Core-Sets for Polytope Distance

Mittagsseminar

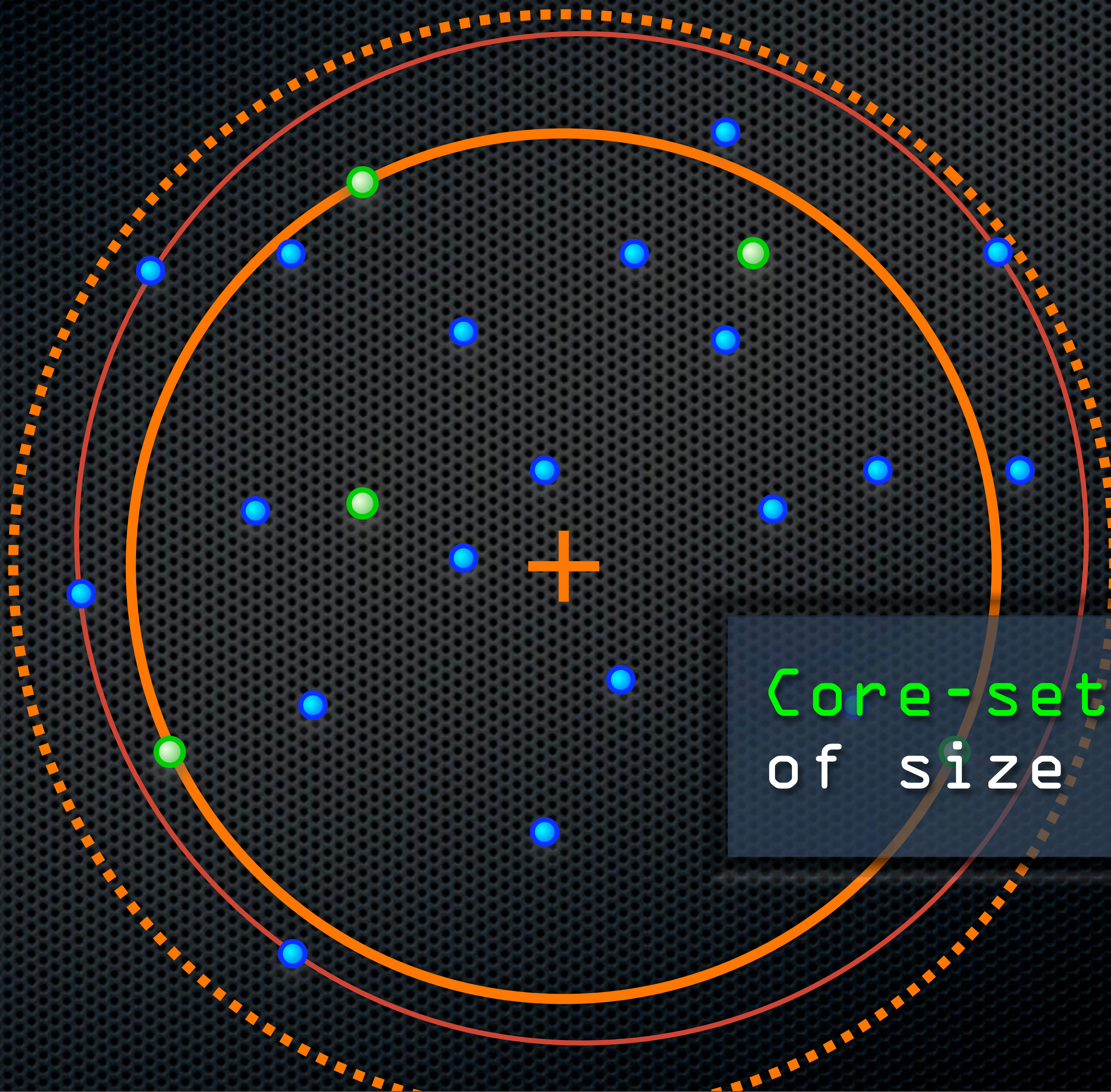
Martin Jaggi, 9.10.2008

Contents

- What are Core-Sets good for?
  - Smallest Enclosing Ball
  - Extent
  - Clustering
- Core-Sets for Polytope Distance
  - Lower bound
  - Upper bound: A new analysis for an old algorithm

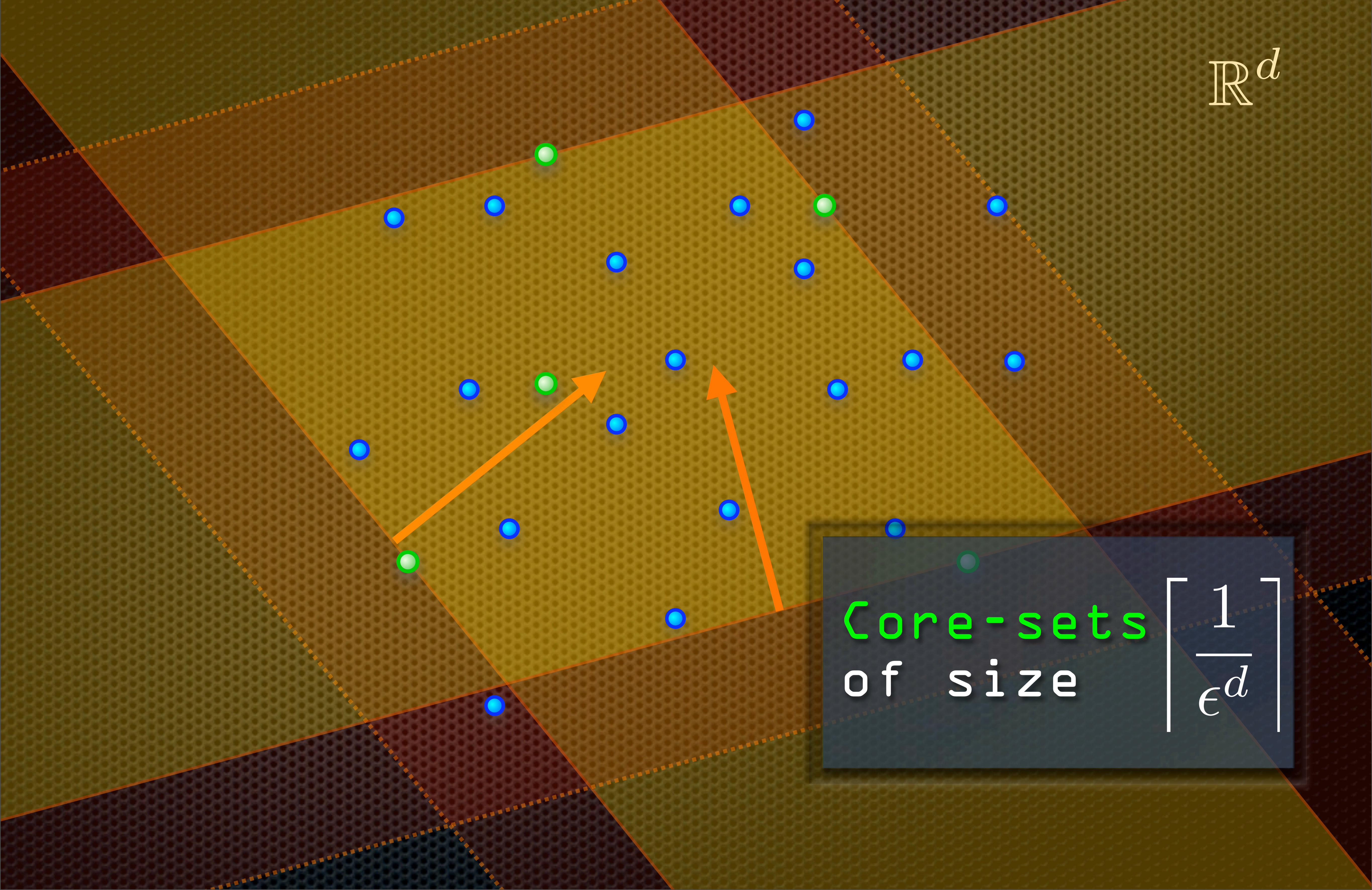
Smallest Enclosing Ball

$\mathbb{R}^d$

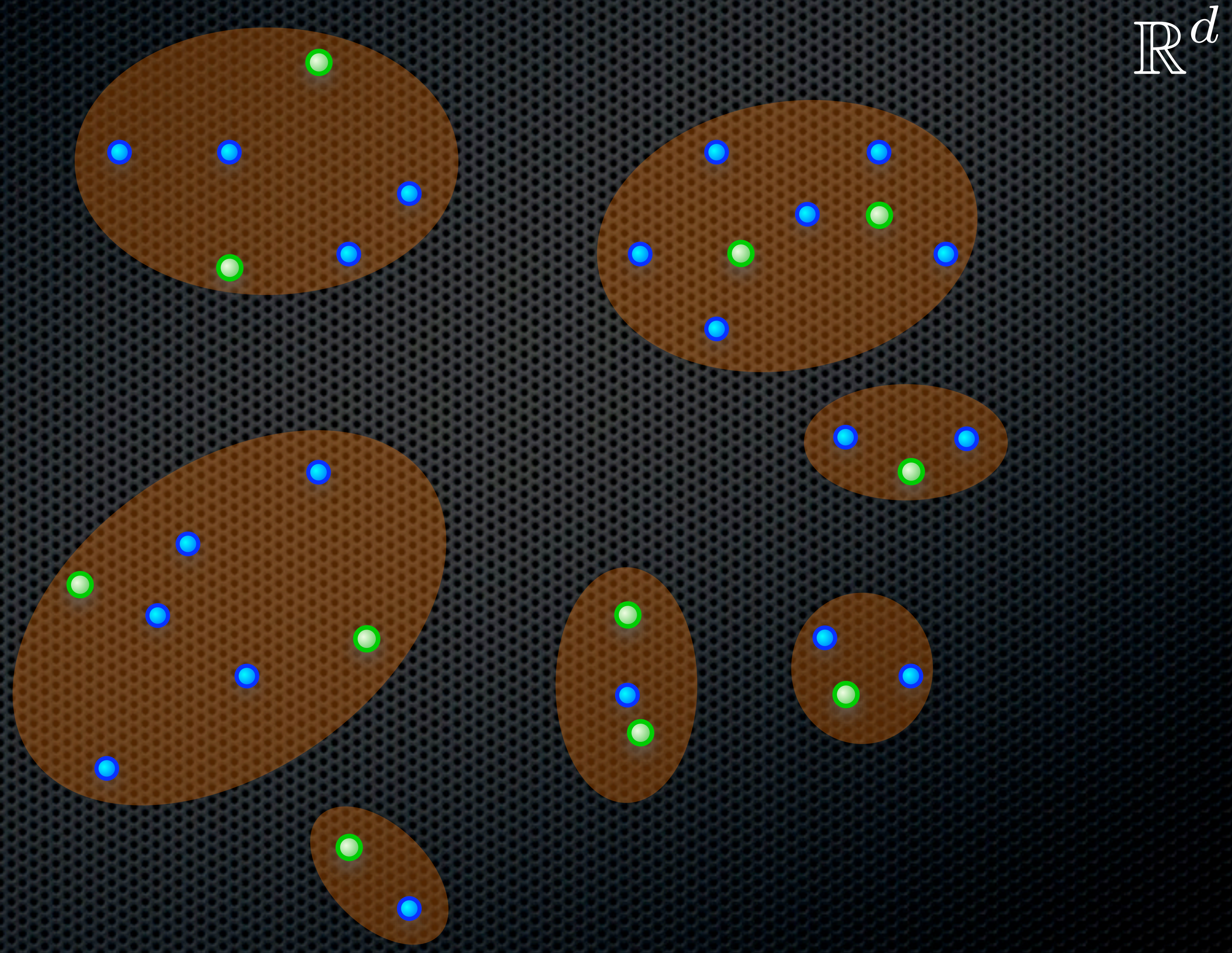


Core-sets of size $\left\lceil \frac{1}{\epsilon} \right\rceil$

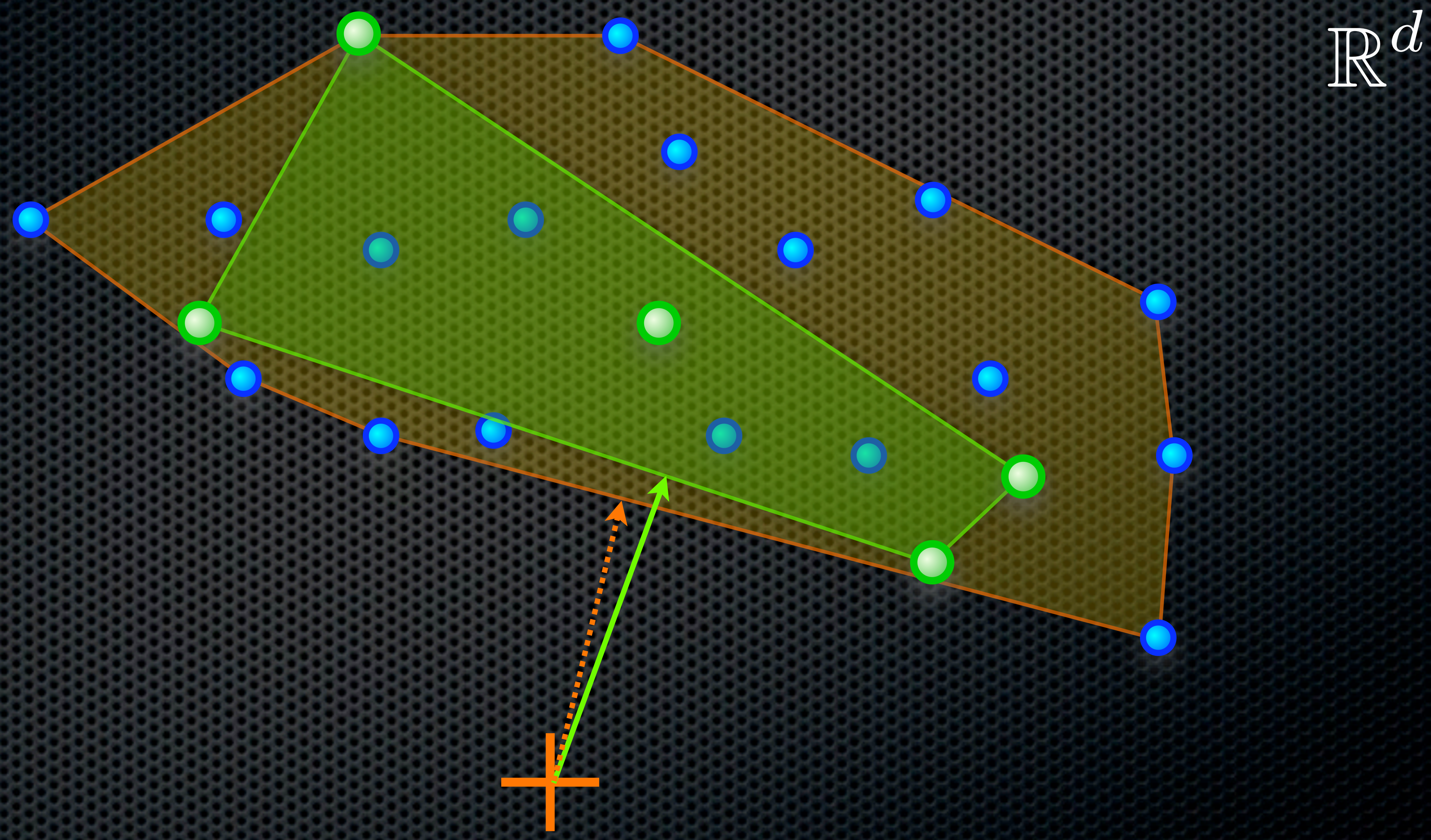
Core-Sets for Extent



Core-Sets for Clustering



Core-Sets for Polytope Distance

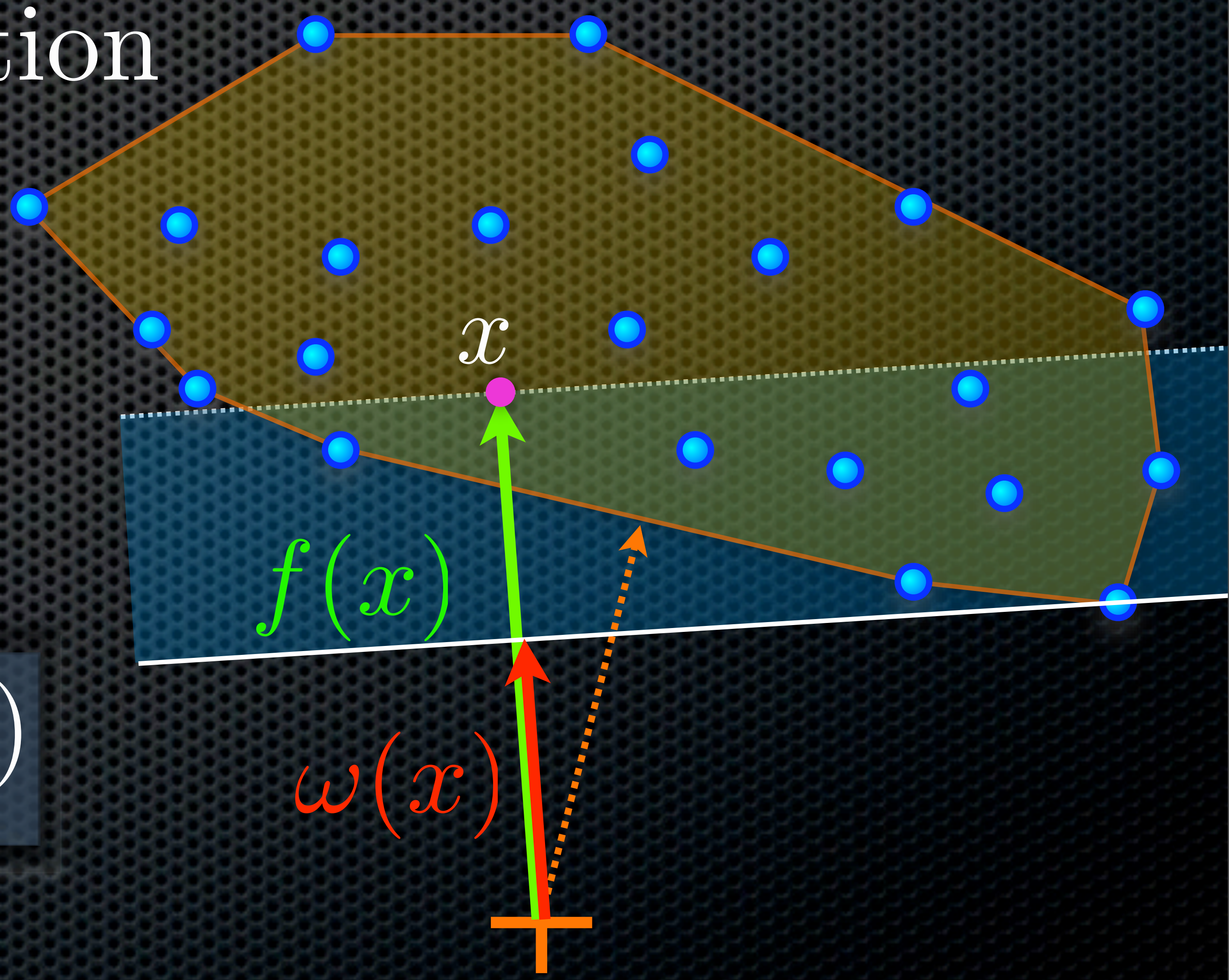


Definition: Approximation

A point x inside the polytope is called a

$0 < \epsilon \leq 1$

$$(1-\epsilon)\text{-approximation}$$



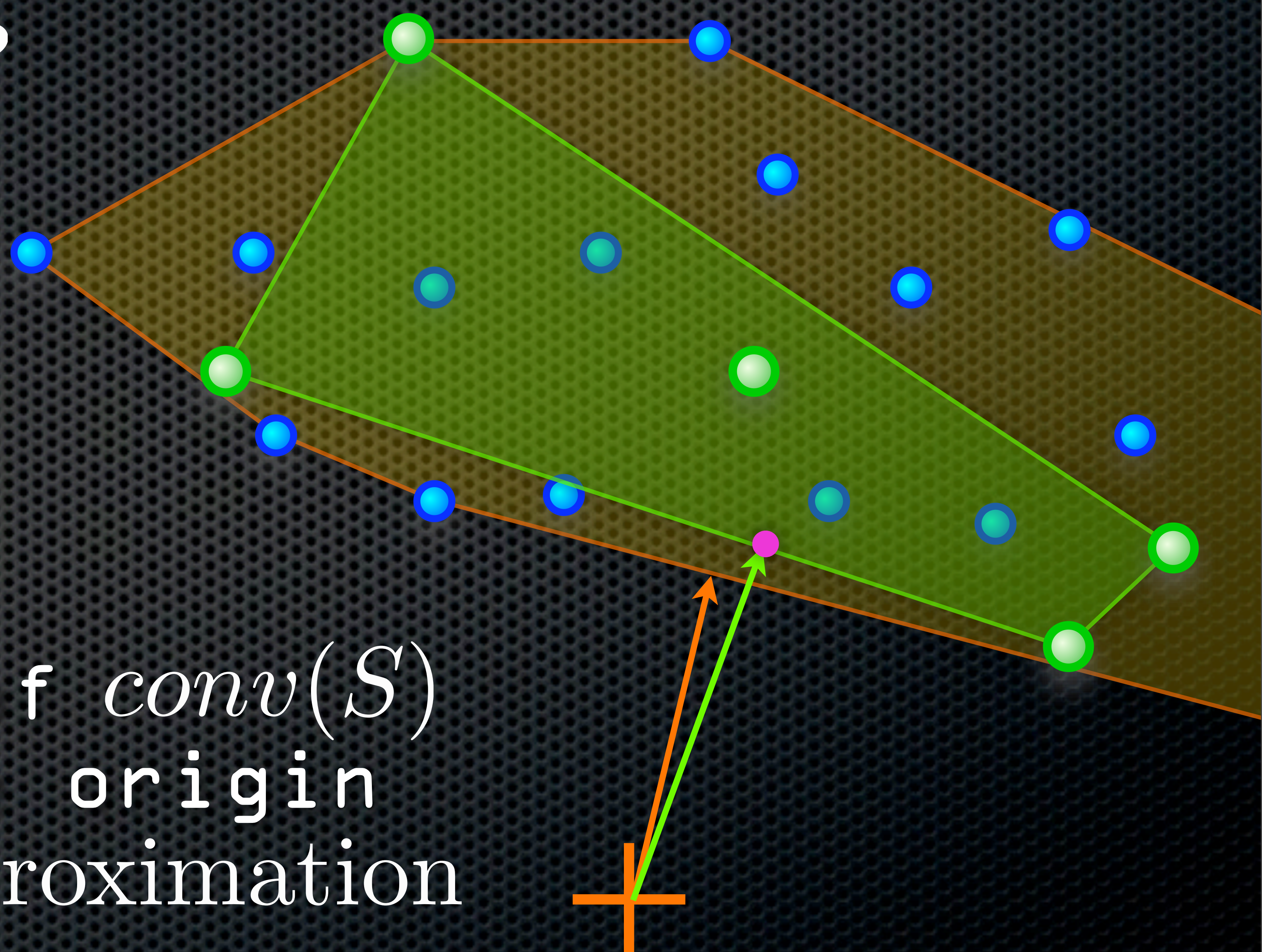
iff

$$f(x) - \omega(x) < \epsilon f(x)$$

Definition: Core-Set

A subset $S \subseteq P$ of the points is called an

ϵ-core-set



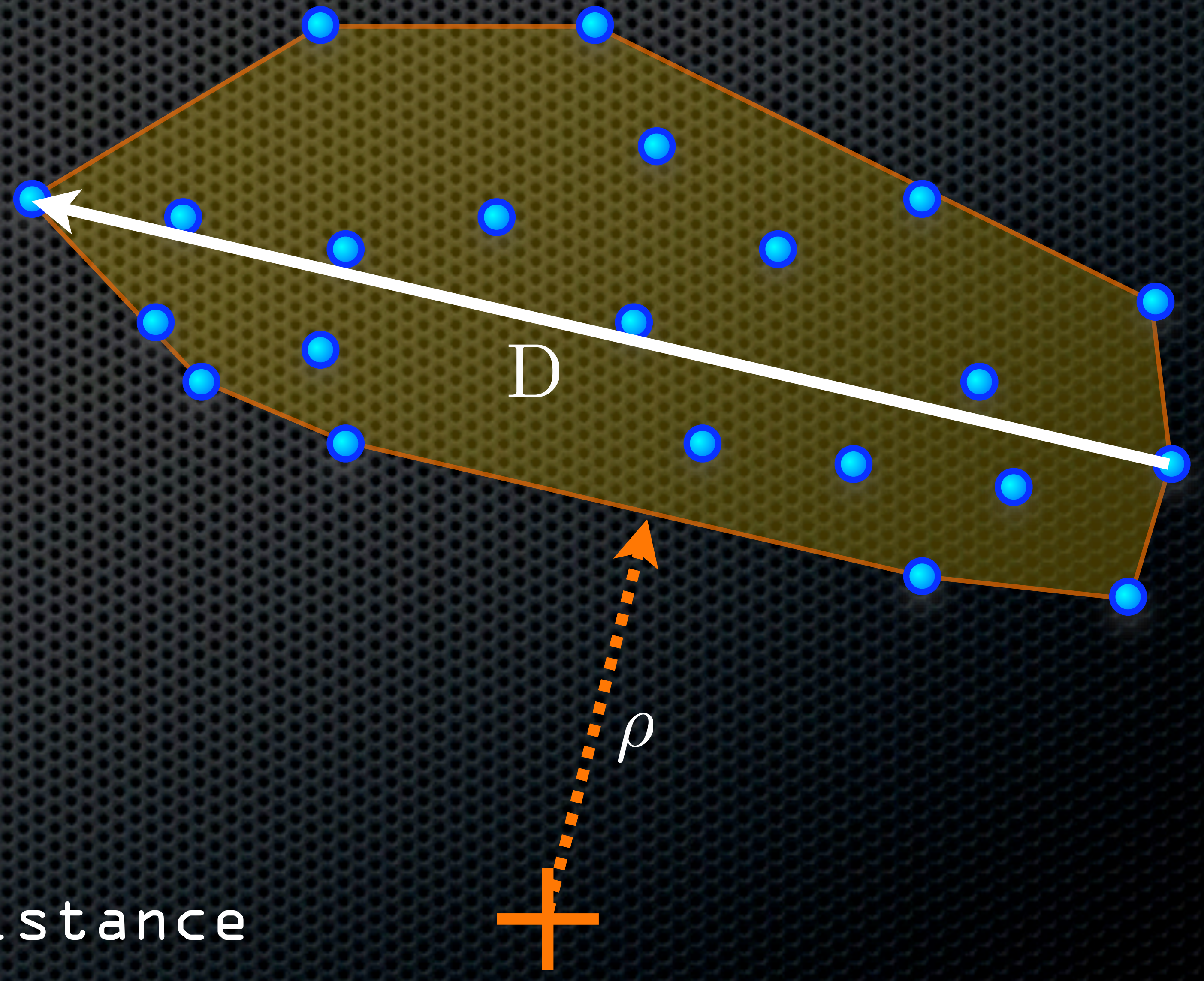
if the point of $conv(S)$ closest to the origin is a $(1-\epsilon)$-approximation

Definition:

Excentricity

The excentricity of a polytope is

$$E := \frac{1}{2}\frac{D^2}{\rho^2}$$



$$D := \max_{p,q \in P} ||p - q||$$

$\rho :=$ true polytope distance

Lower Bound

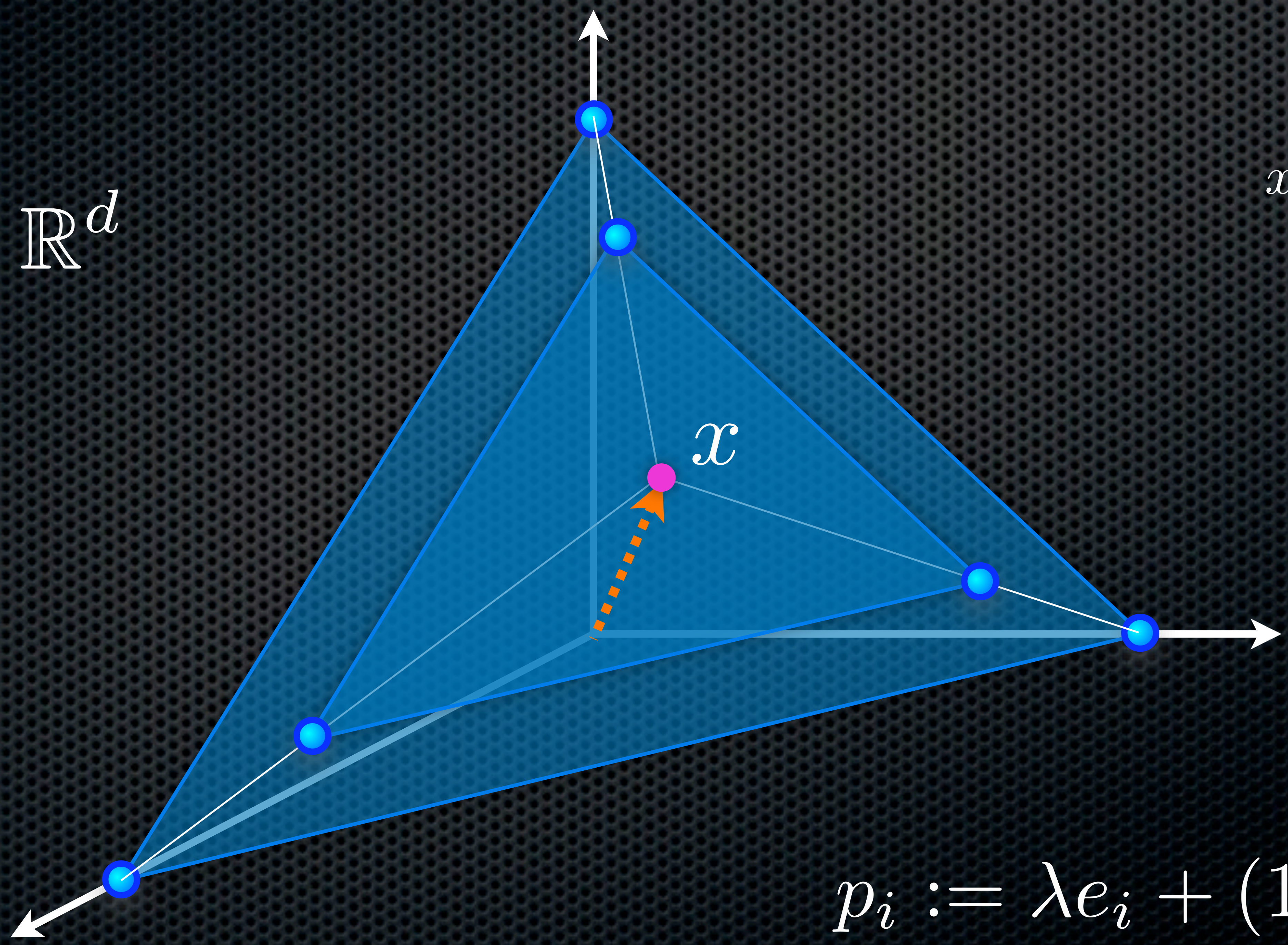
For any $0 < \epsilon \leq 1$ there exists a set of points in $\mathbb{R}^d$ such that any ϵ-core-set has size at least

$$\left\lceil \frac{E}{\epsilon} \right\rceil$$

$d \geq 2$

Lower Bound

Point set such that no strict subset can possibly be an ϵ-core-set ?

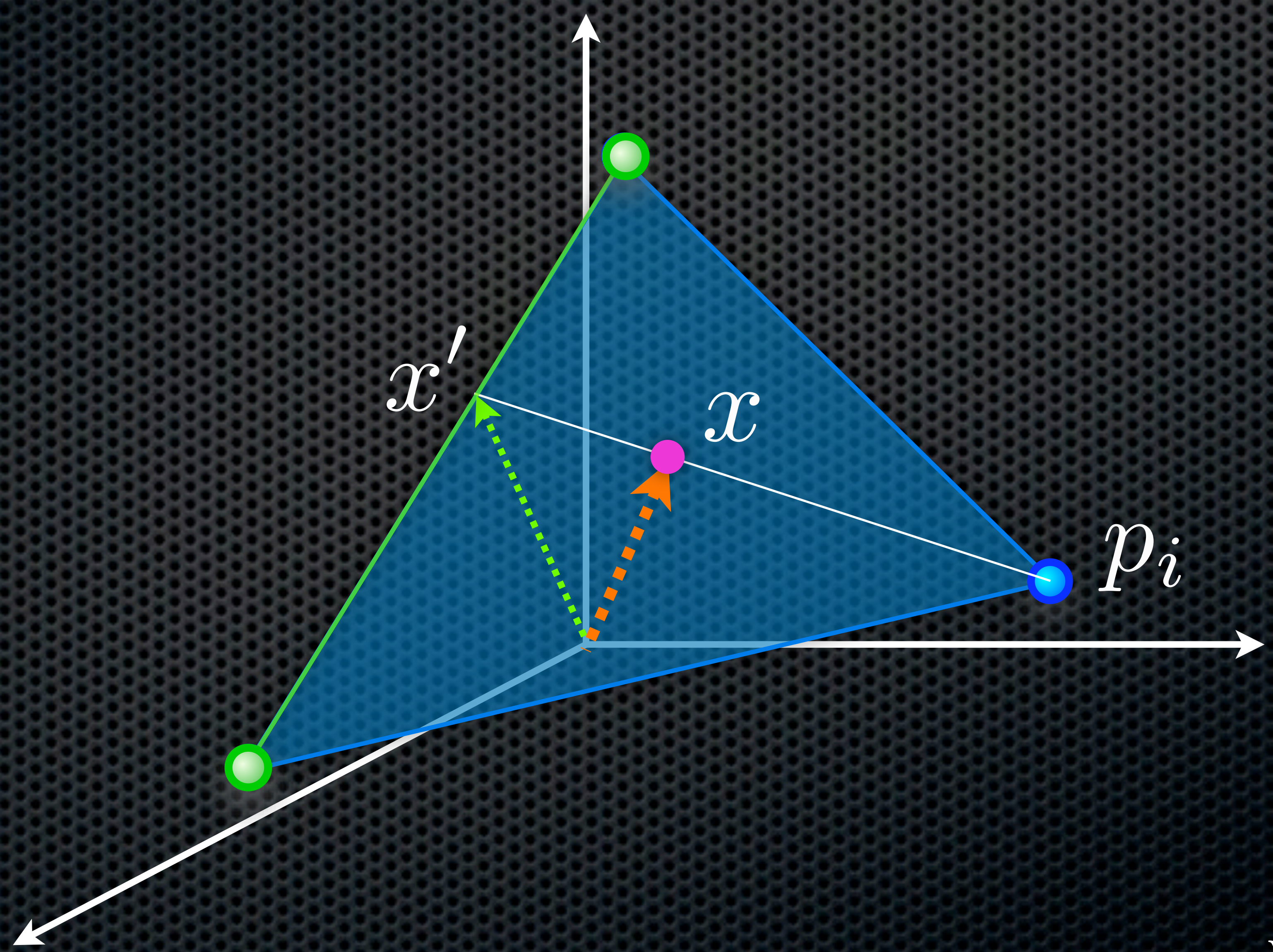


$$x = \begin{pmatrix} 1/d \\ \vdots \\ 1/d \end{pmatrix}$$

$$p_i := \lambda e_i + (1 - \lambda)x$$

Lower Bound

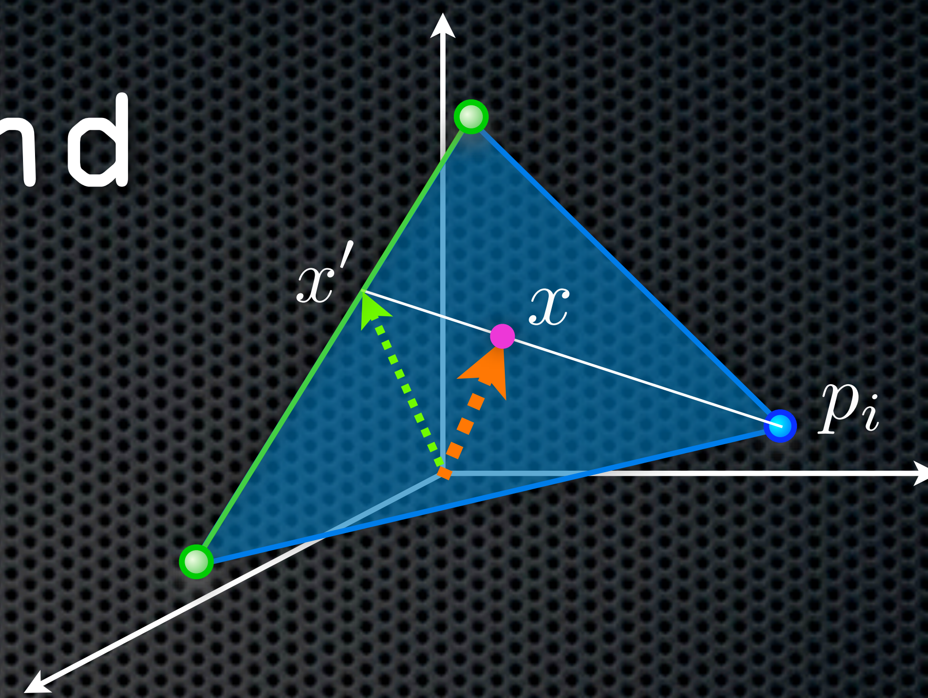
Point set such that no strict subset can possibly be an ϵ-core-set ?

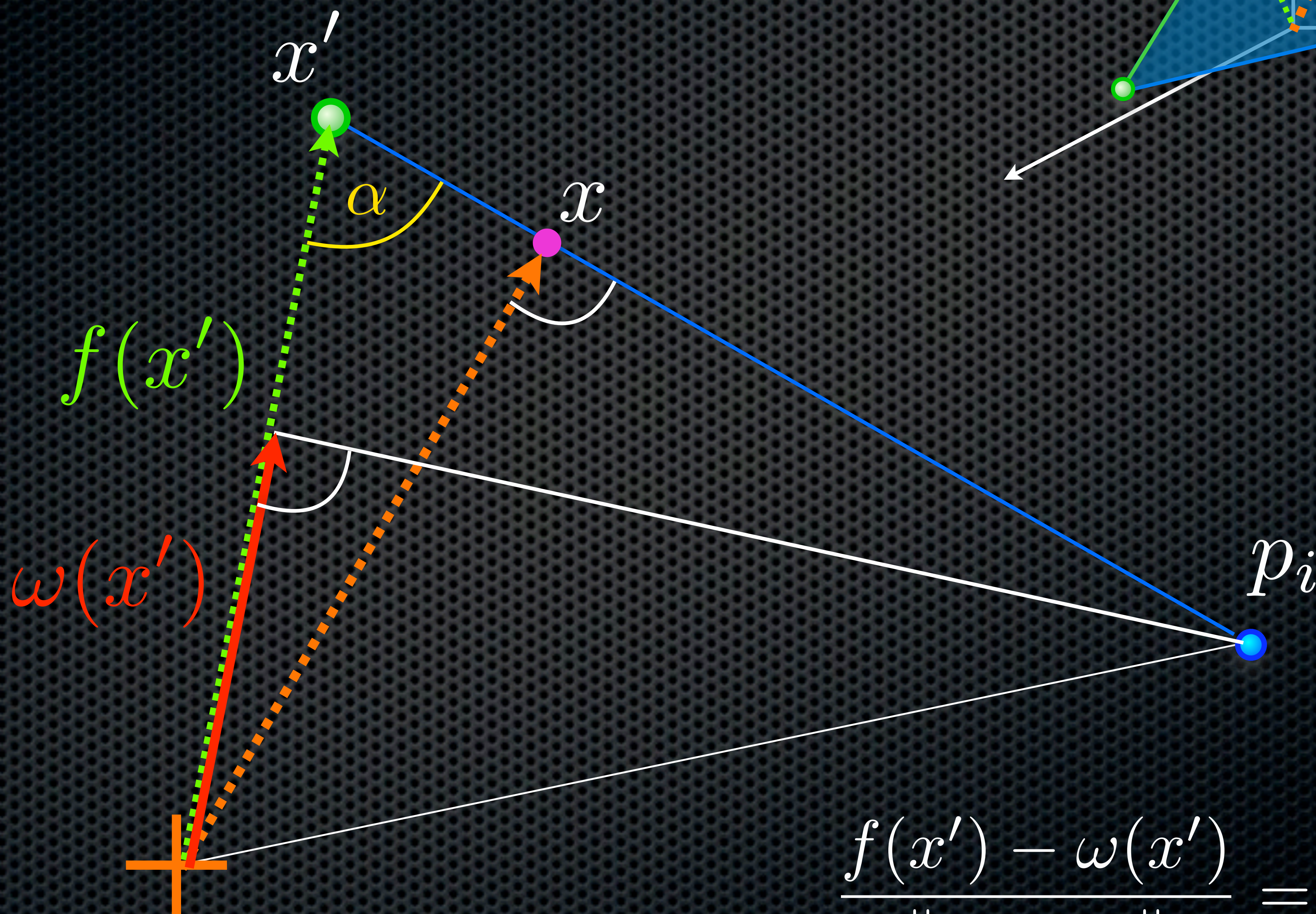


$$x = \begin{pmatrix} 1/d \\ \vdots \\ 1/d \end{pmatrix}$$

$$p_i := \lambda e_i + (1-\lambda)x$$

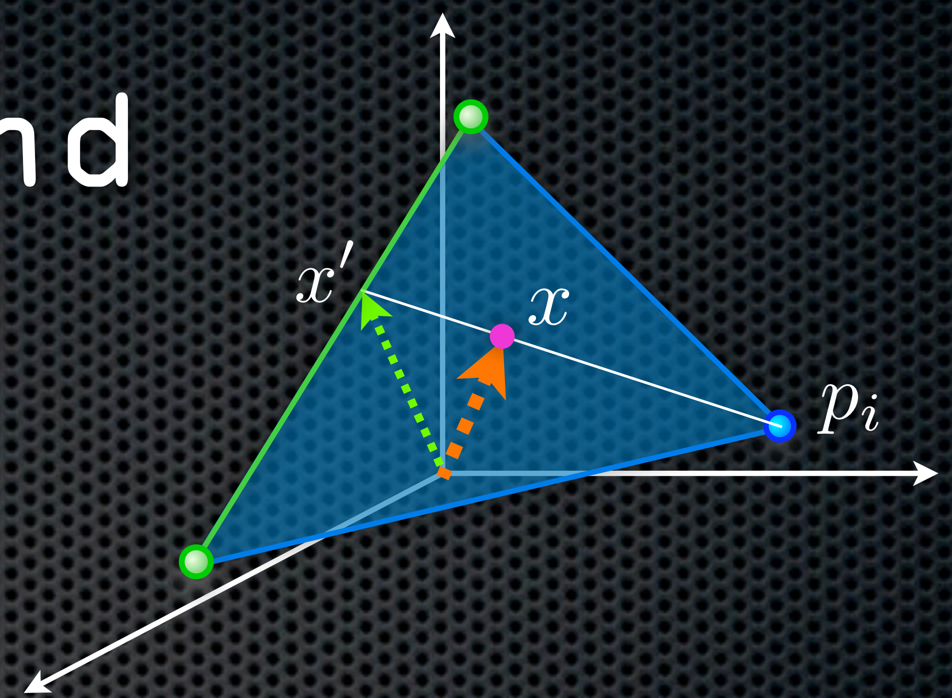
Lower Bound





$$\frac{f(x') - \omega(x')}{\|x' - p_i\|} = \frac{\|x' - x\|}{f(x')}$$

Lower Bound



$$\frac{f(x') - \omega(x')}{\|x' - p_i\|} = \frac{\|x' - x\|}{f(x')}$$

$$\frac{f(x') - \omega(x')}{f(x')} = \frac{\|x' - x\| \|x' - p_i\|}{f(x')^2} = \frac{d\lambda^2}{d - 1 + \lambda^2}$$

$$\stackrel{!}{=} \epsilon$$

$$< \epsilon$$

$$\|x'\|^2 = f(x')^2 = \frac{d - 1 + \lambda^2}{d(d-1)}$$

$$\|x' - x\|^2 = \frac{\lambda^2}{d(d-1)}$$

$$\|x' - p_i\|^2 = \frac{d\lambda^2}{d-1}$$

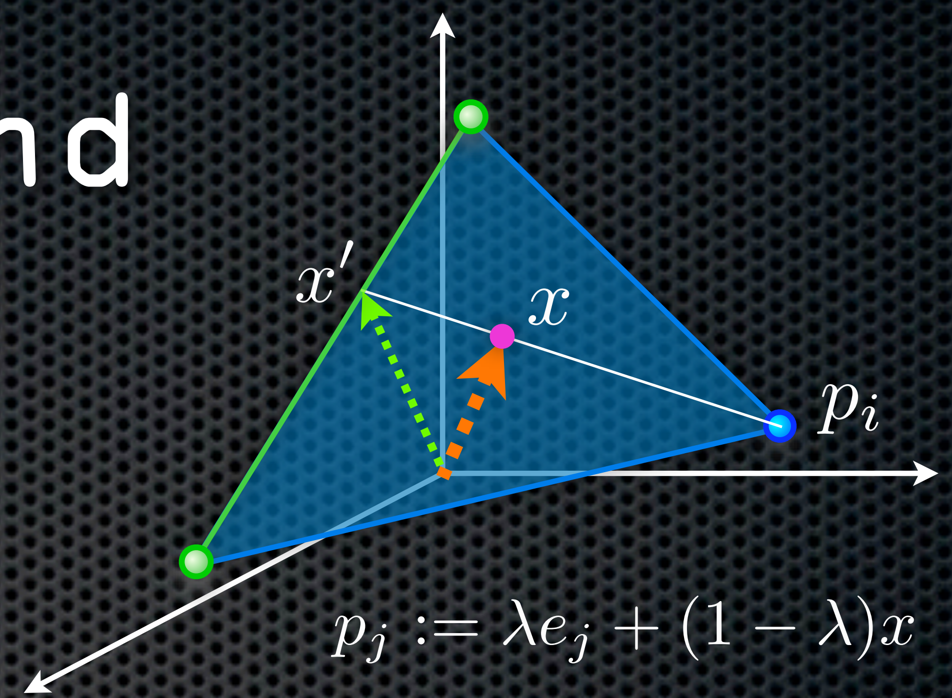
$$\lambda := \sqrt{\frac{(d-1)\epsilon}{d - \epsilon}}$$

$$p_j := \lambda e_j + (1 - \lambda)x$$

ϵ-core-set

$$f(x) - \omega(x) < \epsilon f(x)$$

Lower Bound



If we choose $\lambda := \sqrt{\frac{(d-1)\epsilon}{d-\epsilon}}$

Every ϵ-core-set must have size at least d.

$$E = d\lambda^2$$

$$d = d\lambda^2 \frac{d-\epsilon}{(d-1)\epsilon} = \frac{E}{\epsilon}\left(1 + \frac{1-\epsilon}{d-1}\right)$$

$$E := \frac{1}{2}\frac{D^2}{\rho^2}$$

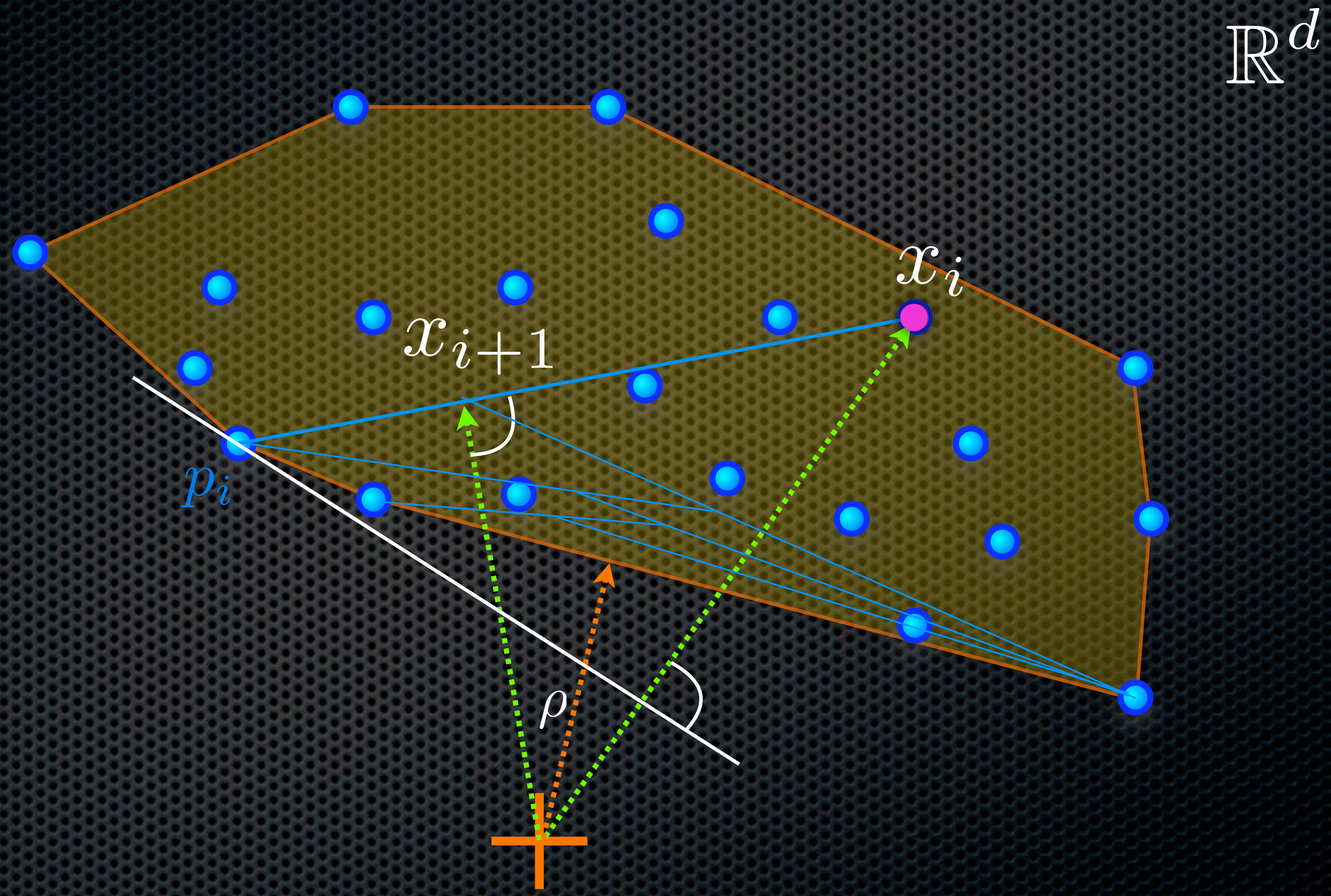
Every ϵ-core-set must have size at least $\left\lceil \frac{E}{\epsilon} \right\rceil$.

Upper Bound

How to construct small core-sets?

- Gilbert's Algorithm `66
  - Frank / Wolfe Approximation Algorithm `56
  - Coordinate-wise Gradient Descent
  - Sparse Greedy Approximation

Gilbert's Algorithm



Theorem

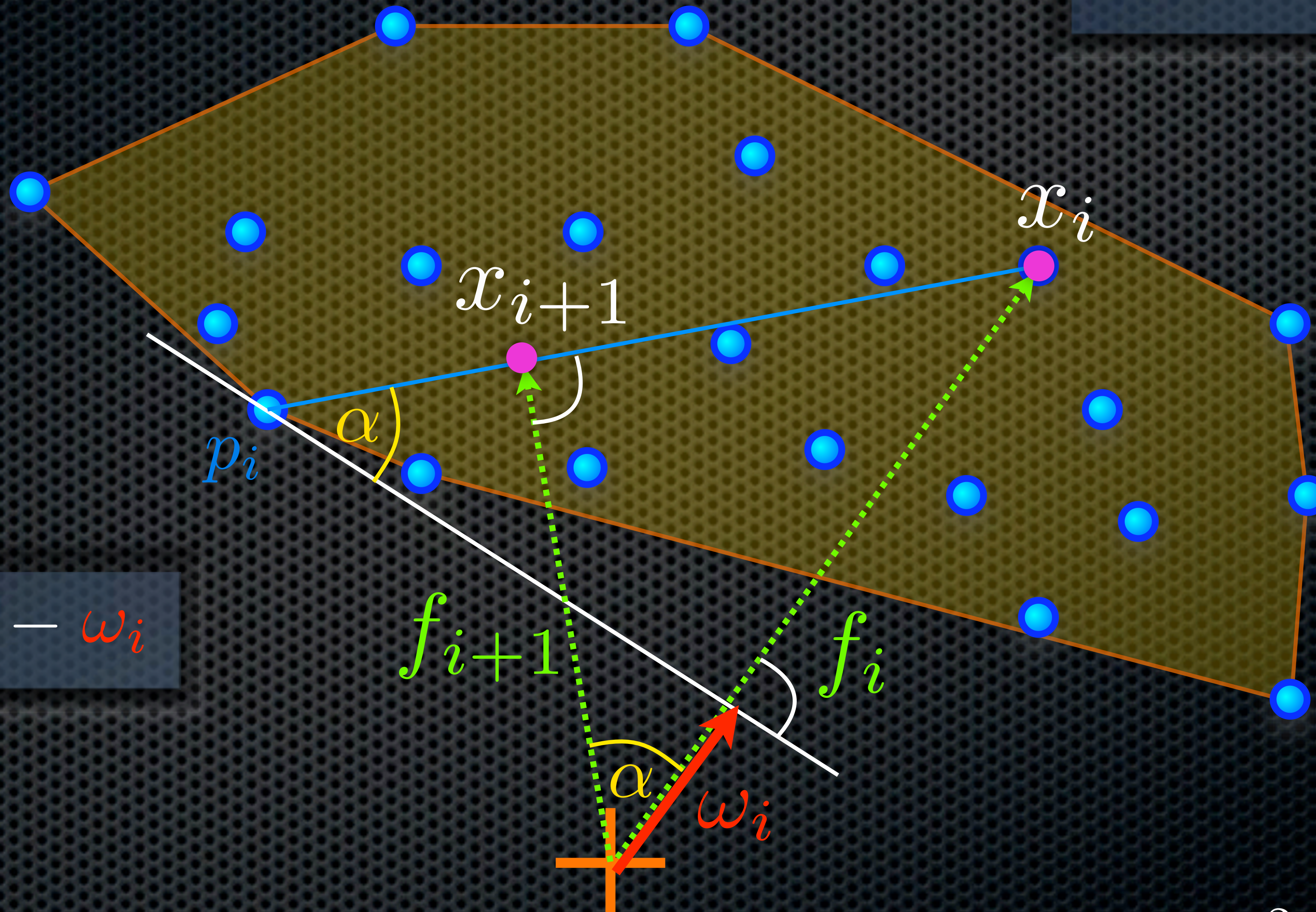
For any polytope, Gilbert's Algorithm returns a $(1-\epsilon)$-approximation

after at most $\left\lceil \frac{8E}{\epsilon} \right\rceil$ many steps.

Analysis

$$f_{i+1} = f_i \cos\alpha$$

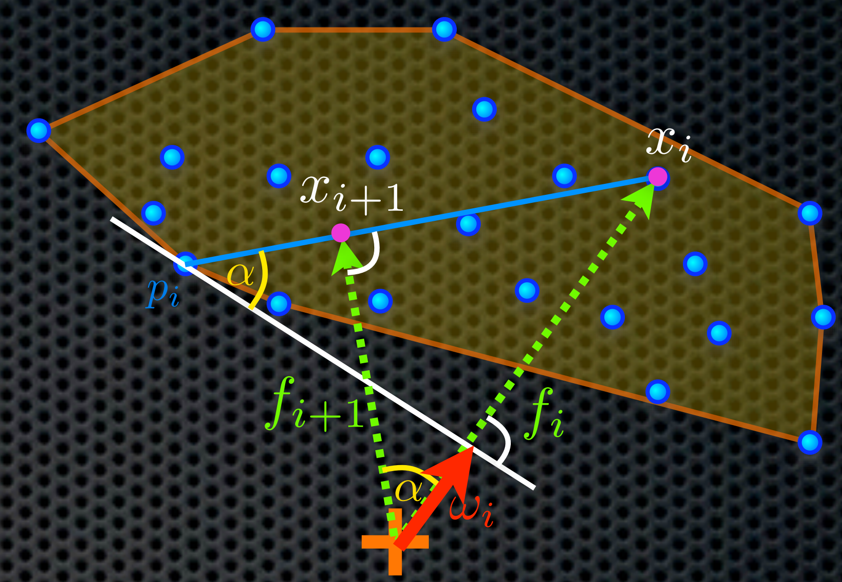
$$1 - \cos\alpha \geq \frac{1}{2}\sin^2\alpha$$
$$\Leftrightarrow (1-\cos\alpha)^2 \geq 0$$



$$g_i := f_i - \omega_i$$

$$f_i - f_{i+1} = (1-\cos\alpha) f_i \geq \frac{1}{2}\sin^2\alpha f_i = \frac{g_i^2}{2\|p_i - x_i\|^2} f_i$$

Analysis



$$f_i - f_{i+1} \geq \frac{g_i^2}{2\|p_i - x_i\|^2} f_i$$

//

$$h_i - h_{i+1}$$

$$\geq \frac{g_i^2}{2D^2}\rho$$

$$= \frac{1}{4E\rho} g_i^2$$

$$\geq \frac{1}{4E\rho} h_i^2$$

$$E := \frac{1}{2}\frac{D^2}{\rho^2}$$

$$g_i := f_i - \omega_i$$

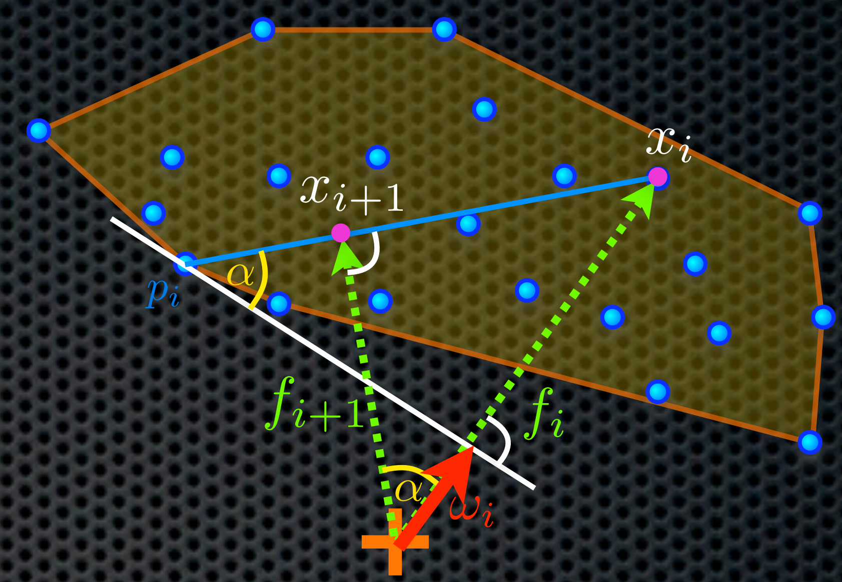
$$h_i := f_i - \rho$$

$$g_i \geq h_i$$

$$h_i' - h_{i+1}' \geq {g_i'}^2 \geq {h_i'}^2$$

$$h_i' := \frac{1}{4E\rho} h_i \qquad g_i' := \frac{1}{4E\rho} g_i$$

Analysis



$$h'_i - h'_{i+1} \geq {g'_i}^2 \geq {h'_i}^2$$

$$h'_{i+1} \leq h'_i(1 - h'_i)$$

$$\leq \frac{h'_i}{1 + h'_i} = \frac{1}{1 + \frac{1}{h'_i}}$$

$$1 - \gamma \leq \frac{1}{1 + \gamma}$$

$$g_i := f_i - \omega_i$$

$$h_i := f_i - \rho$$

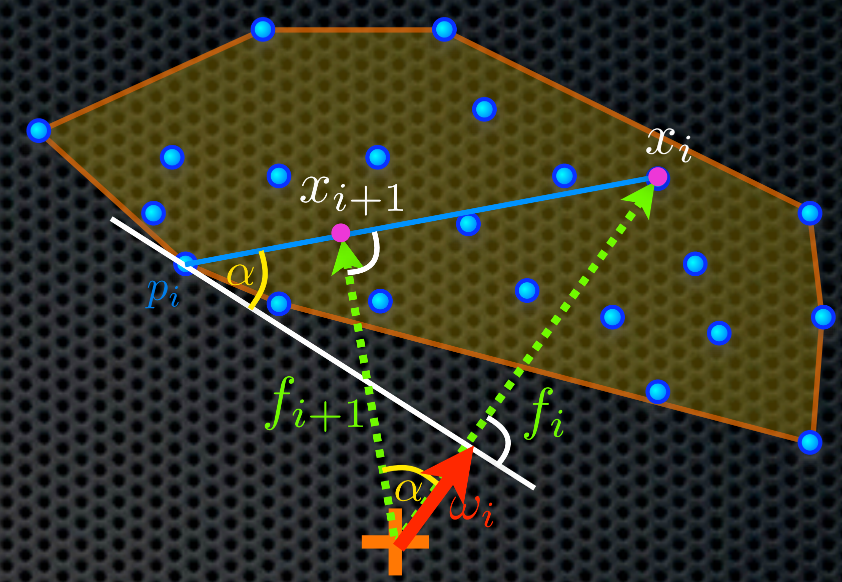
By induction we get:

$$h'_k \leq \frac{1}{k+2} \qquad h'_0 \leq \frac{1}{2}$$

$$g_i \geq h_i$$

$h'_k < \epsilon'$ for $k \geq K := \left\lceil \frac{1}{\epsilon'} \right\rceil$

Analysis



$h'_k < \epsilon'$ for $\quad k \geq K := \left\lceil \frac{1}{\epsilon'} \right\rceil$

$g'_k < \epsilon'$ for some $\quad K \leq k \leq 2K$

$$h'_i - h'_{i+1} \geq {g'_i}^2 \geq {h'_i}^2$$

$g_k < \epsilon\rho$ for some $\quad k \leq \left\lceil \frac{8E}{\epsilon} \right\rceil$

$$g_i := f_i - \omega_i$$

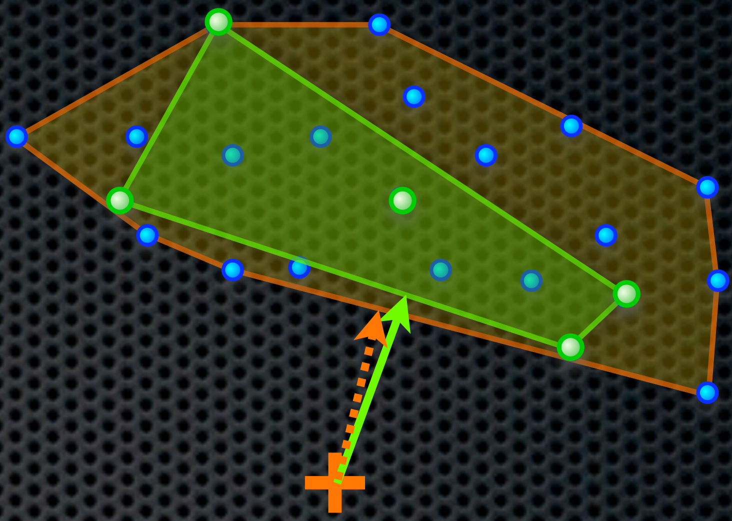
$$h_i := f_i - \rho$$

$$g_k = f_k - \omega_k < \epsilon\rho \leq \epsilon f_k$$

$$g_i \geq h_i$$

So we obtain a $(1-\epsilon)$-approximation after at most $\left\lceil \frac{8E}{\epsilon} \right\rceil$ many steps.

Conclusion



Lower bound

Every ϵ-core-set must have size at least $\left\lceil \frac{E}{\epsilon} \right\rceil$.

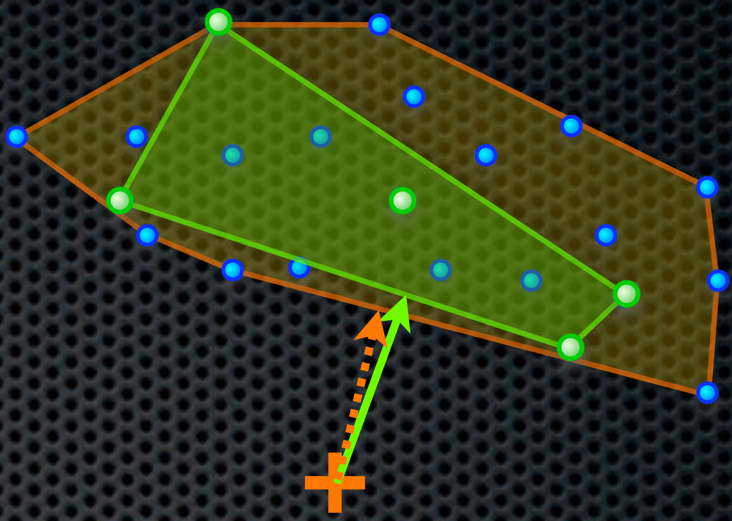
Upper bound

For any polytope, an ϵ-core-set of size $\left\lceil \frac{8E}{\epsilon} \right\rceil$ exists by the analysis of Gilbert's Algorithm.

Open Question:

Why is there still a gap of 2 ?

Conclusion



Lower bound

Every ϵ-core-set must have size at least $\left\lceil \frac{E}{\epsilon} \right\rceil$.

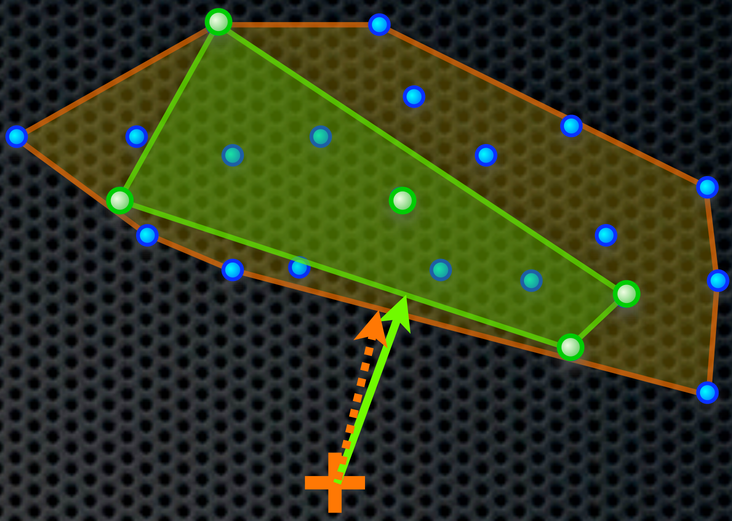
Upper bound

For any polytope, an ϵ-core-set of size $\left\lceil \frac{2E}{\epsilon} \right\rceil$ exists by the analysis of Clarkson's Away Steps Algorithm.

Open Question:

Why is there still a gap of 2 ?

Applications



Running time

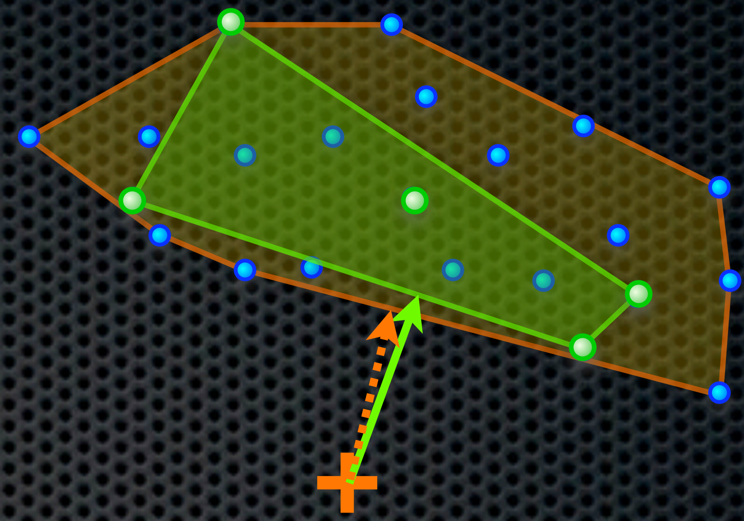
Gilbert's Algorithm runs in time $O(nd)$ for ϵ constant.

Distance between 2 polytopes

Run Gilbert's Algorithm on the Minkowski Difference.

($\Theta(n^2)$ vertices, but still $O(nd)$ running time)

Applications



Core-sets for Support Vector Machines

We can guarantee zero errors on the training data after only $\lceil 8E \rceil$ steps.

$\epsilon := 1$

Sparseness of approximate solutions

Using the upper and lower bounds for the sparseness of approximate solutions, we know the effect of different

- geometric kernels
- regularization parameter values

on this sparsity.

Thanks for your attention!