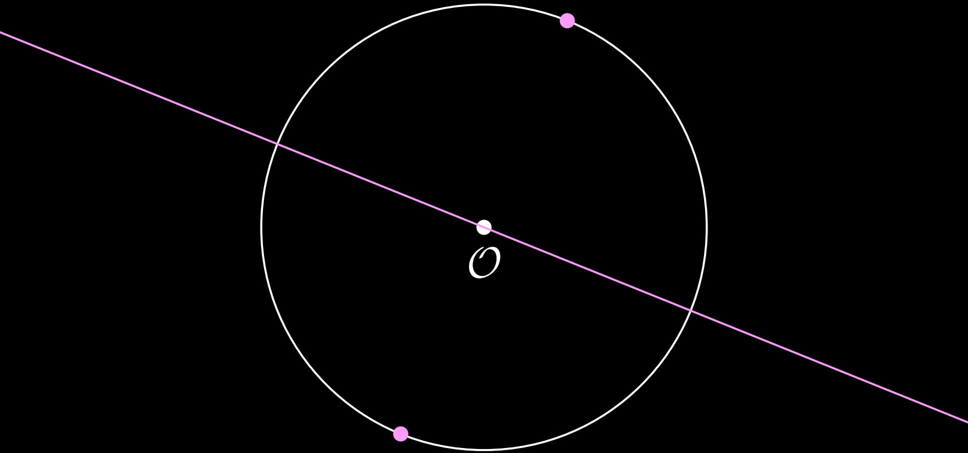
Hyperplane LSH

Opposite point



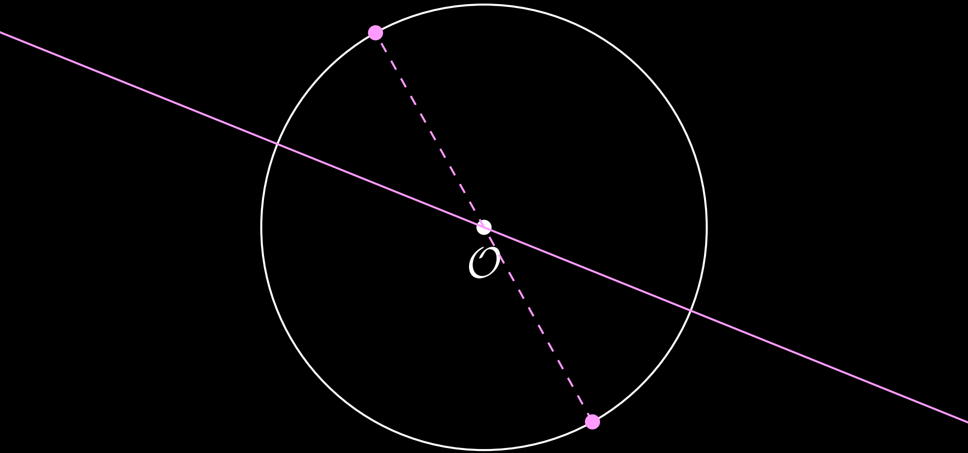
Hyperplane LSH

Two Voronoi cells



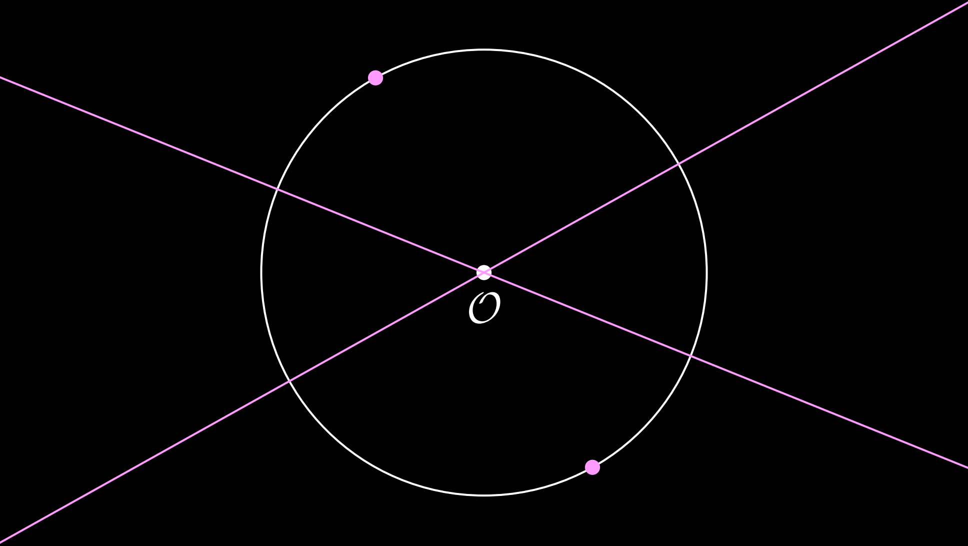
Hyperplane LSH

Another pair of points



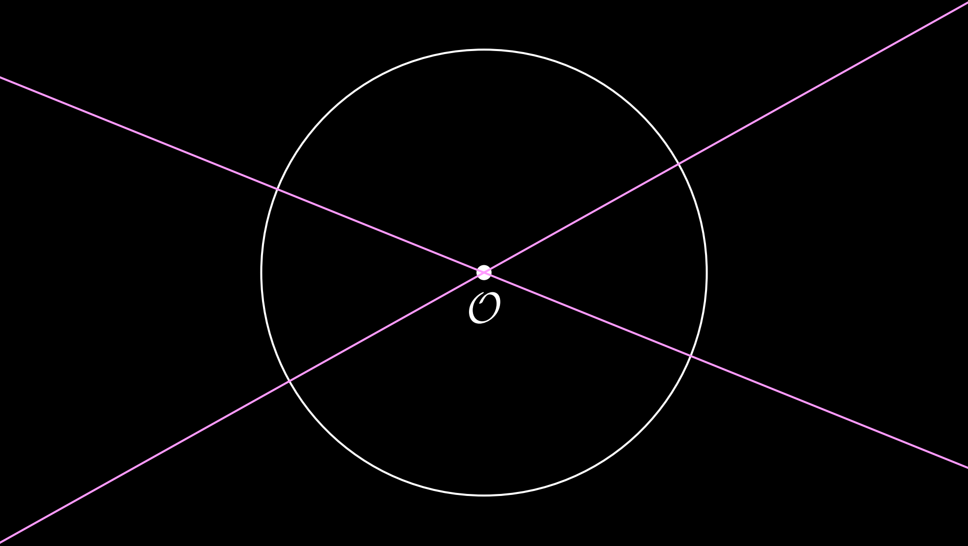
Hyperplane LSH

Another hyperplane



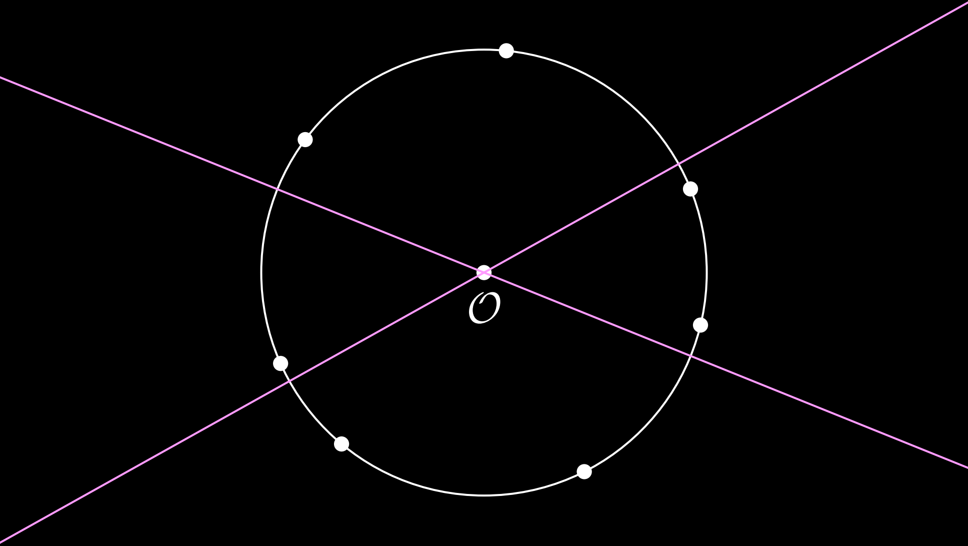
Hyperplane LSH

Defines partition



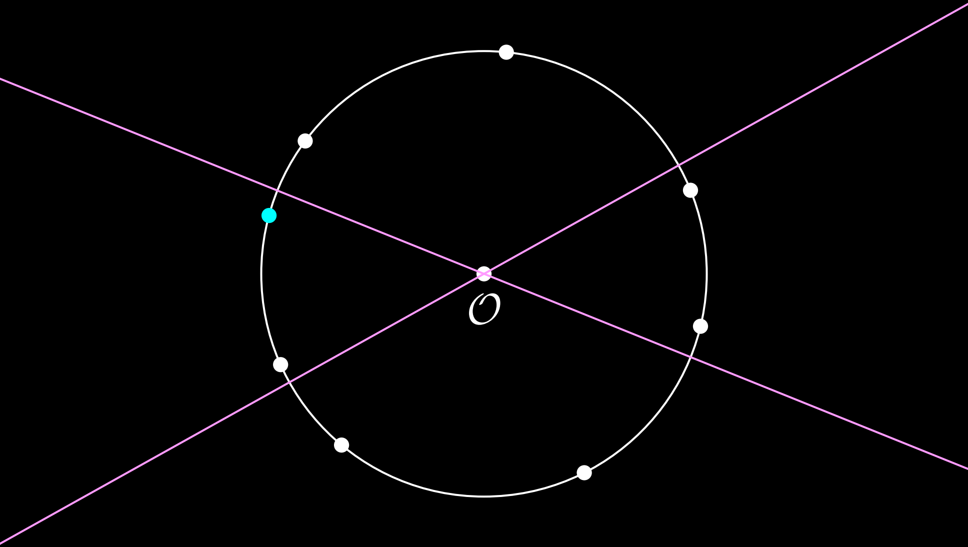
Hyperplane LSH

Preprocessing



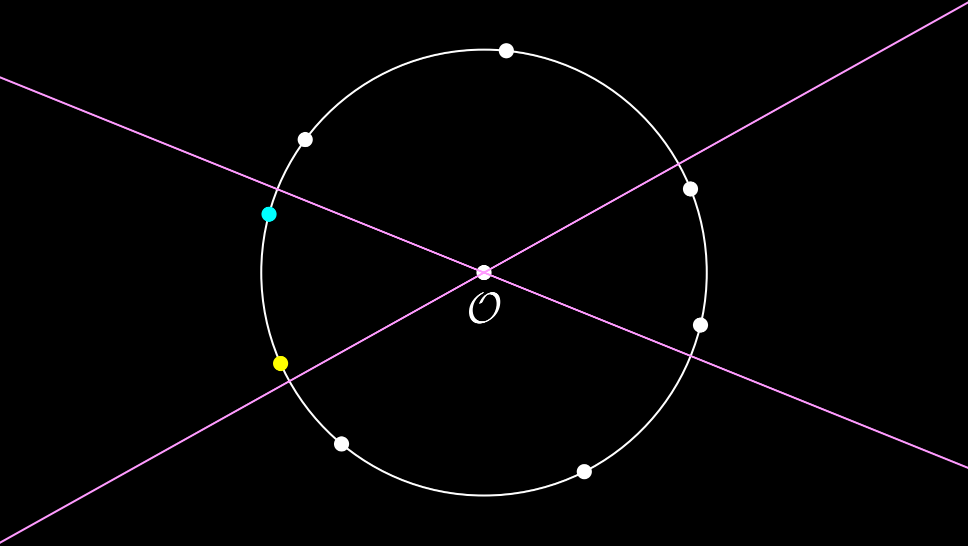
Hyperplane LSH

Query



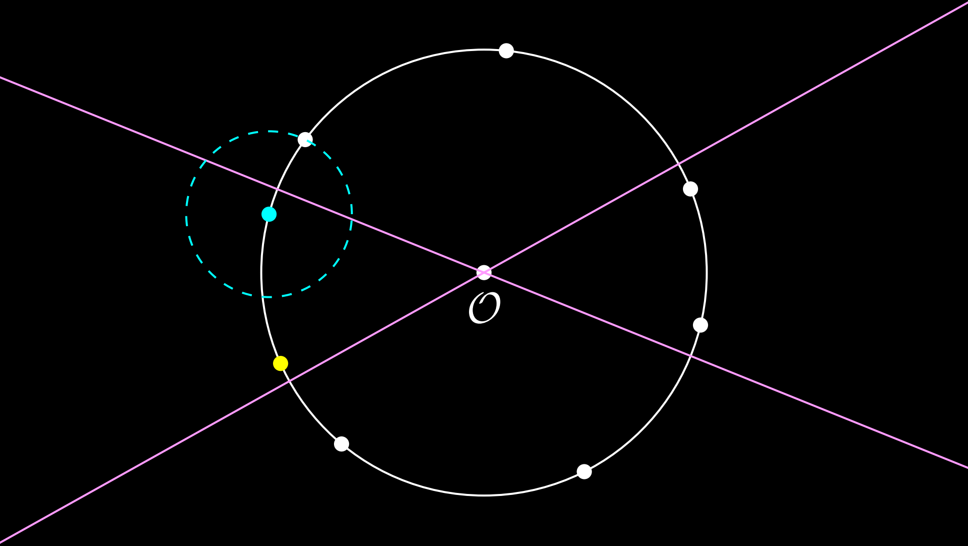
Hyperplane LSH

Collisions



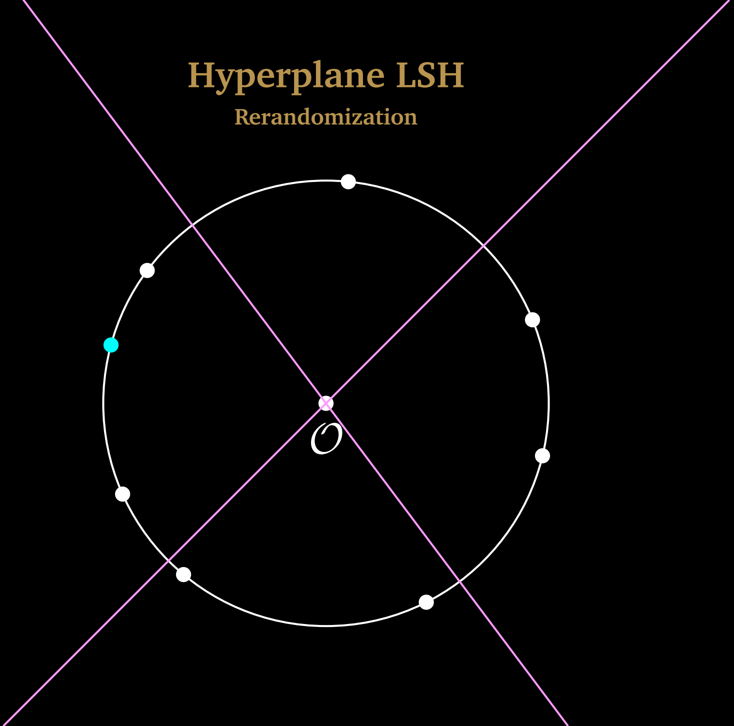
Hyperplane LSH

Failure



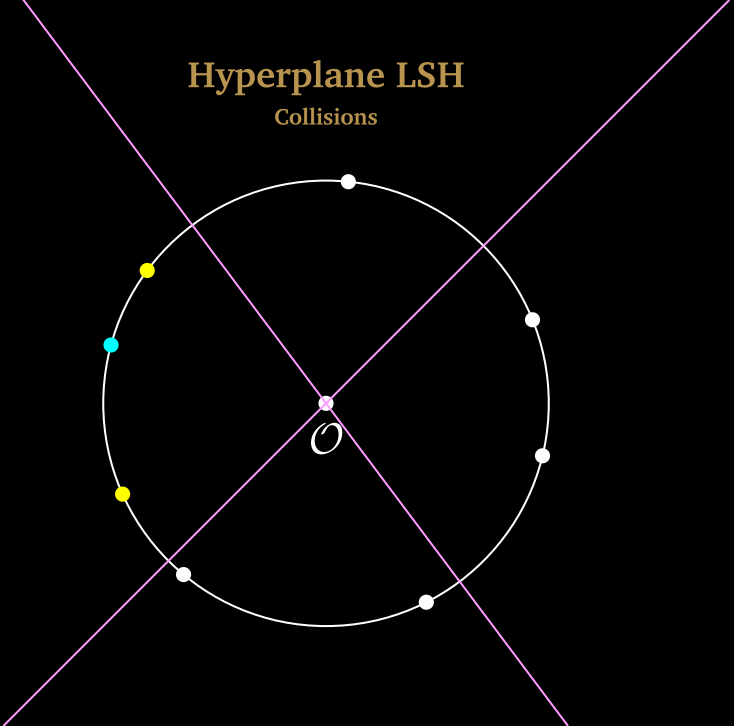
Hyperplane LSH

Rerandomization



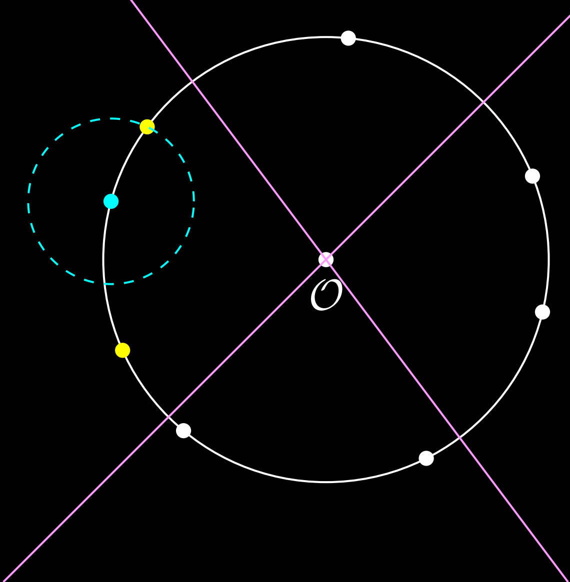
Hyperplane LSH

Collisions



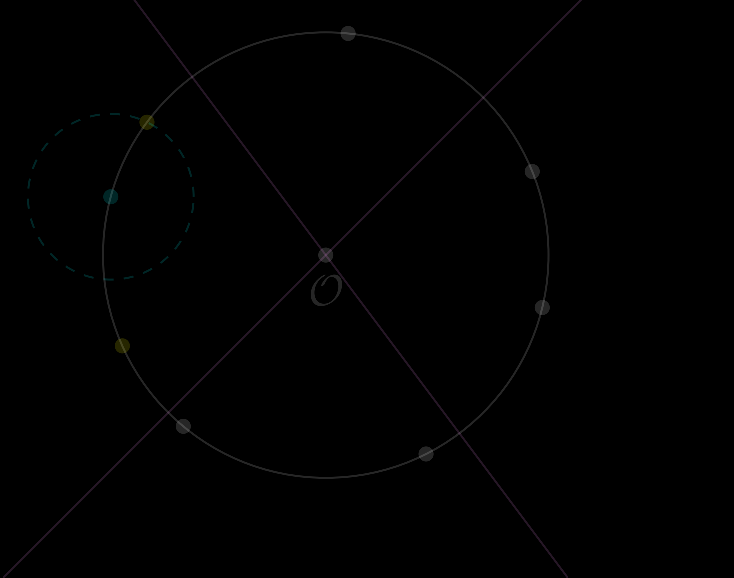
Hyperplane LSH

Success



Hyperplane LSH

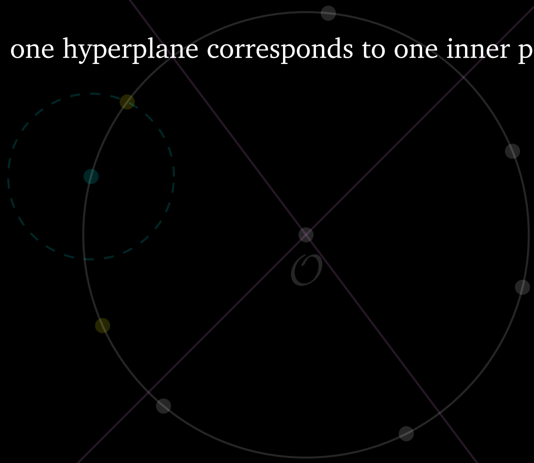
Overview



Hyperplane LSH

Overview

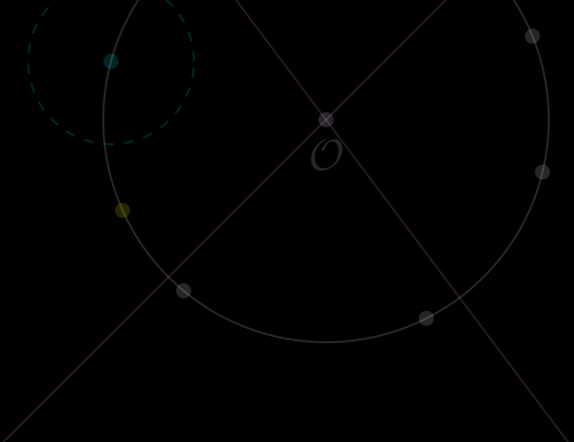
- Simple: one hyperplane corresponds to one inner product



Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ



Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - Random points [AI'06, AINR'14, ...]

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - Random points [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - Random points [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]
  - Classical root lattices A_d, D_d [JASG'08]
  - Exceptional root lattices $E_{6,7,8}$, F_4, G_2 [JASG'08]

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - Random points [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]
  - Classical root lattices A_d, D_d [JASG'08]
  - Exceptional root lattices $E_{6,7,8}$, F_4, G_2 [JASG'08]
  - Cross-polytopes [TT'07, AILRS'15, KW'17]

Hyperplane LSH

Overview

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - Random points [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]
  - Classical root lattices A_d, D_d [JASG'08]
  - Exceptional root lattices $E_{6,7,8}$, F_4, G_2 [JASG'08]
  - Cross-polytopes [TT'07, AILRS'15, KW'17]
  - Hypercubes [TT'07]

Hyperplane LSH

Asymptotically "optimal"

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - Orthogonal hyperplanes [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - **Random points** [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]
  - **Classical root lattices** A_d, D_d [JASG'08]
  - Exceptional root lattices $E_{6,7,8}$, F_4, G_2 [JASG'08]
  - **Cross-polytopes** [TT'07, AILRS'15, KW'17]
  - Hypercubes [TT'07]

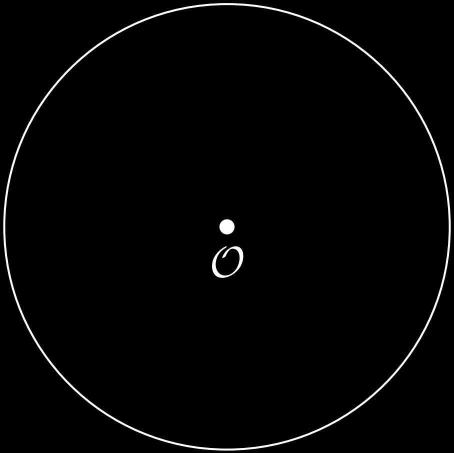
Hyperplane LSH

Topic of this paper

- Simple: one hyperplane corresponds to one inner product
- Easy to analyze: collision probability $1 - \frac{\theta}{\pi}$ for vectors at angle θ
- Can be made very efficient in practice
  - Sparse hyperplane vectors [Ach'01, LHC'06]
  - **Orthogonal hyperplanes** [TT'07]
- Theoretically suboptimal: use "nicer" (lattice-based) partitions
  - **Random points** [AI'06, AINR'14, ...]
  - Leech lattice [AI'06]
  - **Classical root lattices** A_d, D_d [JASG'08]
  - Exceptional root lattices $E_{6,7,8}$, F_4, G_2 [JASG'08]
  - **Cross-polytopes** [TT'07, AILRS'15, KW'17]
  - **Hypercubes** [TT'07]

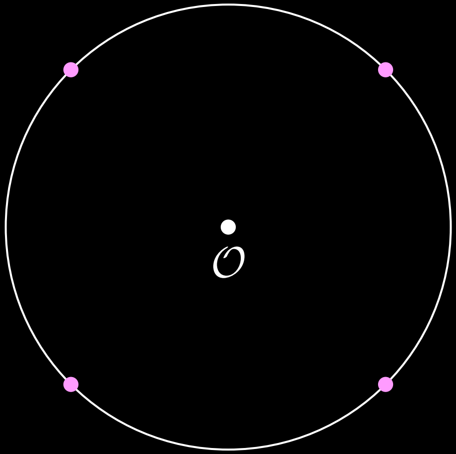
Hypercube LSH

[Terasawa–Tanaka, WADS'07]



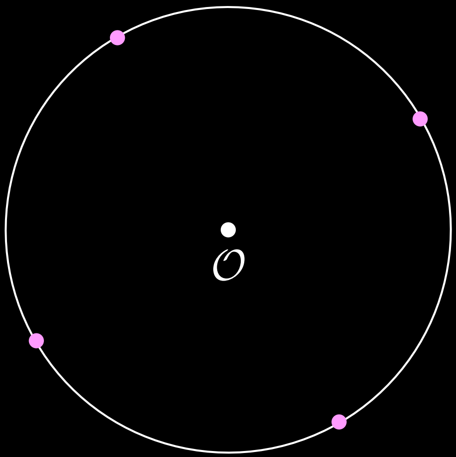
Hypercube LSH

Vertices of hypercube



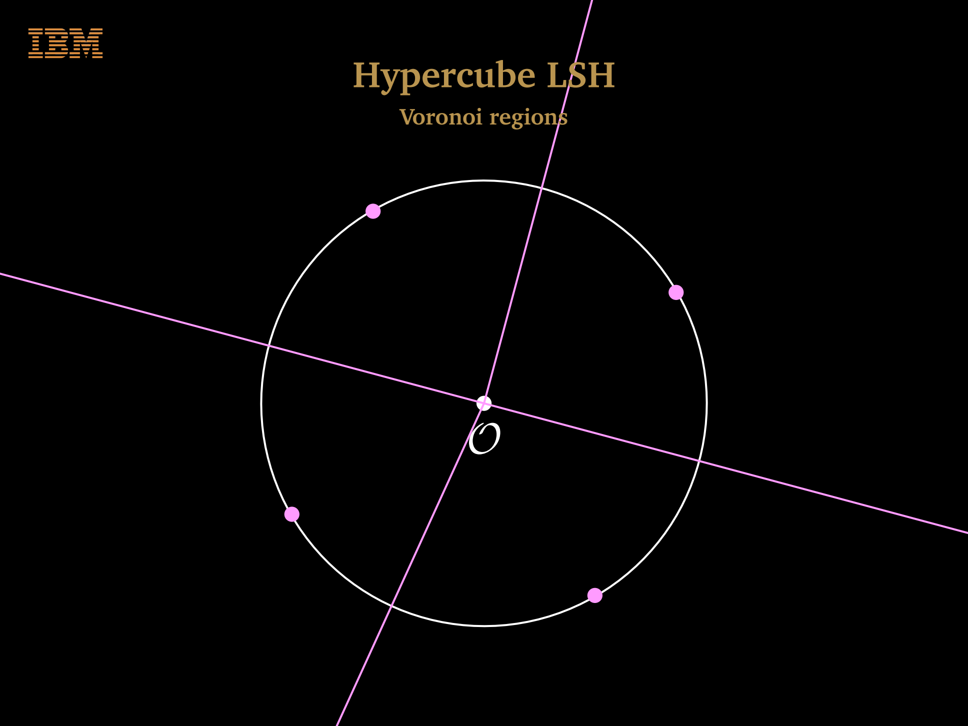
Hypercube LSH

Random rotation



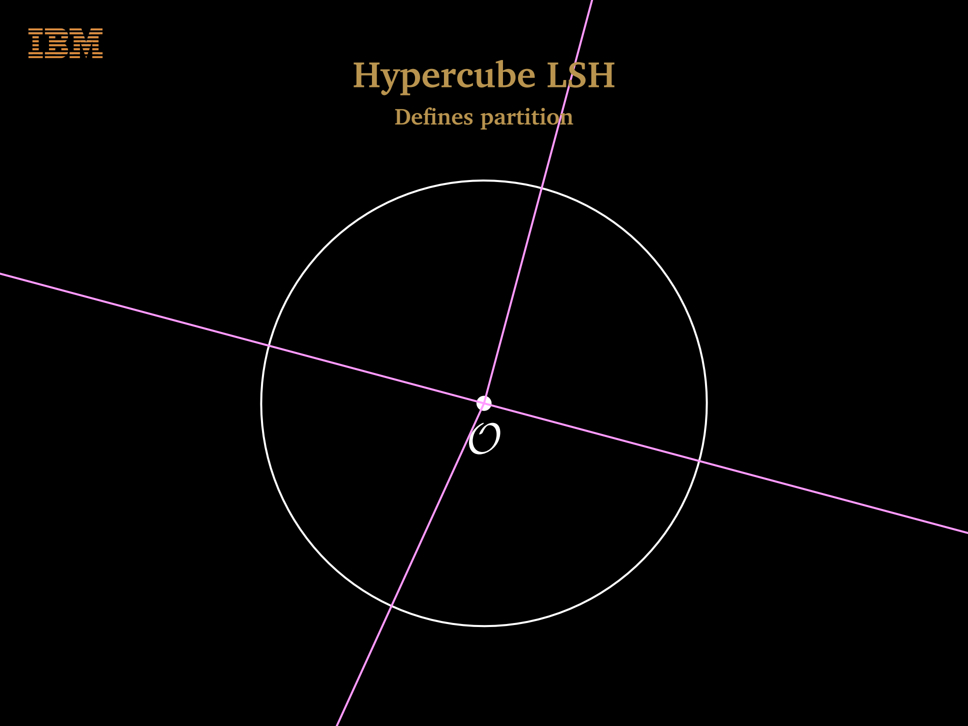
Hypercube LSH

Voronoi regions



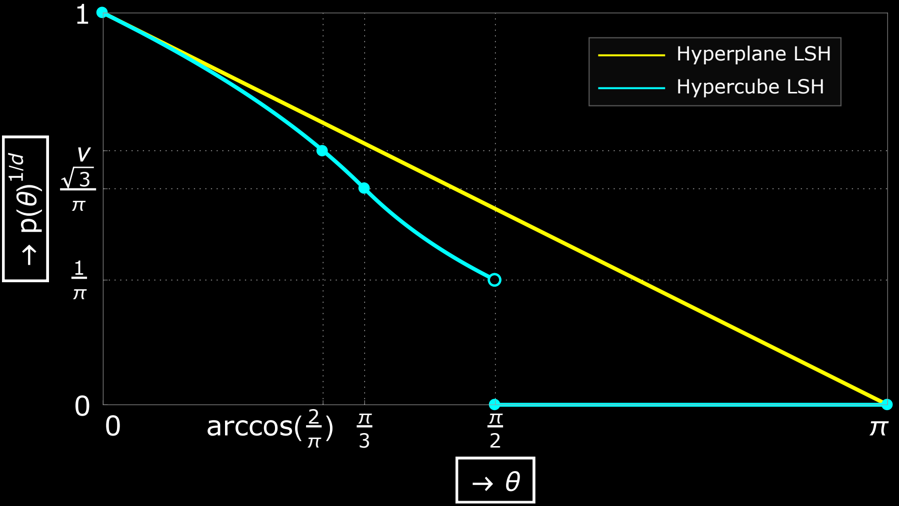
Hypercube LSH

Defines partition



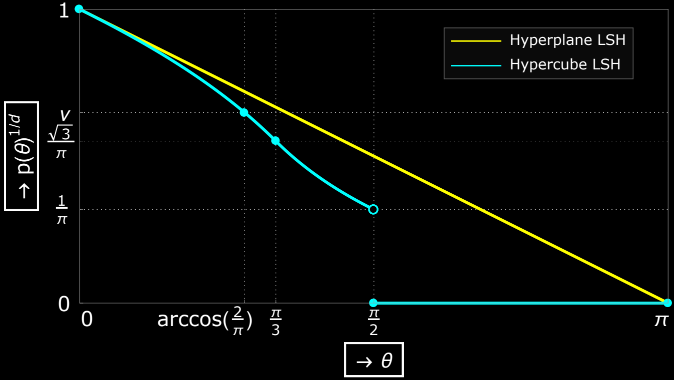
Hypercube LSH

Collision probabilities



Hypercube LSH

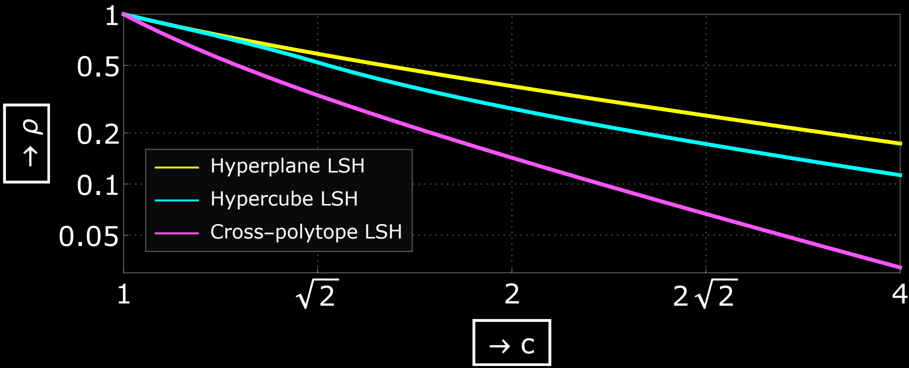
Collision probabilities



- Two vectors at angle $(\frac{\pi}{2})^-$ lie in the same orthant with probability $(\frac{1}{\pi})^d$
- Two vectors at angle $\frac{\pi}{3}$ lie in the same orthant with probability $(\frac{\sqrt{3}}{\pi})^d$

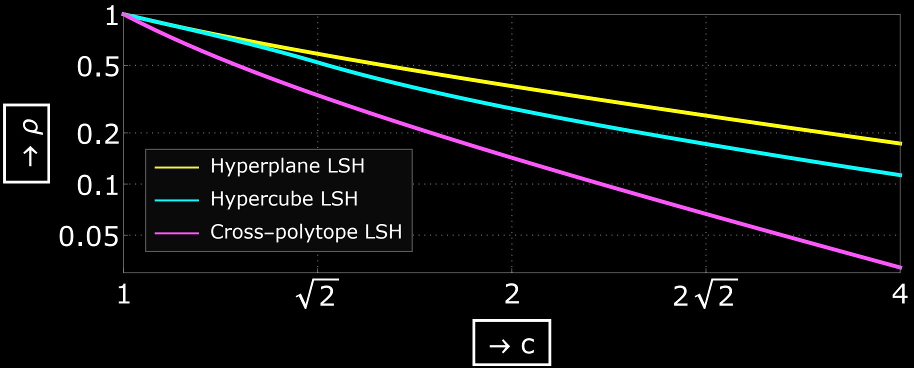
Hypercube LSH

Asymptotic performance (random data)



Hypercube LSH

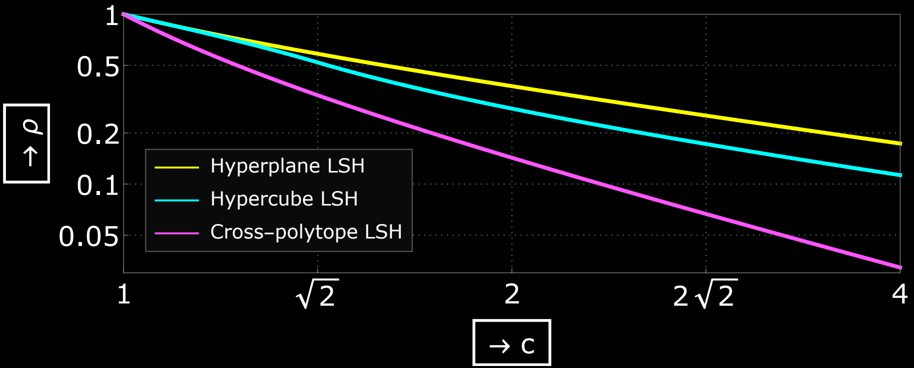
Asymptotic performance (random data)



- Hyperplane LSH: $\rho = \frac{\sqrt{2}}{\pi c \ln 2} + O(\frac{1}{c^2})$

Hypercube LSH

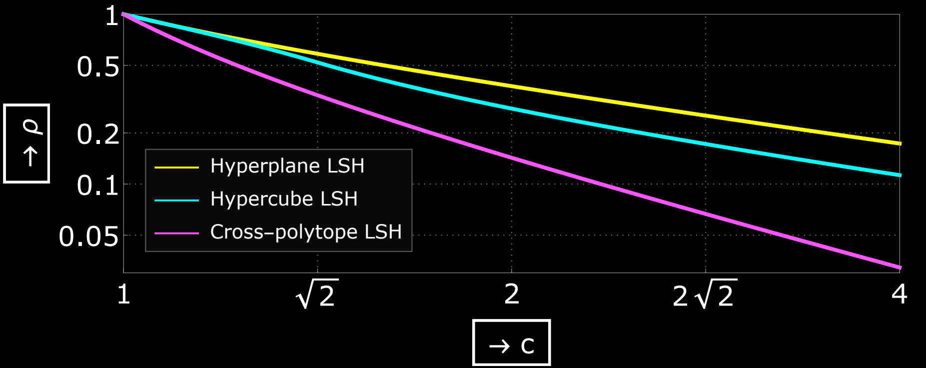
Asymptotic performance (random data)



- Hyperplane LSH: $\rho = \frac{\sqrt{2}}{\pi c \ln 2} + O(\frac{1}{c^2})$
- Hypercube LSH: $\rho = \frac{\sqrt{2}}{\pi c \ln \pi} + O(\frac{1}{c^2})$ – saves factor $\log_2(\pi) \approx 1.65$

Hypercube LSH

Asymptotic performance (random data)



- Hyperplane LSH: $\rho = \frac{\sqrt{2}}{\pi c \ln 2} + O(\frac{1}{c^2})$
- Hypercube LSH: $\rho = \frac{\sqrt{2}}{\pi c \ln \pi} + O(\frac{1}{c^2})$ – saves factor $\log_2(\pi) \approx 1.65$
- Cross-polytope LSH: $\rho = \frac{1}{2c^2 - 1} + o(\frac{1}{c^2})$

Conclusions

Positive results

- Exact asymptotics for full-dimensional hypercube LSH
- Exact asymptotics for partial hypercube LSH when $d' \leq O(d/\log d)$
- Asymptotically superior to hyperplane LSH
- Theoretical justification for using orthogonal hyperplanes

Conclusions

Positive results

- Exact asymptotics for full-dimensional hypercube LSH
- Exact asymptotics for partial hypercube LSH when $d' \leq O(d/\log d)$
- Asymptotically superior to hyperplane LSH
- Theoretical justification for using orthogonal hyperplanes

Negative results

- Asymptotically inferior to e.g. cross-polytope LSH
- Need large hypercubes to beat hyperplane LSH

Conclusions

Positive results

- Exact asymptotics for full-dimensional hypercube LSH
- Exact asymptotics for partial hypercube LSH when $d' \leq O(d/\log d)$
- Asymptotically superior to hyperplane LSH
- Theoretical justification for using orthogonal hyperplanes

Negative results

- Asymptotically inferior to e.g. cross-polytope LSH
- Need large hypercubes to beat hyperplane LSH

Open problems

- Exact asymptotics for all of partial hypercube LSH
- Other, better partition families?

Conclusions

Positive results

- Exact asymptotics for full-dimensional hypercube LSH
- Exact asymptotics for partial hypercube LSH when $d' \leq O(d/\log d)$
- Asymptotically superior to hyperplane LSH
- Theoretical justification for using orthogonal hyperplanes

Negative results

- Asymptotically inferior to e.g. cross-polytope LSH
- Need large hypercubes to beat hyperplane LSH

Open problems

- Exact asymptotics for all of partial hypercube LSH
- Other, better partition families?

Thank you for your attention!