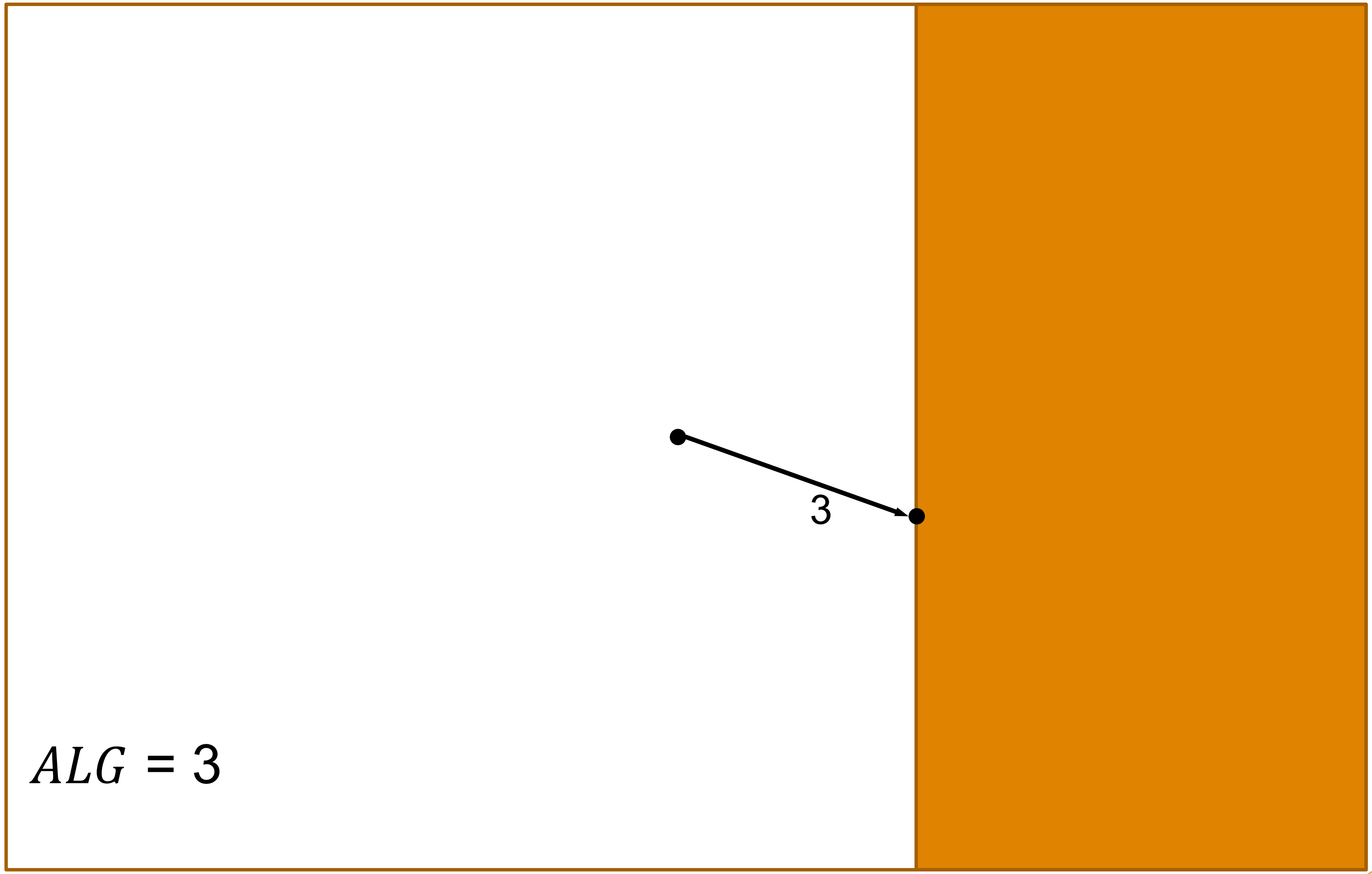
Linear-Competitive Convex Body Chasing

C.J. Argue[1] / Mark Sellke

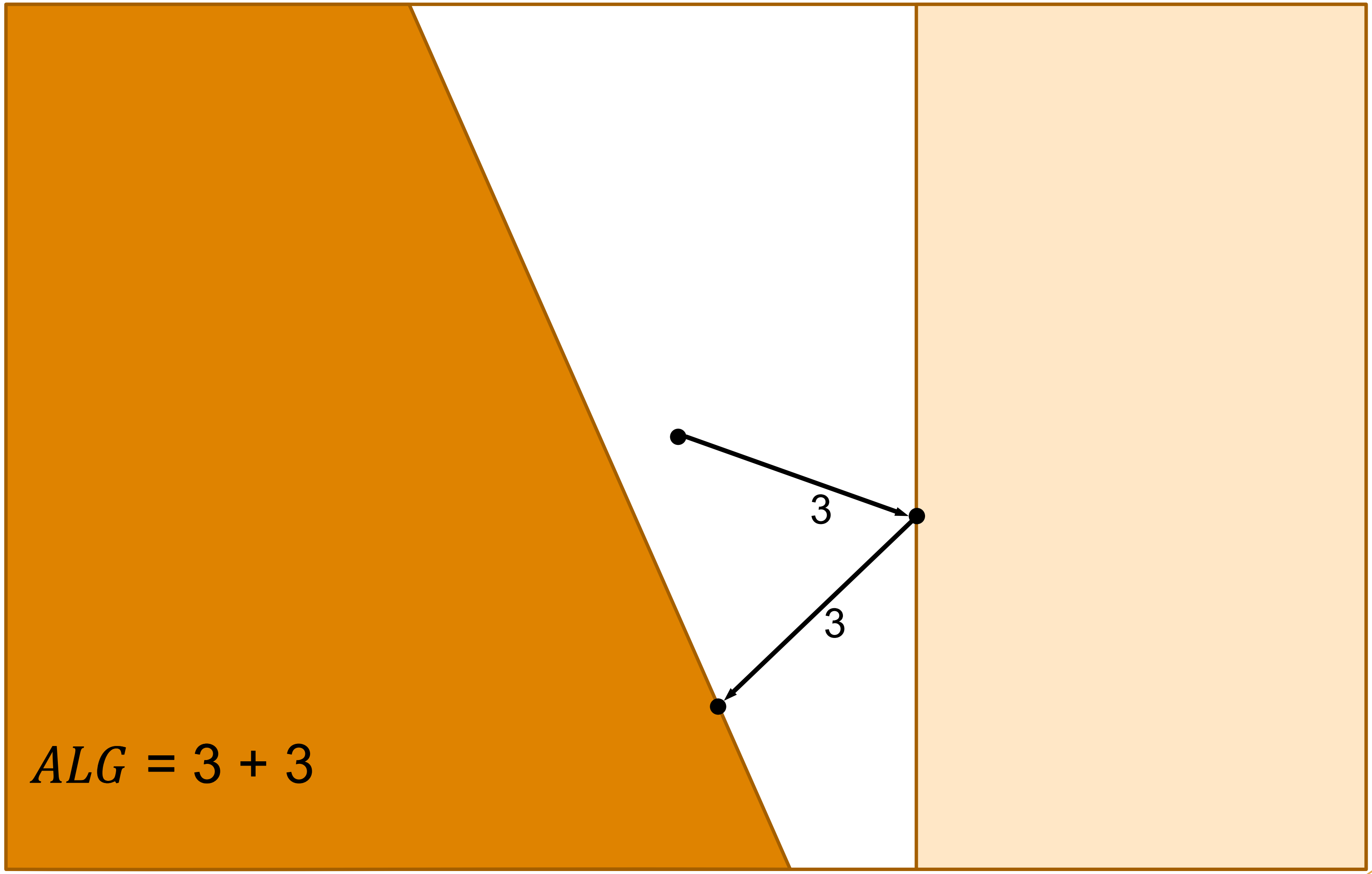
[1]Joint with Anupam Gupta, Guru Guruganesh, Ziye Tang

some slides courtesy of Guru Guruganesh

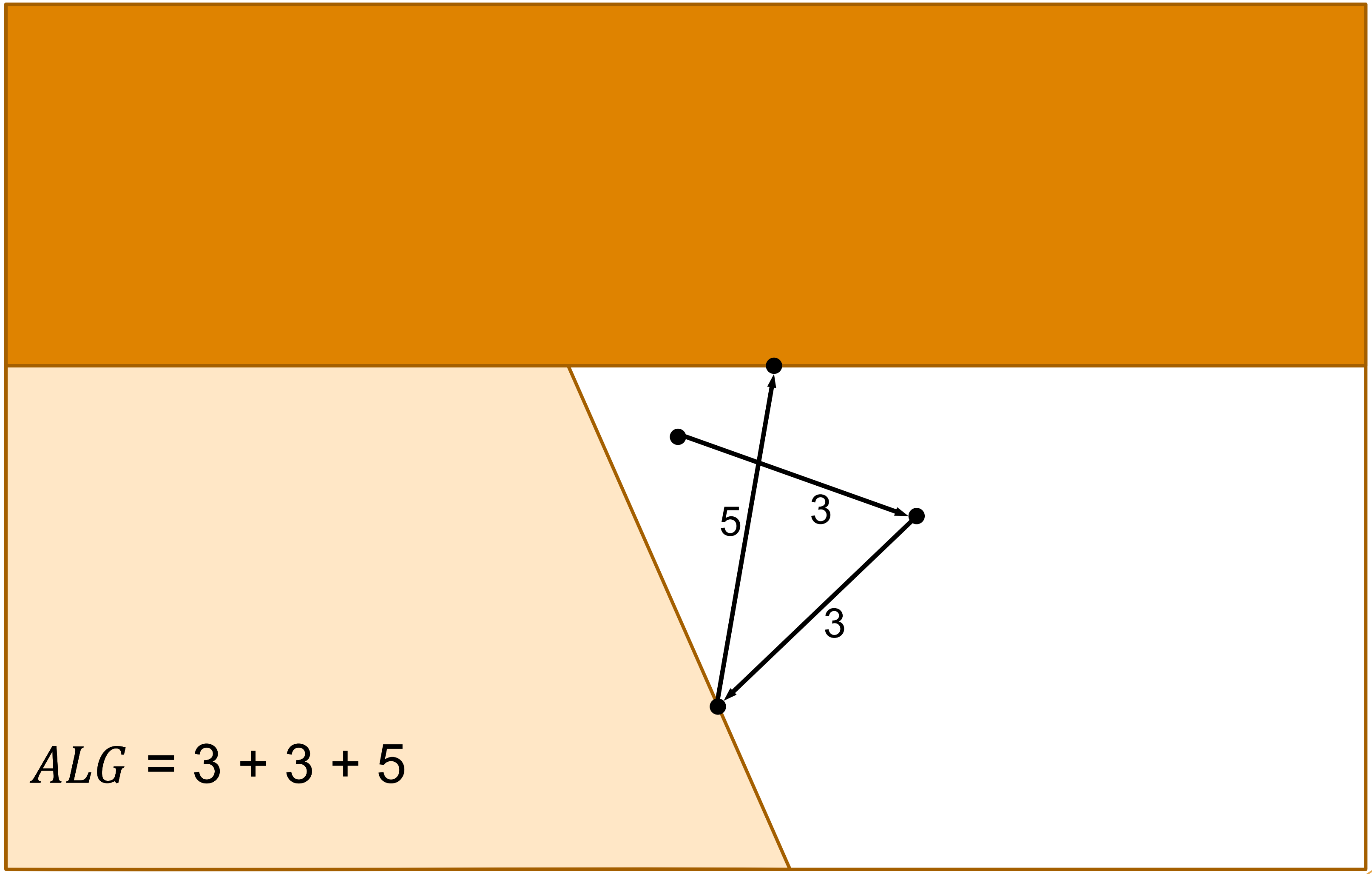
Convex Body Chasing – The Problem



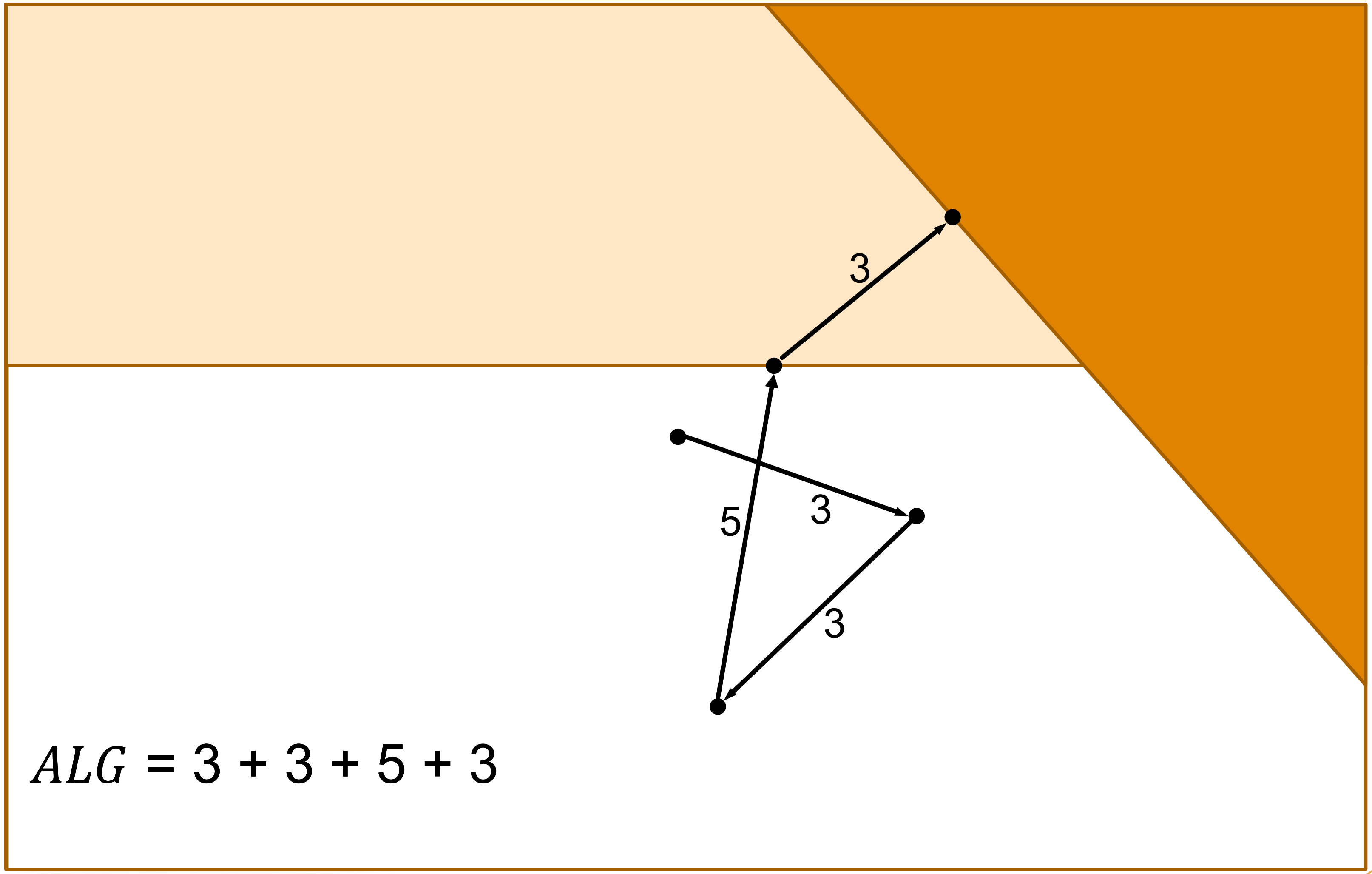
Convex Body Chasing – The Problem



Convex Body Chasing – The Problem

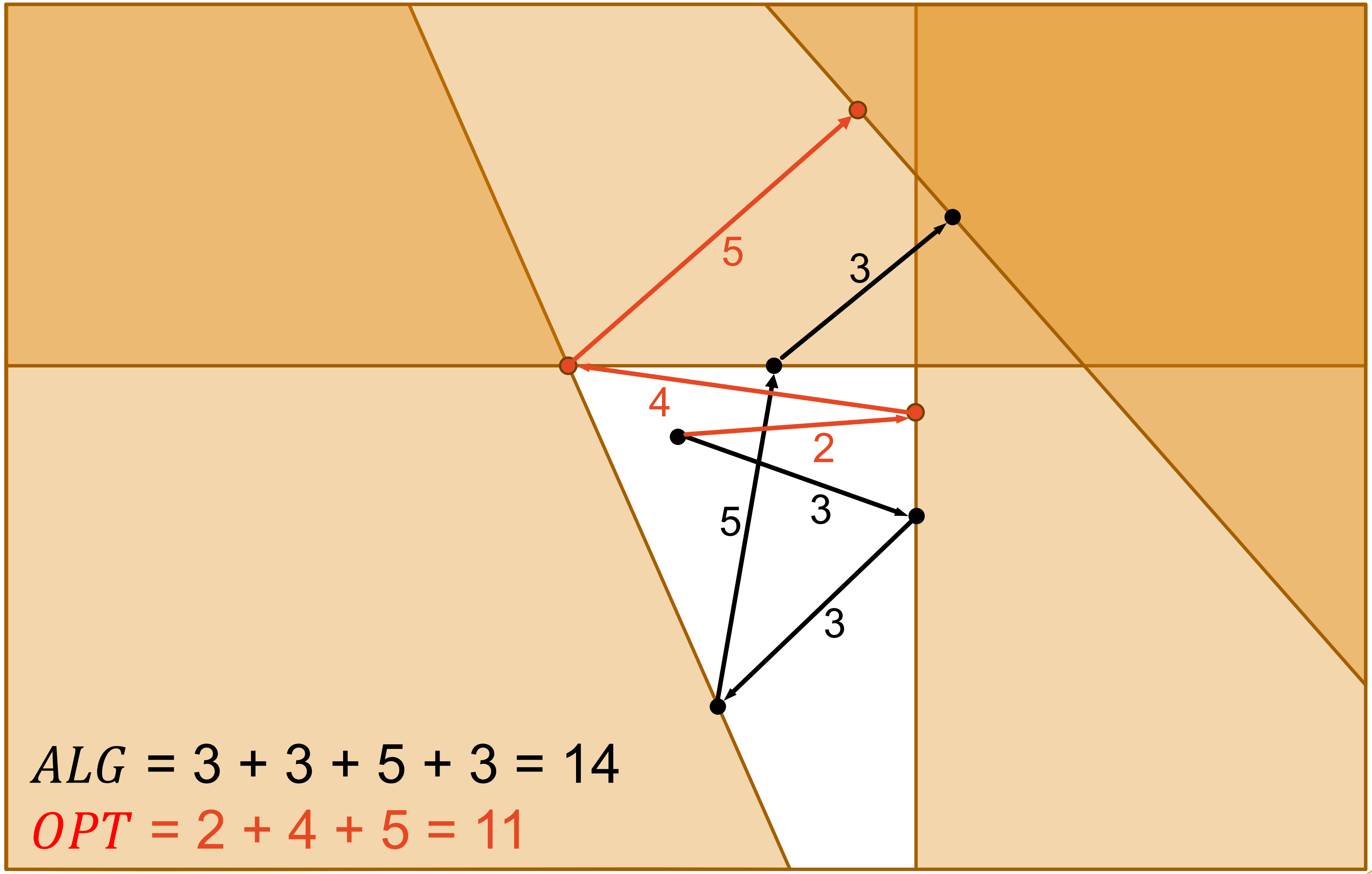


Convex Body Chasing – The Problem



$ALG = 3 + 3 + 5 + 3$

Convex Body Chasing – The Problem



$ALG = 3 + 3 + 5 + 3 = 14$

$OPT = 2 + 4 + 5 = 11$

Formal Definition

$r > 0,$ "$r = OPT$"

- Input: ^ convex sets $K_1, K_2, K_3, \dots, K_T \subseteq \mathbb{R}^d$
- Choose *online* $x_i \in K_i$
- Cost $ALG = \sum_{i=1}^{T} ||x_i - x_{i-1}||$
- Goal – minimize ~~competitive ratio~~ $f(d)$ s.t. $ALG \leq f(d) \cdot r$

~~$$\mathrm{cr}(ALG) := \max_{\text{instance } \sigma} \frac{ALG(\sigma)}{OPT(\sigma)}$$~~

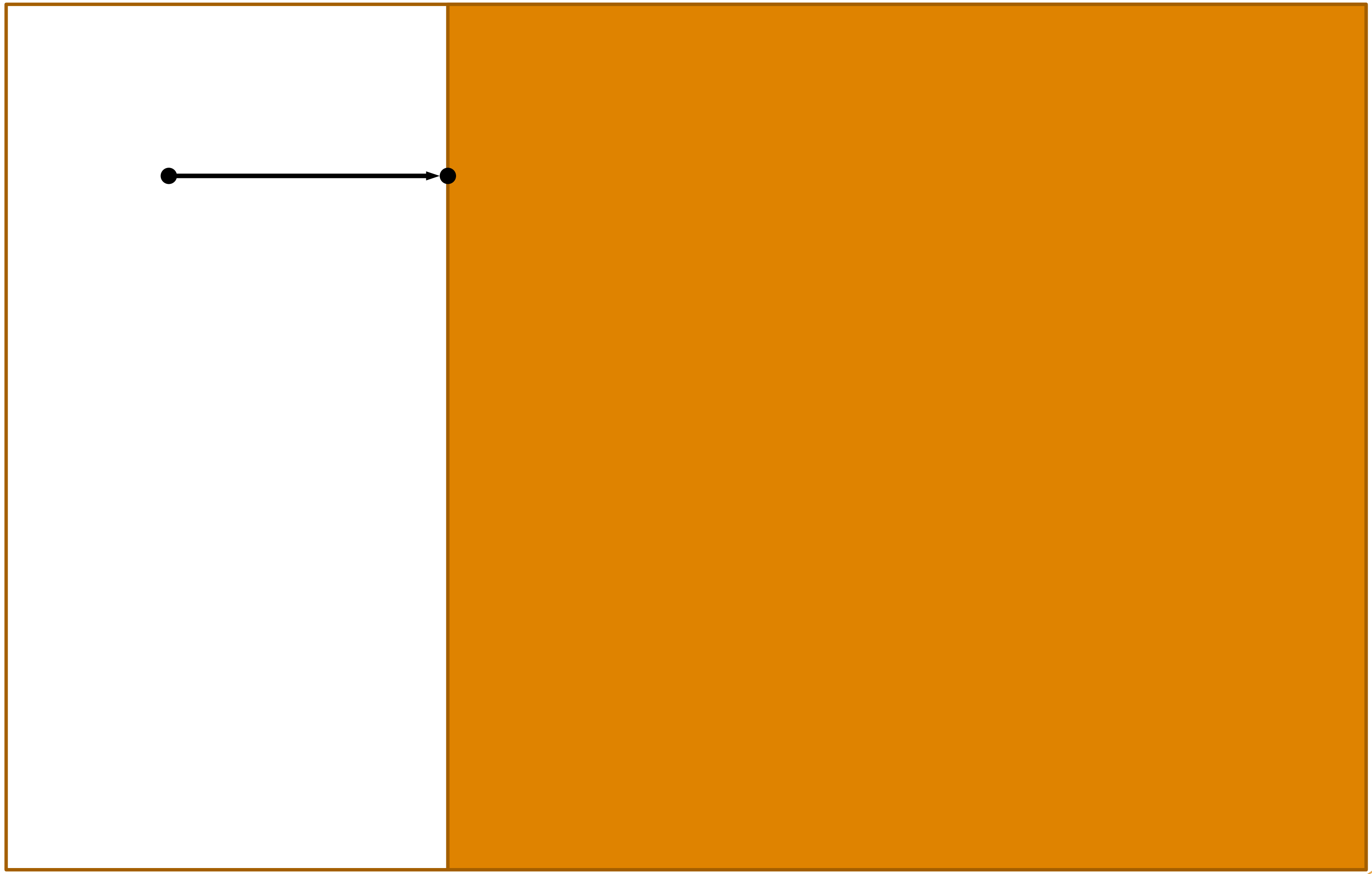
- Equivalent problem
  - Guess-and-double

Motivation

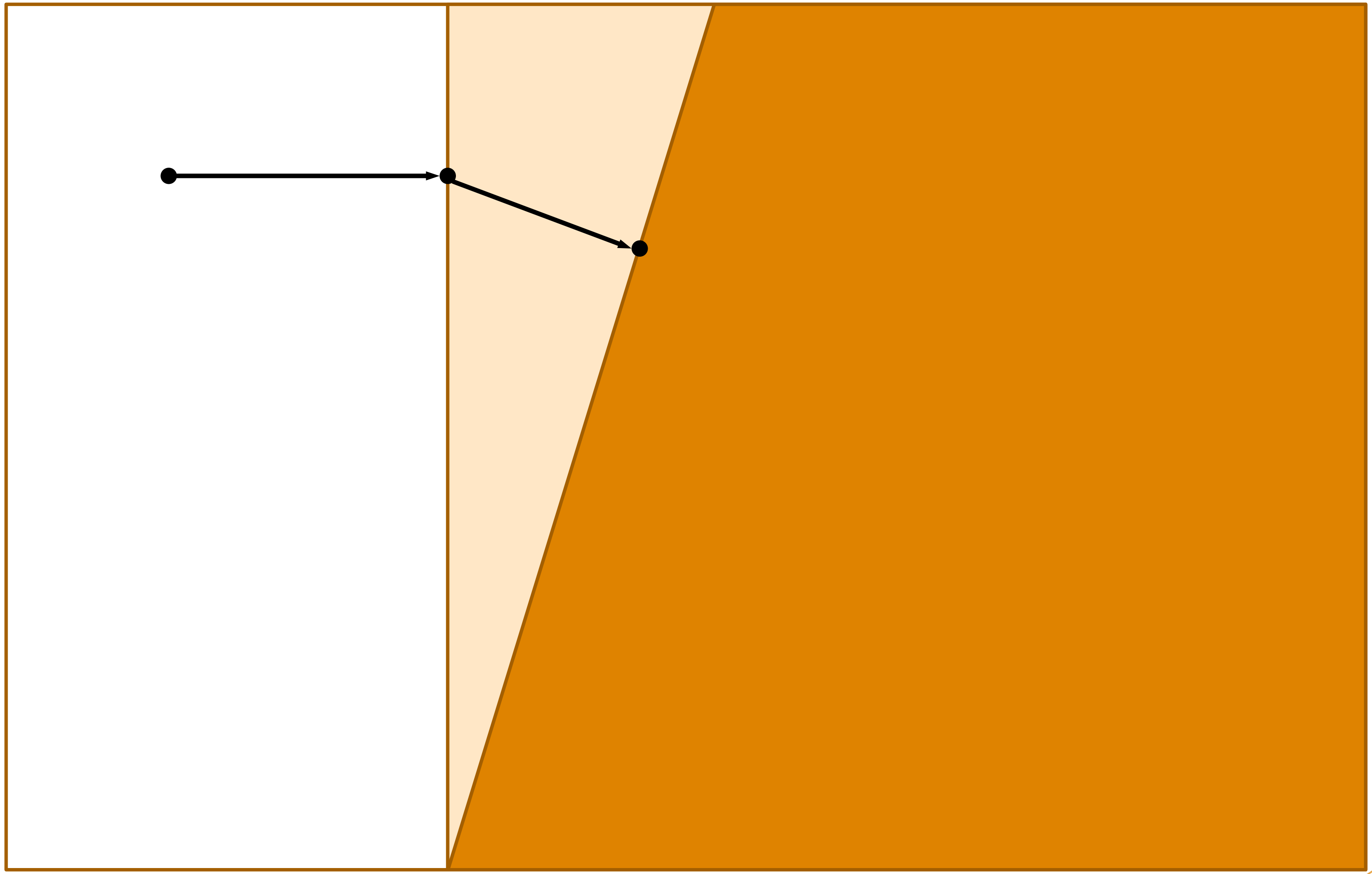
- Function chasing / Smooth online convex optimization
  - Function chasing ≅ body chasing

    [Bubeck, Lee, Li, Sellke 19]
- Metrical task systems
- Paging, k-Server (fractional)

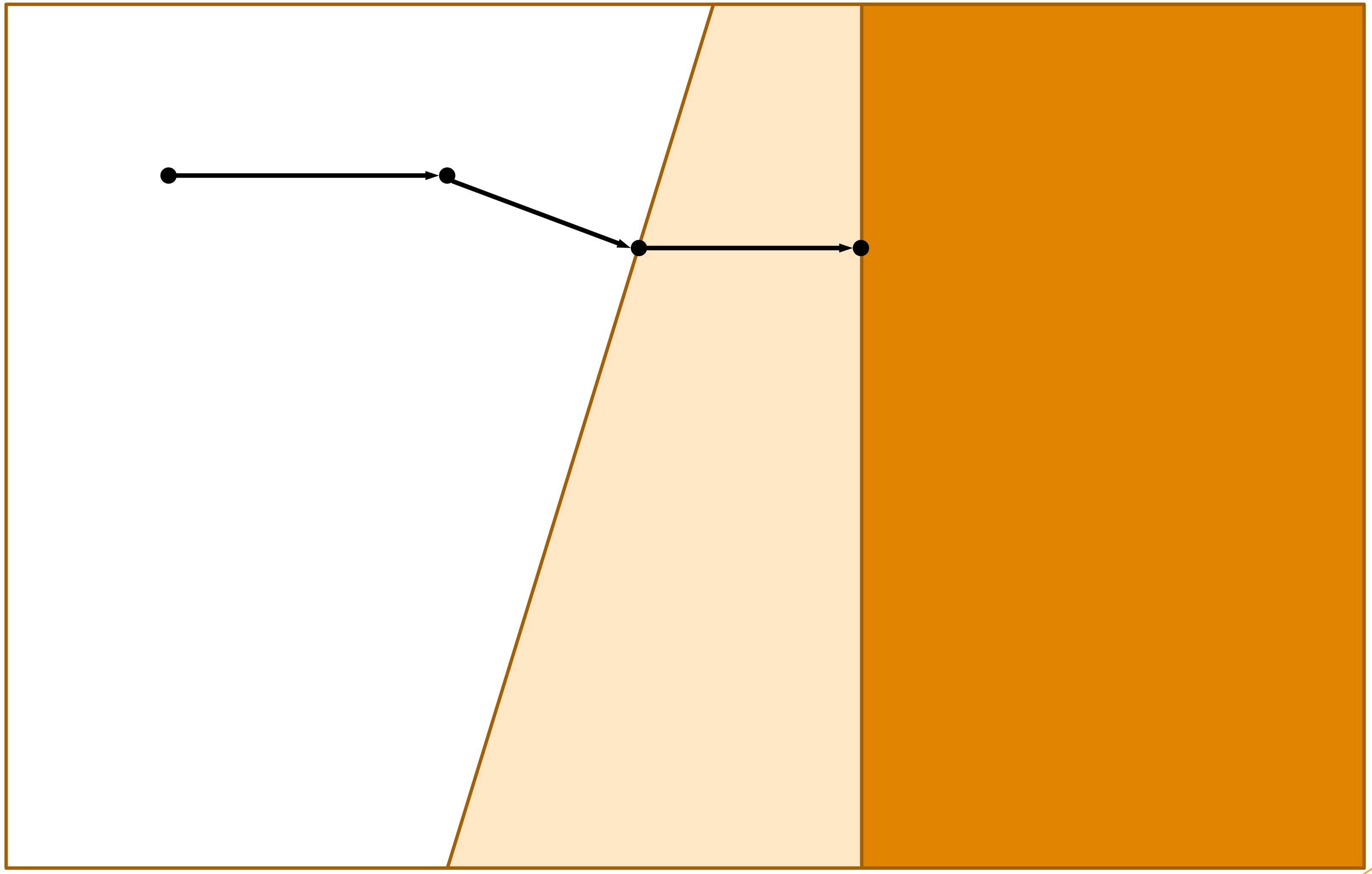
Nested – Important Special Case



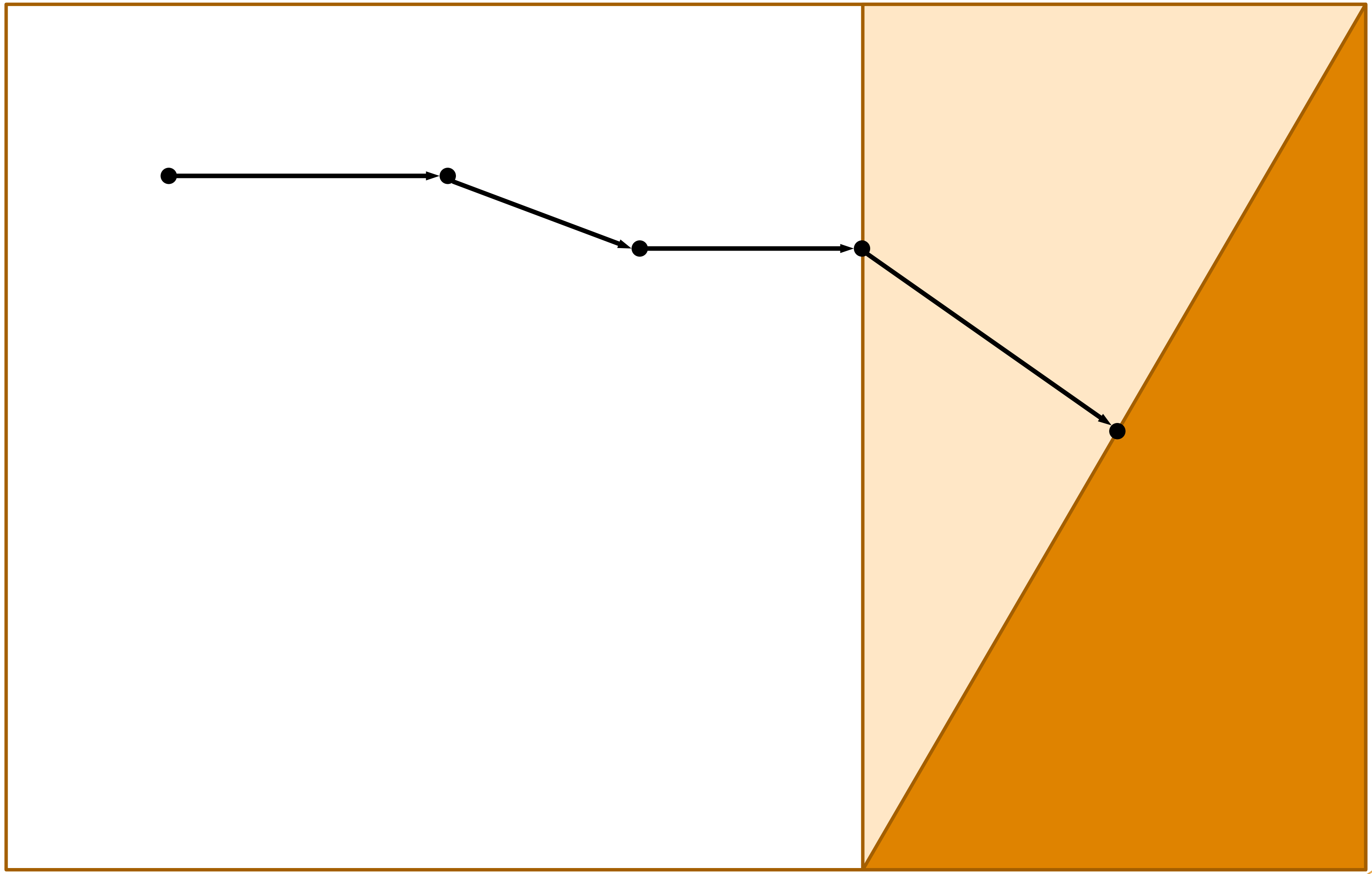
Nested – Important Special Case



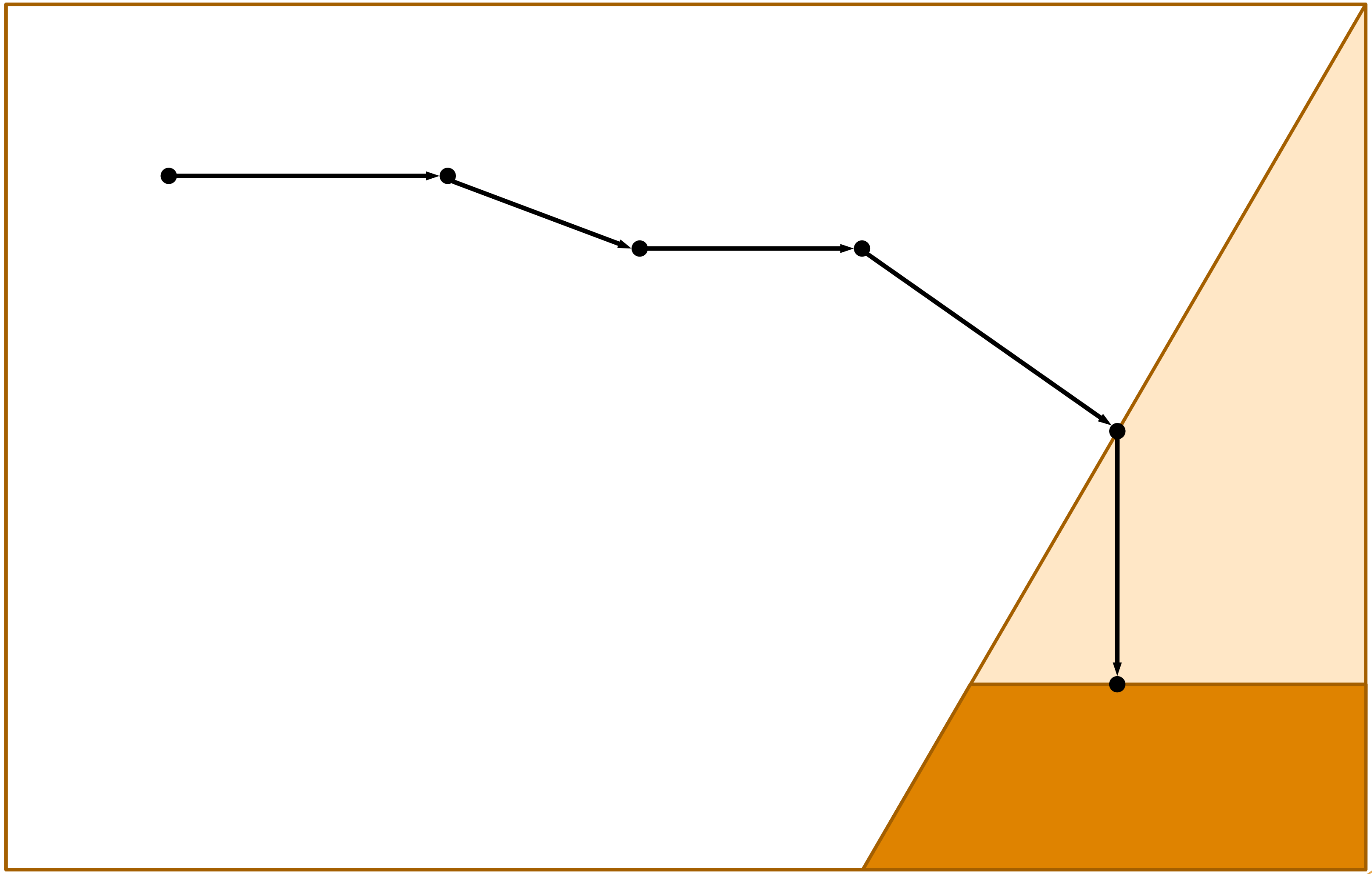
Nested – Important Special Case



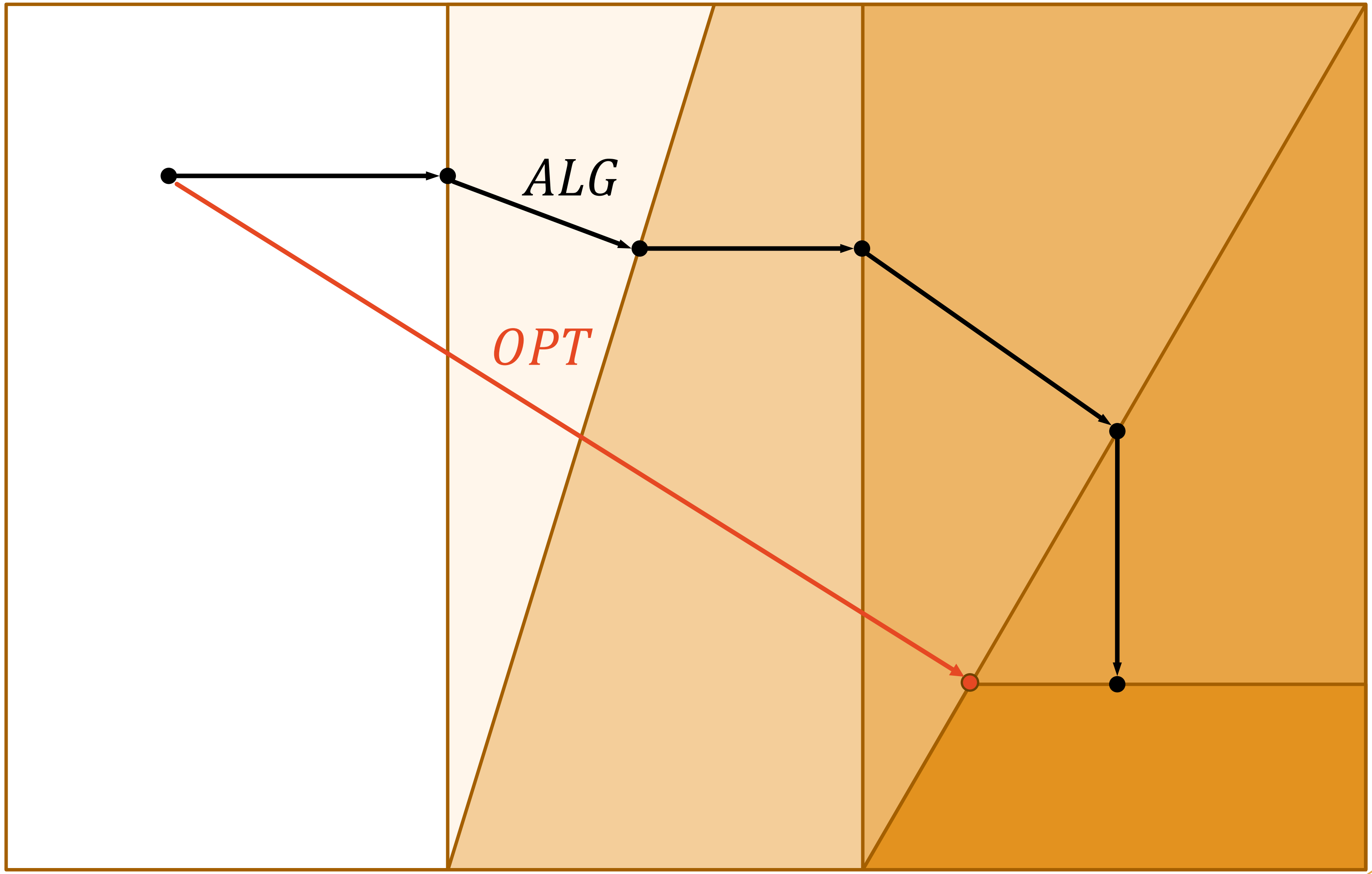
Nested – Important Special Case



Nested – Important Special Case



Nested – Important Special Case

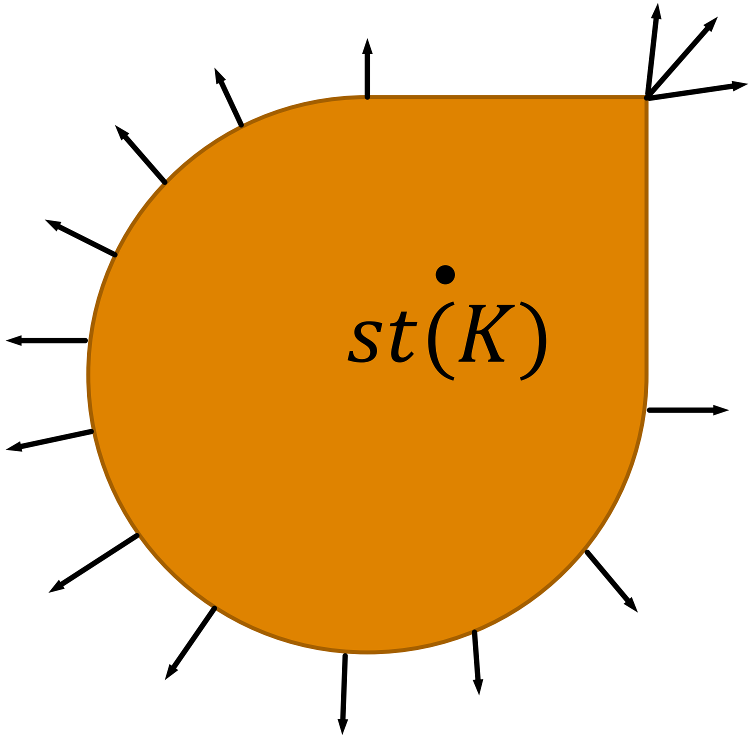


Progress

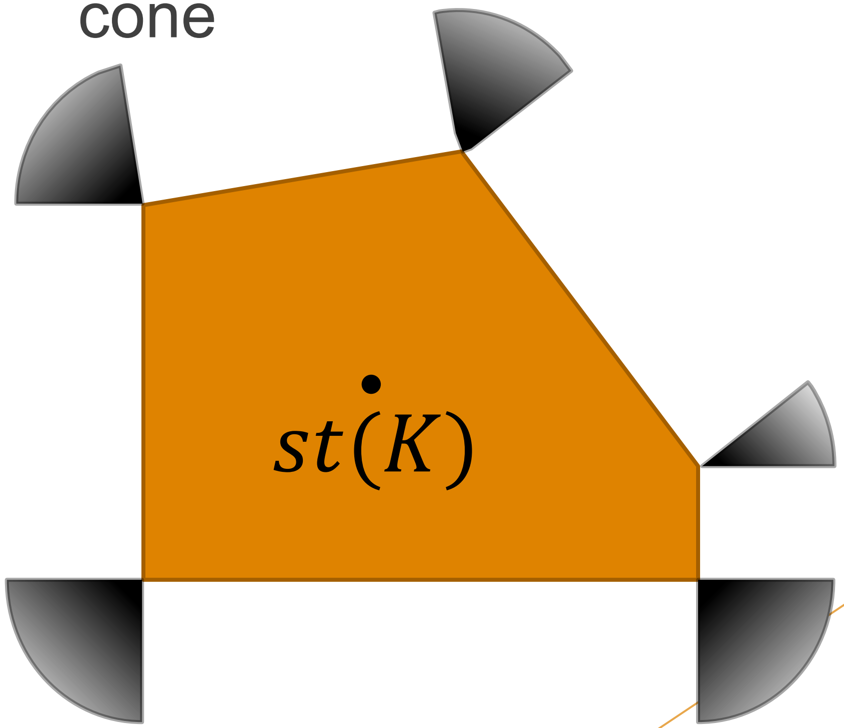
- Previous best known
  - Lower bound: $\Omega(\sqrt{d})$ [Friedman, Linial 93]
  - Nested: $O(\sqrt{d \log d})$, simple $O(d)$ [Bubeck, Klartag, Lee, Li, Sellke 20]
  - General: $2^{O(\mathrm{d})}$ [Bubeck, Lee, Li, Sellke 19]
- This talk
  - General: simple $O(d)$

Steiner Point

- Average of extreme points in all directions

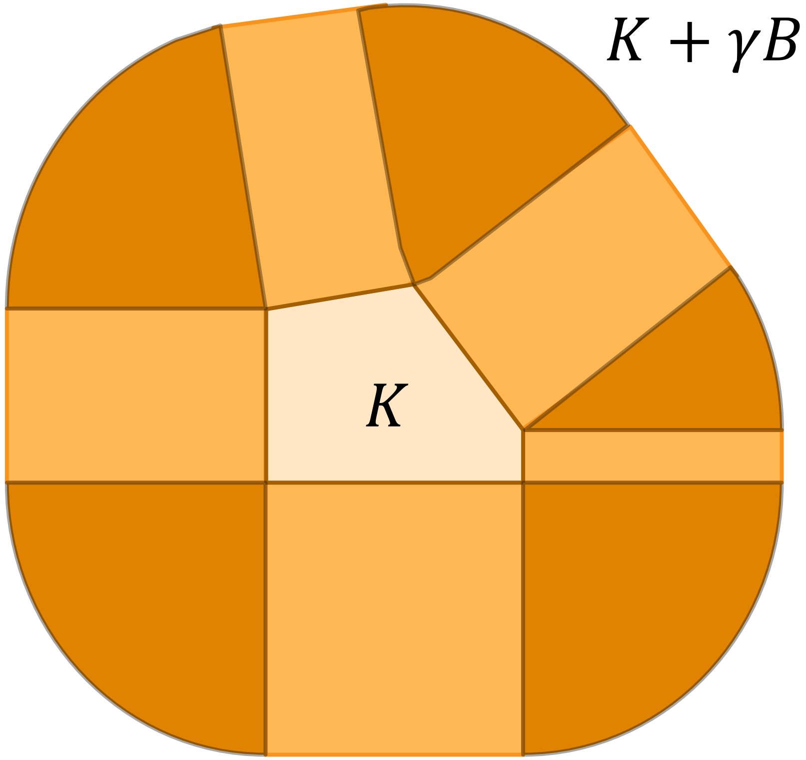


- Average of extreme points weighted by size of normal cone



Steiner Point Definition

- $st(K) := \lim_{\gamma \to \infty} cg(K + \gamma B)$
  - B := unit ball
- More equivalent definitions in Part 2



Steiner Point Algorithm (Nested)

[Bubeck, Klartag, Lee, Li, Sellke 20]

- $x_t = st(K_t)$
- $O(d)$ competitive
- Simple and beautiful!!

Reducing General to Nested

- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

- Given:
  - General instance: $r > 0, K_1, \ldots, K_T$
  - $O(d)$ competitive nested algo $NEST$
- Goal: Construct sets $\Omega_1, \ldots, \Omega_T$ s.t.
  - Ω_t convex
  - $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
  - $NEST(\Omega_1, \ldots, \Omega_t) \leq O(d) \cdot r$
  - $NEST(\Omega_i) \in K_i$

Work Function

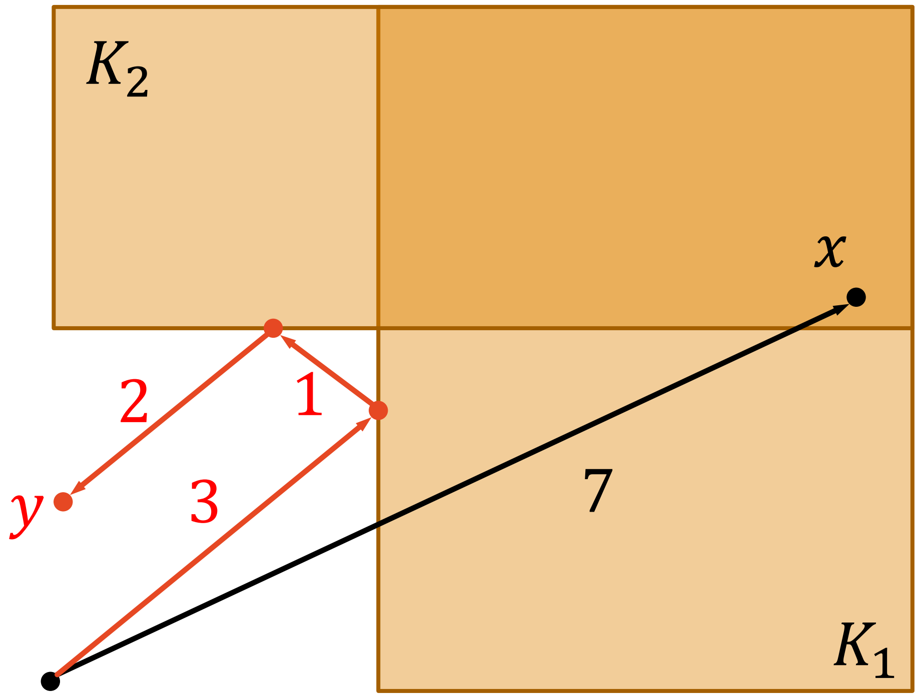
- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

- Classical technique for related problems
- $w_t(x) :=$ min cost to satisfy requests $1, \dots, t$ and end at x

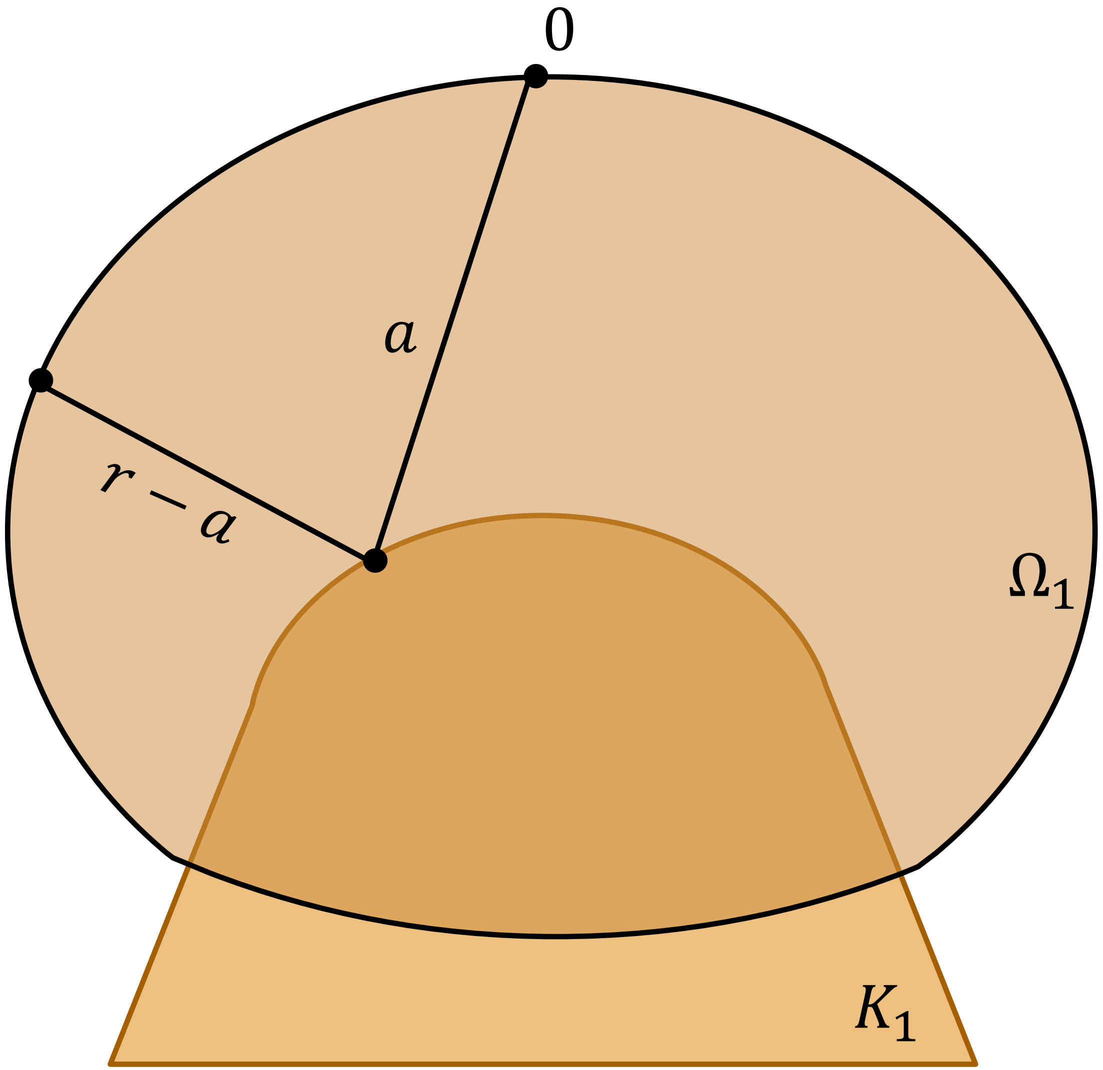
$$= \min_{y_i \in K_i} \sum_{i=1}^{t} \|y_i - y_{i-1}\| + \|y_t - x\|$$

$w_2(x) = 7$

$w_2(y) = 3 + 2 + 1$



Defining Ω_t



- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \lesssim O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

$$\Omega_t \coloneqq \{x \mid w_t(x) \leq r\}$$

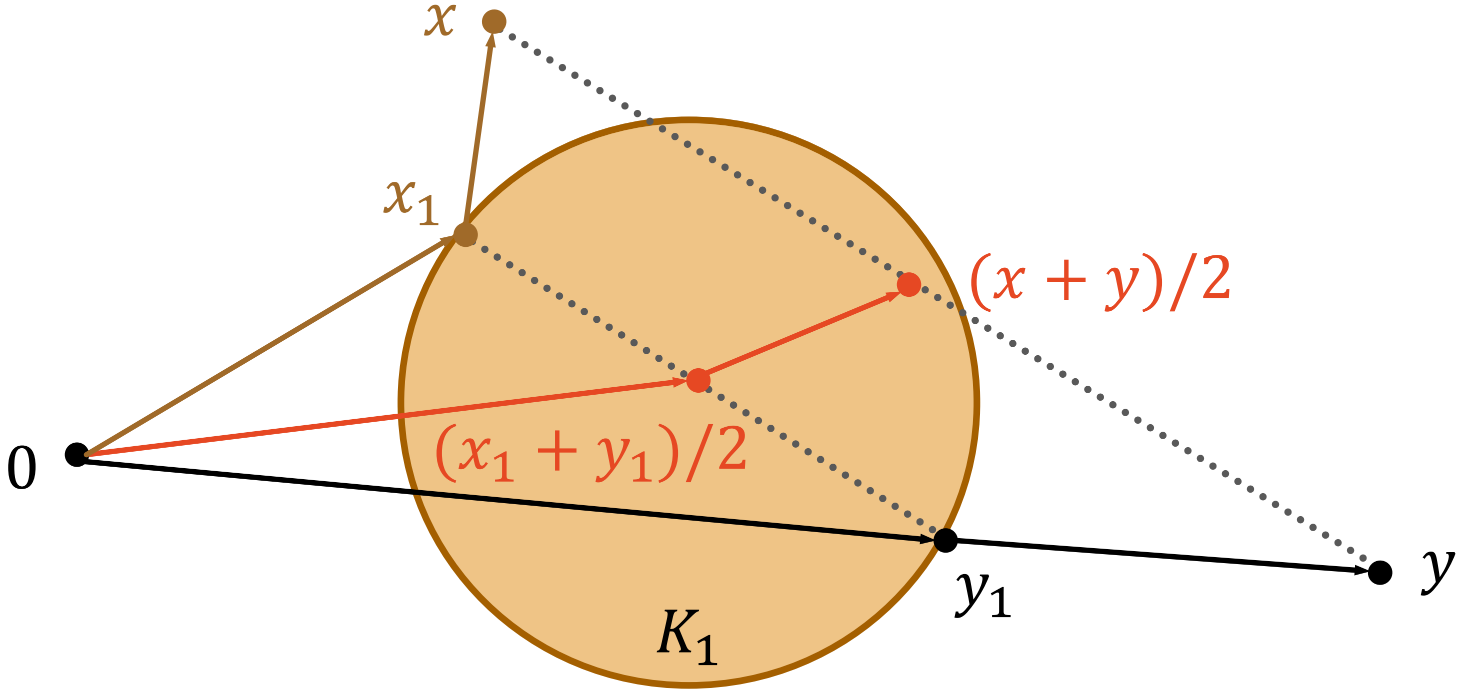
Sublevel set of work function

Convexity

- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

w_t convex

$\Longrightarrow \Omega_t = \{x \mid w_t(x) \leq r\}$ convex



Nestedness

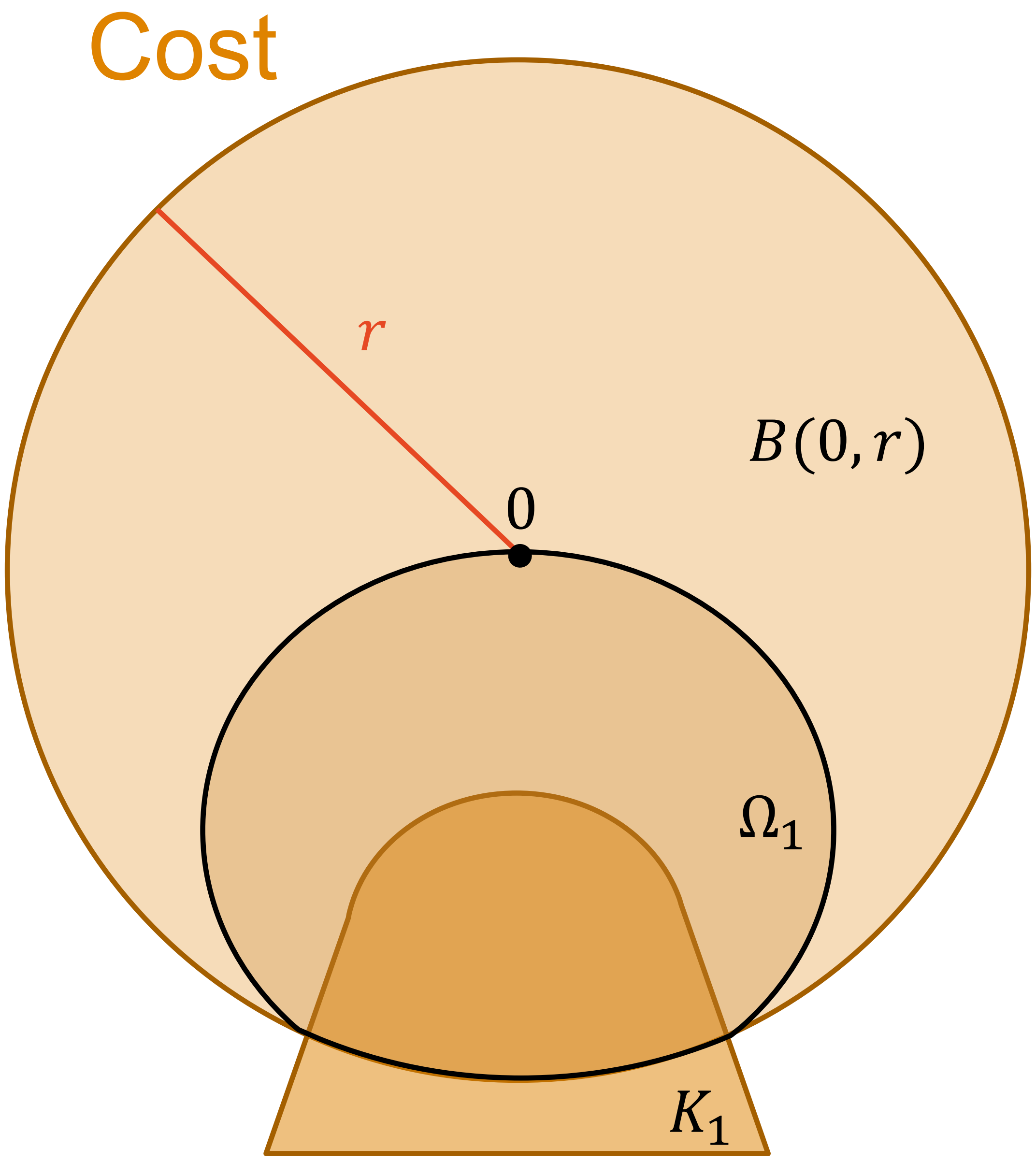
- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \lesssim O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

$$w_t(x) \leq w_{t+1}(x) \Longrightarrow \{x \mid w_t(x) \leq r\} \supseteq \{x \mid w_{t+1}(x) \leq r\}$$

$$\Omega_t \supseteq \Omega_{t+1}$$

Cost to satisfy requests $1, \dots, t+1$ and end at x

Cost to satisfy requests $1, \dots, t$ and end at x



- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

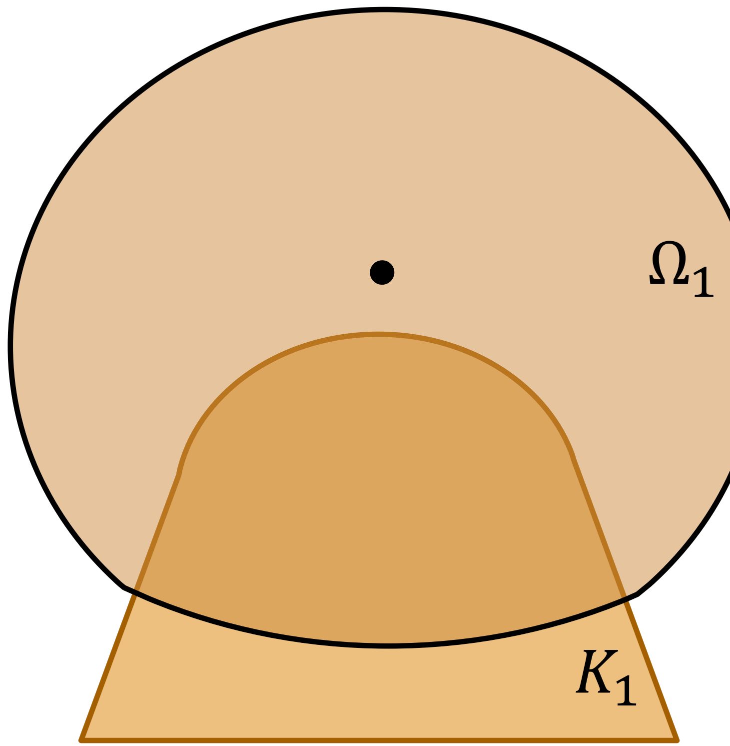
$$\Omega_1 = \{ x \mid w_1(x) \leq r \} \subseteq B(0, r)$$

$$\begin{aligned} cost &\leq O(d) \cdot OPT(\Omega_1, \dots, \Omega_T) \\ &\leq O(d) \cdot diam(\Omega_1) \\ &\leq O(d) \cdot r \end{aligned}$$

(In)feasibility

- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $NEST(\Omega_t) \in K_i$

- $\Omega_t \nsubseteq K_t$
  - May play infeasible point



Feasibility

- Ω_t convex
- $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
- $cost \leq O(d) \cdot r$
- $st(\Omega_t) \in K_i$
- $NEST(\Omega_t) \in K_i$

Feasibility Lemma: $st(\Omega_t) \in K_t$

Main Theorem: $x_t = st(\Omega_t)$ is $O(d)$-competitive

[Argue, Gupta, Guruganesh, Tang 20]

Proof of Feasibility Lemma

$K_t = \{\, x \mid \langle a, x\rangle \geq b \,\}$ (w.l.o.g.)

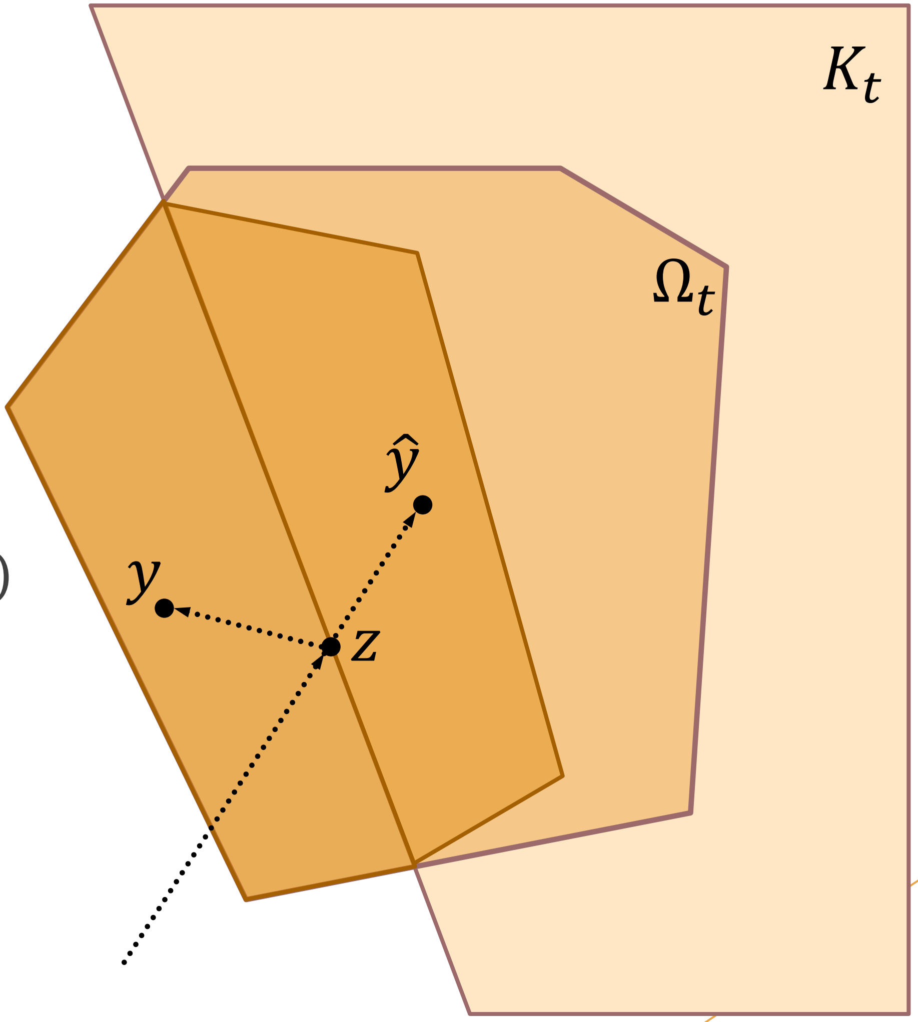
For $y \notin K_t$, $\hat{y} \coloneqq reflect(y)$

Claim: If $y \in \Omega_t$ then $\hat{y} \in \Omega_t$

$$w_t(y) = \min_{z \in K_t} \|y - z\| + w_{t-1}(z)$$

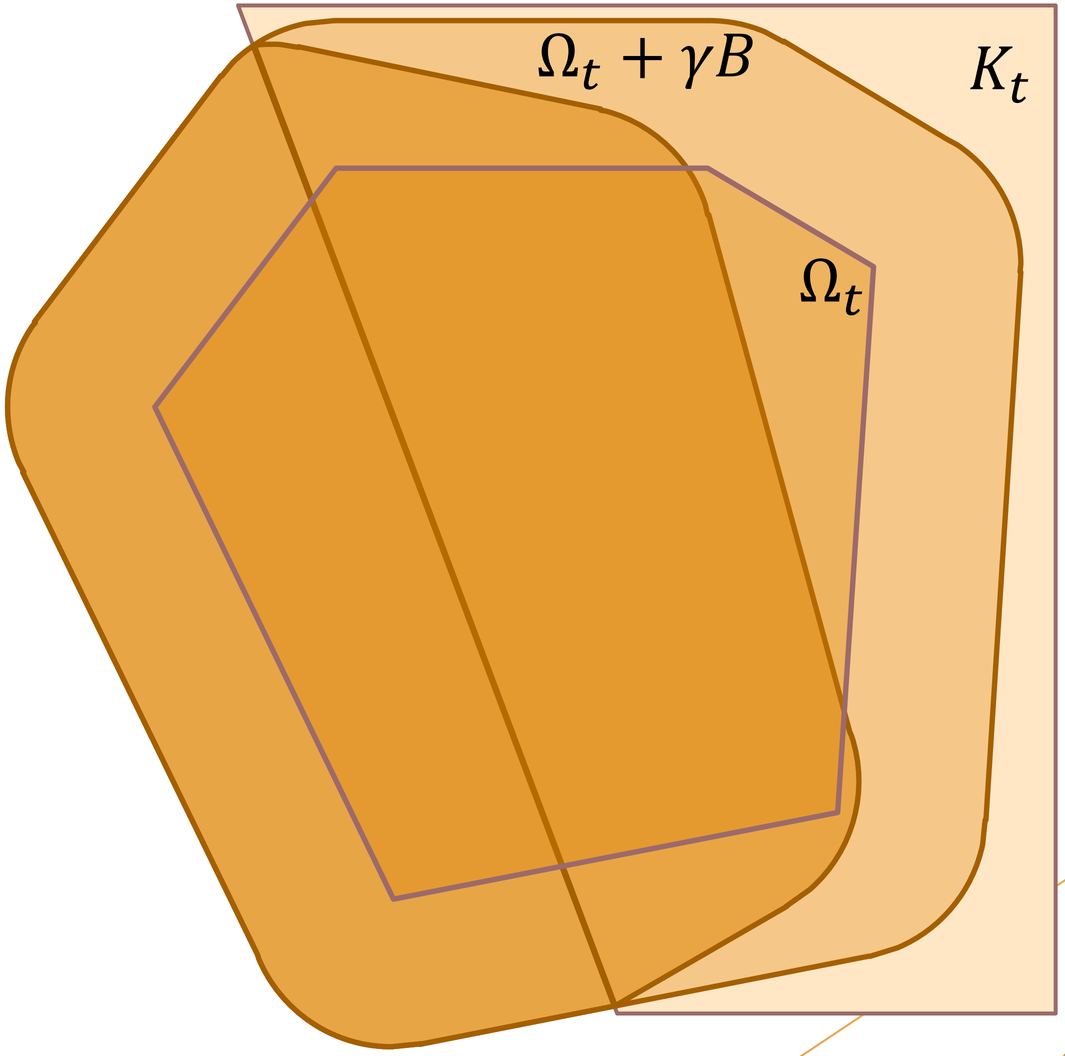
$$\Rightarrow \; w_t(\hat{y}) \leq w_t(y) \leq r$$

☺



Goal: $st(\Omega_t) \in K_t$
$\Omega_t = \{x \mid w_t(x) \leq r\}$
$st(\Omega_t) = \lim_{\gamma \to \infty} cg(\Omega_t + \gamma B)$

Proof of Feasibility Lemma

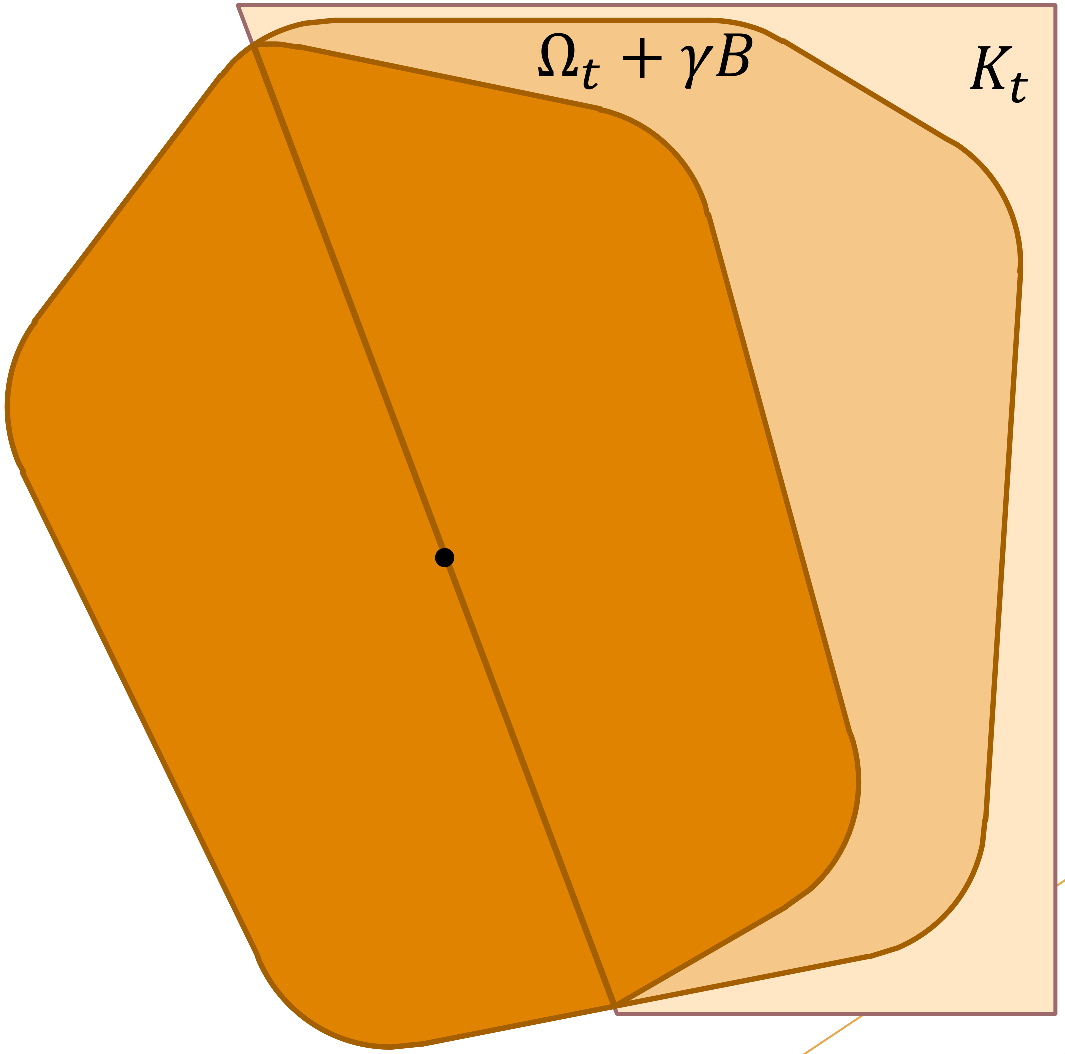


Goal: $st(\Omega_t) \in K_t$

$\Omega_t = \{x \mid w_t(x) \leq r\}$

$st(\Omega_t) = \lim_{\gamma \to \infty} cg(\Omega_t + \gamma B)$

Proof of Feasibility Lemma

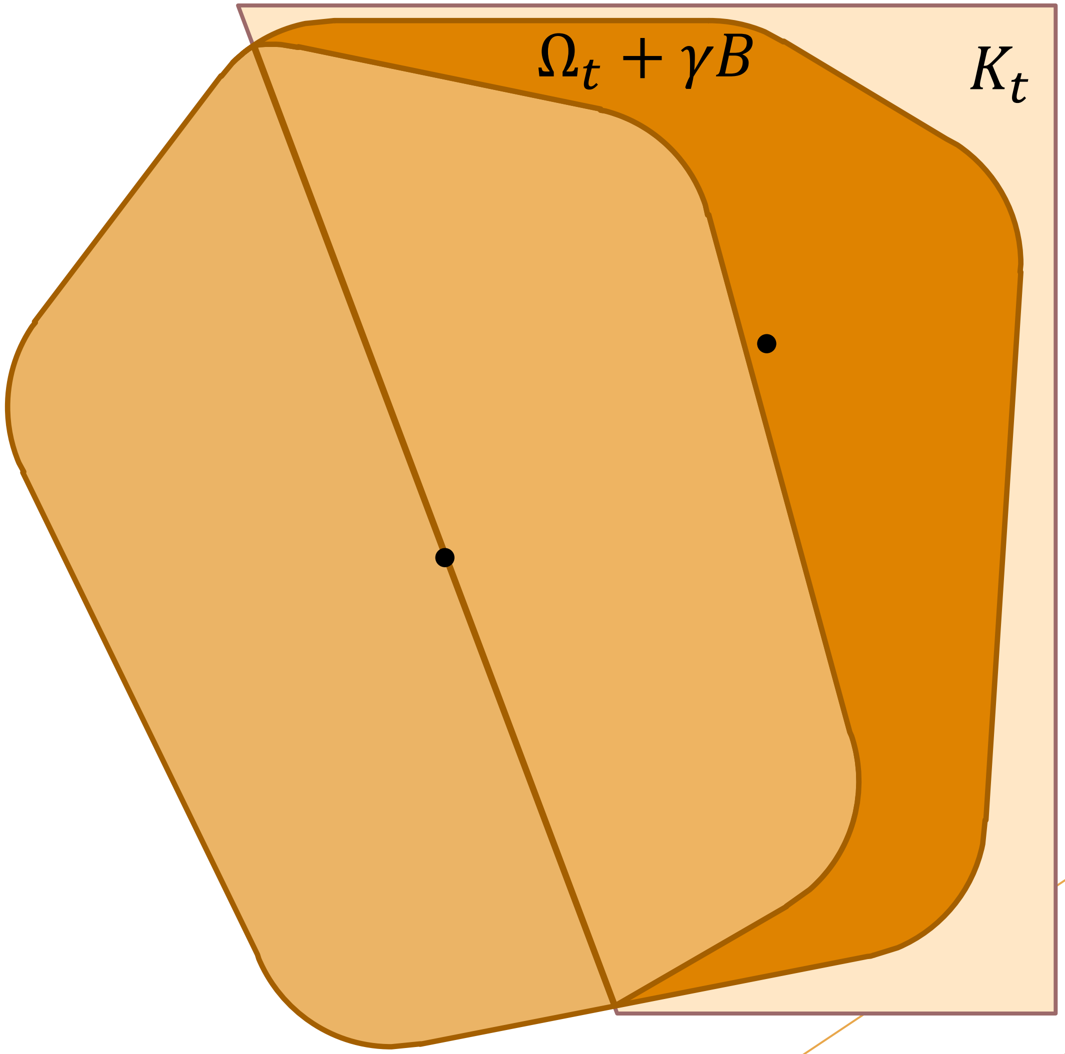


Goal: $st(\Omega_t) \in K_t$

$\Omega_t = \{x \mid w_t(x) \leq r\}$

$st(\Omega_t) = \lim_{\gamma\to\infty} cg(\Omega_t + \gamma B)$

Proof of Feasibility Lemma



Goal: $\boldsymbol{st(\Omega_t) \in K_t}$

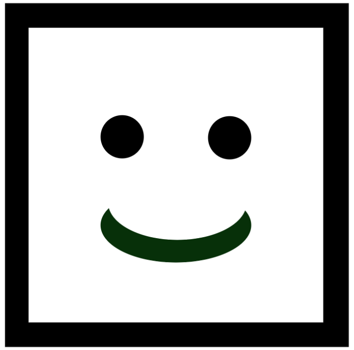
$\Omega_t = \{x \mid w_t(x) \le r\}$

$st(\Omega_t) = \lim_{\gamma \to \infty} cg(\Omega_t + \gamma B)$

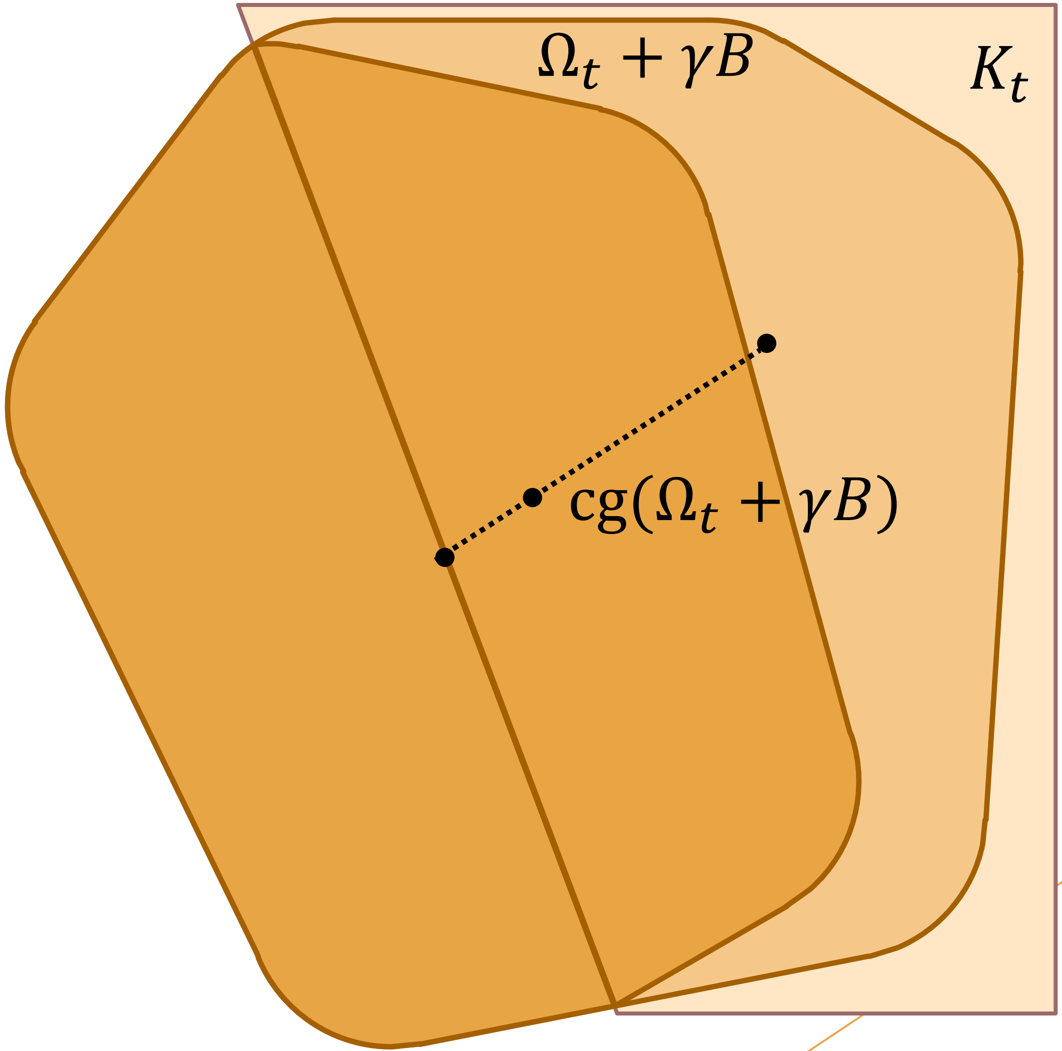
Proof of Feasibility Lemma

$\mathrm{cg}(\Omega_t + \gamma B) \in K_t$
for all $\gamma \geq 0$

$\mathrm{st}(\Omega_t) \in K_t$

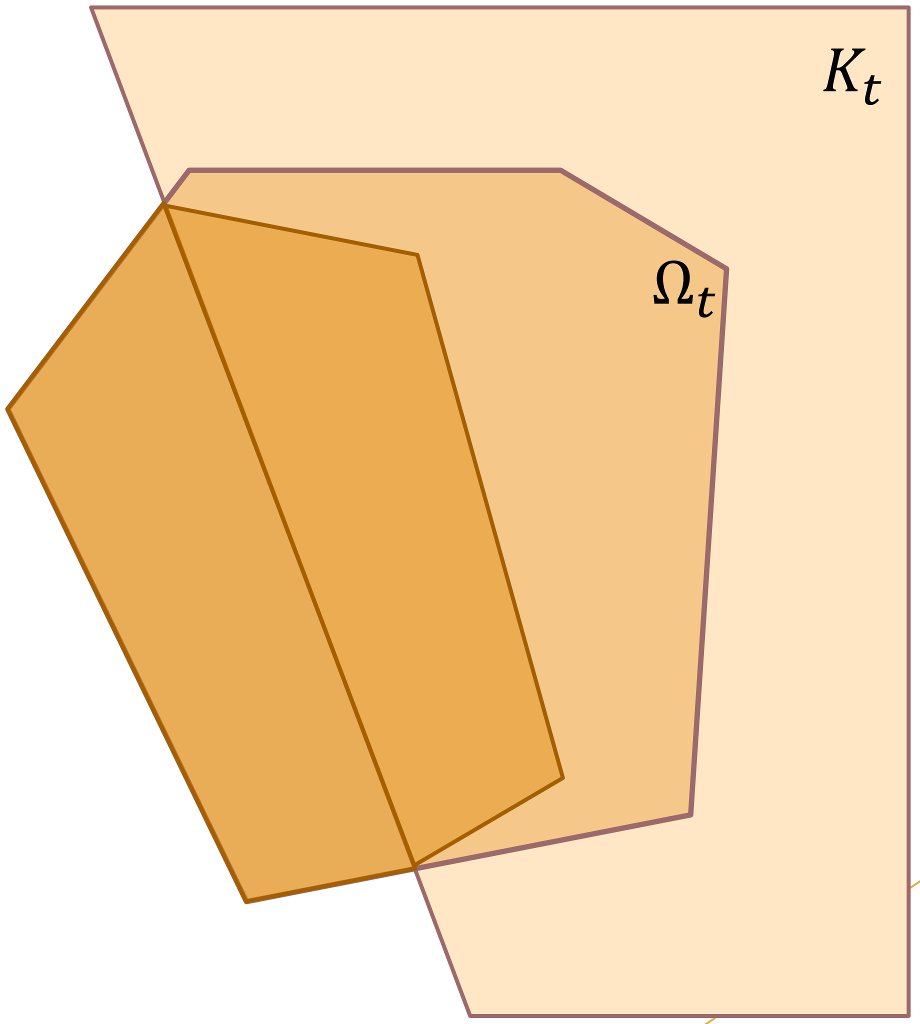


Goal: $st(\Omega_t) \in K_t$
$\Omega_\mathrm{t} = \{x \mid w_t(x) \leq r\}$
$st(\Omega_t) = \lim_{\gamma \to \infty} cg(\Omega_t + \gamma B)$



Recap of Part 1

- Algo: $x_t = st(\Omega_t)$
  - $\Omega_t = \{x \mid w_t(x) \leq r\}$
- $O(d)$ competitiveness
  - Ω_t convex, $\Omega_1 \supseteq \Omega_2 \supseteq \cdots \supseteq \Omega_T$
  - Feasibility: $x_t \in K_t$
  - $ALG \leq O(d) \cdot r$



Part 2: Functional Steiner Point

- Instead of using Steiner out-of-the-box, redefine it.
- Define the *Functional Steiner Point* of a convex function, and apply to work function.
- Again two formulas via divergence theorem. Support function becomes Fenchel dual.
- Same $\min(d, O(\sqrt{d \cdot log(T)}\,)$ competitive ratio.
- Same proofs as nested chasing with Steiner point in previous talk.
- Coincides with Steiner point of a large level set.

Steiner Point: Two Equivalent Definitions

- Definition ([Ste 1840]): the **Steiner point** $st(K) \in K$ of a convex set $K \in \ell_d^2$ is:

$$st(K) = \fint_{|v|<1} f_K(v)dv = d \cdot \fint_{|\theta|=1} h_K(\theta)\theta d\theta$$

- Both integrals are normalized to be **expectations** over the unit ball and sphere in $\mathbb{R}^d$. And:

Support Function (Scalar)

Extreme Point (Vector)

$$f_K(v) = argmax_{x \in K}\langle v, x\rangle,$$

$$h_K(\theta) = \max_{x \in K}\langle \theta, x\rangle = \langle \theta, f_K(\theta)\rangle.$$

- First definition is **primal**: $f_{K_t}(v) \in K_t$ implies $st(K_t) \in K_t$ by convexity.
- Second definition is **dual**: used to upper bound movement.

Why Do The Definitions Agree?

$$st(K_t) = ⨍_{|v|<1} f_{K_t}(v)dv = d \cdot ⨍_{|\theta|=1} h_{K_t}(\theta)\theta d\theta$$

$$f_K(v) = \underset{x \in K}{\operatorname{argmax}}(\langle v, x\rangle), \qquad h_K(\theta) = \max_{x \in K}(\langle \theta, x\rangle) = \langle \theta, f_K(\theta)\rangle$$

- Key: $f_K = \nabla h_K$, and $\theta = \hat{n}(\theta)$ is the outward normal to the sphere at θ.
- General **Gauss-Green** Theorem (variant of Divergence Theorem):

$$\int_U \nabla h(v)dv = \int_{\partial U} h(v)\hat{n}(v)dv.$$

Both sides are $\nabla_x \int_{U+x} h(v)dv$

- Factor d from change in total measure – the colored integrals are normalized.

Nested Chasing with Steiner Point

- Start with K_1 a unit ball, request sequence $K_1 \supseteq K_2 \supseteq K_3 \ldots$. Set $x_t = st(K_t)$.
- Claim: total movement $\leq d$.
- Nested condition is equivalent to support function decreasing:

$$h_{K_1}(\theta) \geq h_{K_2}(\theta) \geq h_{K_3}(\theta) \ldots$$

- Triangle inequality now says:

$$|\mathrm{st}(K_{t-1}) - \mathrm{st}(K_t)| \leq d \cdot \fint_{|\theta|=1} h_{K_{t-1}}(\theta) - h_{K_t}(\theta) d\theta.$$

- Summing over t for total movement, RHS telescopes! Hence upper bound of d.
- To get $O\left(\sqrt{d \cdot \log(T)}\right)$: only very small sets of the sphere can correlate much.

Defining Functional Steiner Point

- Two definitions of Steiner point, equivalent by Gauss-Green and $\nabla h = f$:

$$st(K_t) = \fint_{|v|<1} f_{K_t}(v)dv = d \cdot \fint_{|\theta|=1} h_{K_t}(\theta)\theta d\theta$$

$$f_K(v) = argmax_{x\in K}(\langle v, x\rangle) = \nabla h_K(\theta), \qquad h_K(\theta) = \max_{x\in K}(\langle \theta, x\rangle)$$

- We replace h_K with the Fenchel dual W_t^* of W_t to define the *functional Steiner point:*

$$st(W_t) = \fint_{|v|<1} v_t^* dv = (-d) \cdot \fint_{|\theta|=1} W_t^*(\theta) \cdot \theta d\theta$$

$$v_t^* = \operatorname*{argmin}_{x\in\mathbb{R}^d}(W_t(x) - \langle v, x\rangle) = -\nabla W_t^*(v)$$

$$W_t^*(\theta) = \min_{x\in\mathbb{R}^d}(W_t(x) - \langle \theta, x\rangle) = W_t(v_t^*) - \langle \theta, v_t^*\rangle$$

- First defn: $\mathbb{E}^{|v|<1}[\, argmin_x(W_t(x) - \langle v, x\rangle)]$. Aka follow the perturbed leader.
- $W_t^*(\theta)$ measures the height of a θ –slope tangent plane to W_t at input 0.

Functional Steiner Point is an Online Selector

$$st(W_t) = \fint_{|v|<1} v_t^* dv\,; \qquad v_t^* = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} (W_t(x) - \langle v, x \rangle)$$

Lemma: $st(W_t) \in K_t$.

By construction, $st(W_t)$ is a weighted average of v with $|\nabla W_t(v)| < 1$.

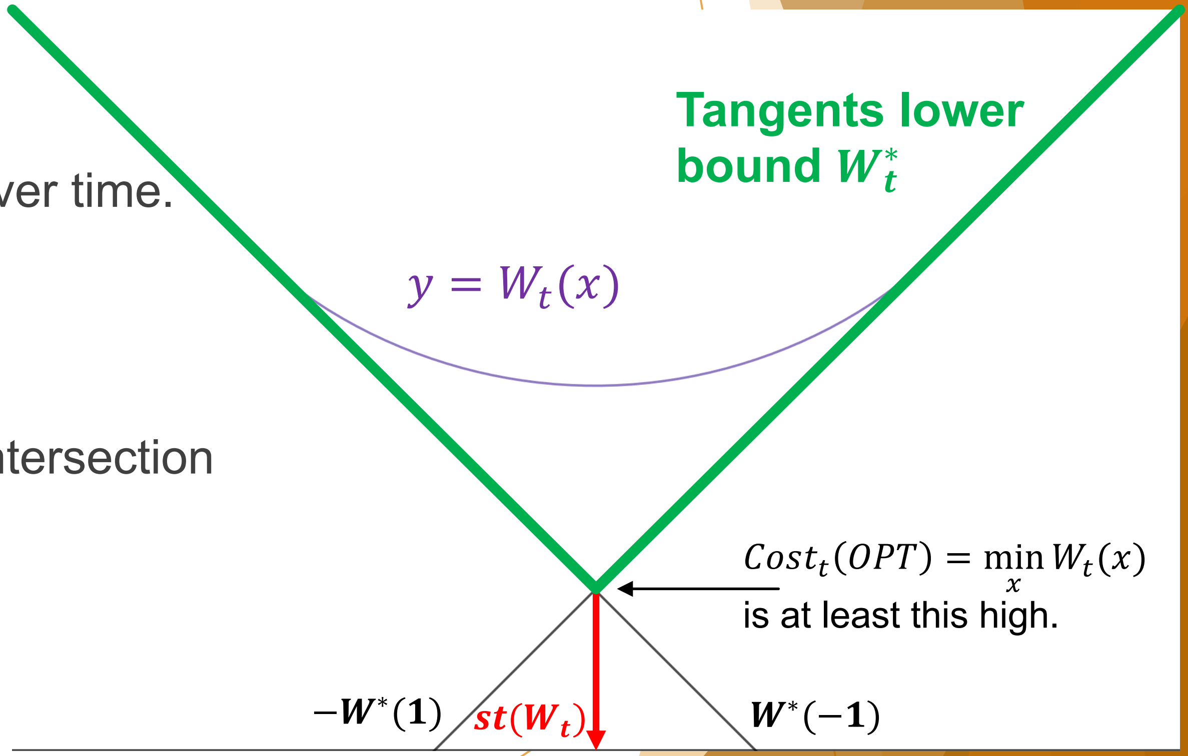
$|\nabla W_t(v)| < 1$ implies $v \in K_t$.

If $v \notin K_t$, the best path ending at v came from $\mathrm{w} \in K_t$. $\nabla W_t(v)$ points in the direction $\overline{vw}$.

Lemma follows by convexity of K_t.

The Dual Definition in 1 Dimension

- Functional Steiner Point in 1 dimension: intersect tangent lines with slopes ±1.
- Equivalent to $st(W_t) = \frac{W_t^*(1) - W_t^*(-1)}{2}$. Tangents move up over time.
- Movement of $st(W_t) \leq$ tangents' total upward movement.
- Tangents' total upward movement = height of tangents' intersection
- **Height of tangents' intersection $\leq \min_x W_t(x)$.**
- Combining, Functional Steiner is 1-competitive.



Functional Steiner Point is d-Competitive

Recall: $st(W_t) = (-d) \cdot ⨍_{|\theta|=1} W_t^*(\theta)\theta d\theta$; $W_t^*(\theta) = \min_{x \in \mathbb{R}^d}(W_t(x) - \langle \theta, x \rangle)$

Properties of work and dual work function:

1. $W_t^*(\theta)$ is **concave**, **increasing in time from $W_0^*(\theta) = 0$**.
2. $\min_x (W_t(x)) = \mathrm{Cost}(\mathrm{OPT}_t)$.

Therefore:

$$\sum_{t \le T} |st(W_t) - st(W_{t-1})| \le d \sum_{t \le T} ⨍_{|\theta|=1} |W_t^*(\theta) - W_{t-1}^*(\theta)| d\theta = d \cdot ⨍_{|\theta|=1} W_T^*(\theta) d\theta$$

$$\le d \cdot W_T^*(0) \quad = \mathrm{d} \cdot \min_x (W_T(x)) = \mathrm{d} \cdot \mathrm{Cost}(\mathrm{OPT_T}) \ \blacksquare$$

For small T, concentration of measure in the first inequality gives $O\left(\sqrt{d \cdot \log(T)}\right)$.

Chasing Convex Functions

- Chasing convex functions: same problem but with soft constraint.
- *Given online positive convex functions f_t, be competitive for:*

$$Cost(ALG) = \sum_{t=1}^{T} ||x_t - x_{t-1}|| + f_t(x_t)$$

- Previously known to be equivalent to CBC, reduction simple but ad-hoc.
- Functional Steiner point works directly here too. No reduction needed!
- Movement d-competitive, service cost $\int_0^T f_t(x_t)dt$ is 1-competitive. Overall d+1 competitive.

Other Norms

- Steiner and Functional Steiner point work in any normed space.
  - In general, integrate over v, θ in the **dual** ball/sphere.
  - Definition depends on the norm. Less obvious what measure to put on sphere.

- Theorem: Functional Steiner Point is d-competitive for chasing convex bodies in **any** normed space.

- $O\left(\sqrt{d \cdot \log(T)}\right)$ is specific to ℓ_2. Concentration of measure depends on norm.

Functional Steiner Point via Level Sets

Consider again a (convex) level set $\Omega_{t,R} = \{x: W_t(x) \leq R\}$ of W_t.

We know:

1. $\mathrm{st}(\Omega_{t,R}) \in K_t$ (first half)
2. $\mathrm{st}(W_t) \in K_t$ (this half)

Theorem: for R large enough that $K_t \subseteq \Omega_{t,R}$, we have $st(W_t) = st(\Omega_{t,R})$.

Takeaway: the two solutions in this talk are essentially equivalent!

Proof outline: all tangents with slope $|\theta| = 1$ touch the graph of W_t above $\Omega_{t,R}$.

Hence $W_t^*(\theta) = h_{\Omega_{t,R}}(\theta) - R$.

Since $\int_{|\theta|=1} R\theta d\theta = 0$, dual definitions of $st(W_t), st\left(\Omega_{t,R}\right)$ are equal.

Open Questions

- $O(\sqrt{d})$-competitive chasing.
- Mildly non-convex problems
  - [Bubeck-Rabani-S 20+]: If $d, k \geq 2$, **no** competitive algorithm to chase convex sets with k servers.
  - Quasi-convex functions?
- New Applications?
  - [Bubeck-Li-Luo-Wei 19] apply CBC to a bandit problem.
  - Do these techniques carry over to other MTS?

Thank you!

Questions?

References

- "Chasing Convex Bodies with Linear Competitive Ratio"
  Argue, Gupta, Guruganesh, Tang, *SODA '20 [This talk]*
- "A Nearly-Linear Bound for Chasing Nested Convex Bodies"
  Argue, Bubeck, Cohen, Gupta, Lee, *SODA '19*
- "Chasing Nested Convex Bodies Nearly Optimally,"
  Bubeck, Klartag, Lee, Li, Sellke, *SODA '20*
- "Competitively Chasing Convex Bodies"
  Bubeck, Lee, Li, Sellke, *STOC '19*
- "Chasing Convex Bodies and Functions"
  Friedman, Linial, *Discrete and Computational Geometry '93*
- "Chasing Convex Bodies Optimally"
  Sellke, *SODA '20 [This talk]*

Formal Definition

- Input: convex sets $K_1, K_2, K_3, \dots, K_T \subseteq \mathbb{R}^d$
- Choose *online* $x_i \in K_i$
- Cost $ALG = \sum_{i=1}^{T} ||x_i - x_{i-1}||$
- Goal – minimize competitive ratio

$$\mathrm{cr}(ALG) \coloneqq \max_{\text{instance } \sigma} \frac{ALG(\sigma)}{OPT(\sigma)}$$

Reduction – Bounded Sets, Bound Cost

- Input: $r > 0$, convex sets $K_1, K_2, K_3, \ldots, K_T \subseteq B(0, r)$
- Choose *online* $x_i \in K_i$
- Cost $ALG = \sum_{i=1}^{T} ||x_i - x_{i-1}||$
- Goal – minimize $ALG \leq f(d) \cdot r \approx f(d) \cdot diam(K_1)$

- Equivalent problem
  - Imagine $OPT = \Theta(r)$
  - Guess and double

Steiner Point Definitions

Visually intuitive

$$st(K) = \int_{\|\theta\|=1} \nabla s_K(\theta)\, d\theta$$

$$\nabla s_K(\theta) := \operatorname*{argmax}_{x \in K} \langle \theta, x \rangle$$

Useful for the proof in the previous talk

$$= d \cdot \int_{\|\theta\|=1} s_K(\theta) \cdot \theta\, d\theta$$

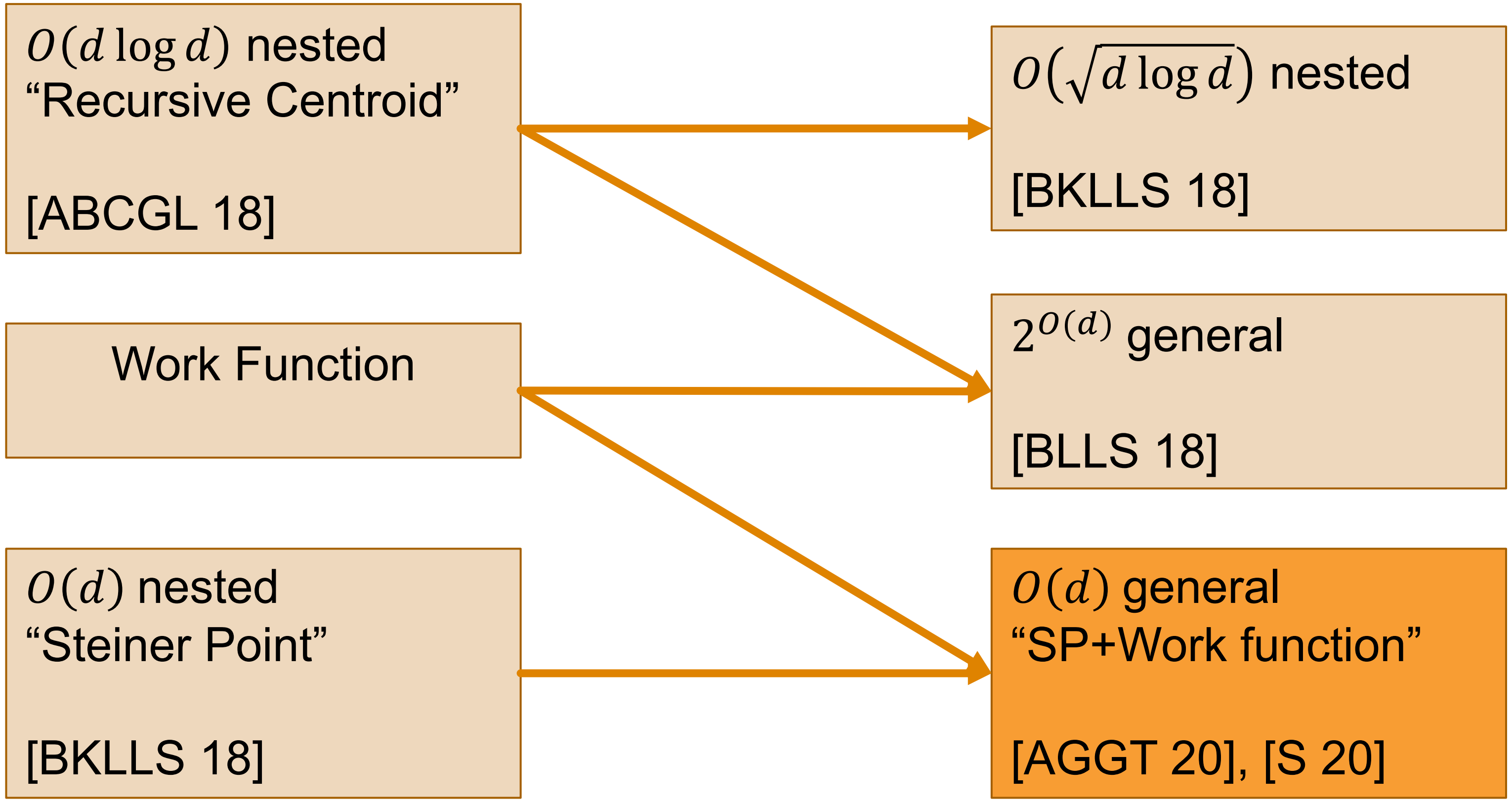
$$s_K(\theta) := \max_{x \in K} \langle \theta, x \rangle$$

Useful for the proof in this talk

$$= \lim_{\gamma \to \infty} cg(K + \gamma B)$$

$$B = B(0,1)$$

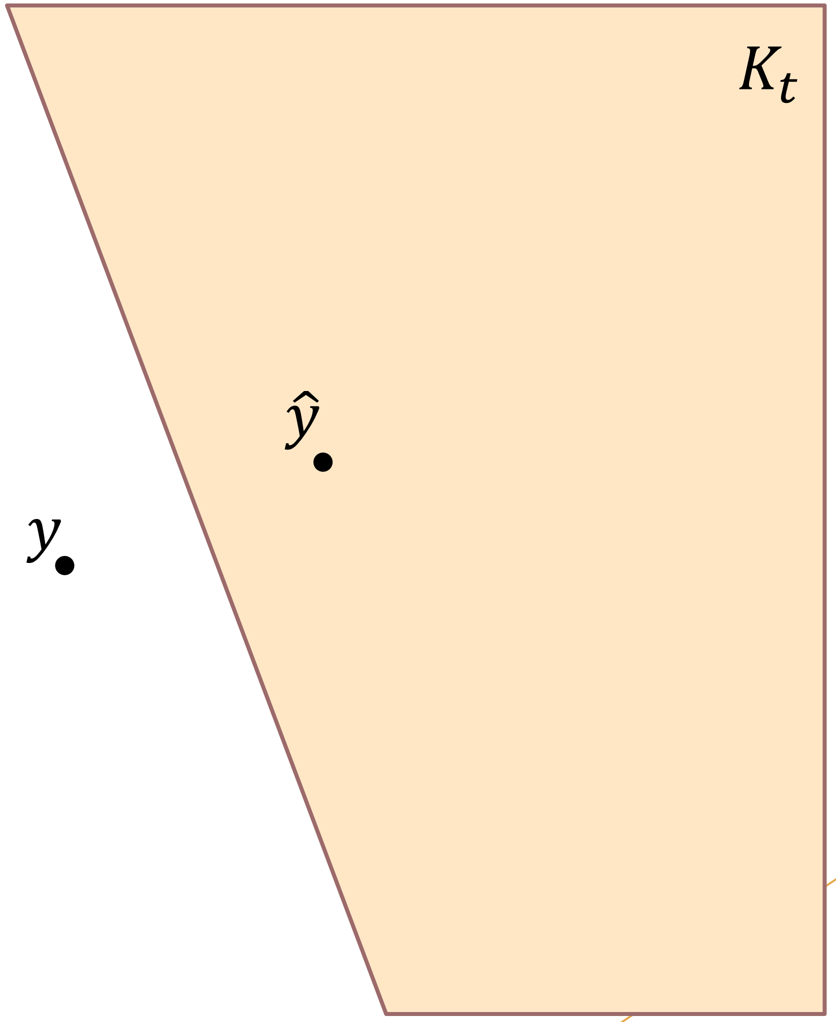
Progress


$O(d \log d)$ nested
"Recursive Centroid"
[ABCGL 18]
Work Function
$O(d)$ nested
"Steiner Point"
[BKLLS 18]
$O(\sqrt{d \log d})$ nested
[BKLLS 18]
$2^{O(d)}$ general
[BLLS 18]
$O(d)$ general
"SP+Work function"
[AGGT 20], [S 20]

Proof of Feasibility Lemma

- $K_t = \{\, x \mid \langle a, x \rangle \geq b \,\}$ (w.l.o.g.)
- Define

$$\hat{y} = \begin{cases} reflect(y) & y \notin K_t \\ y & y \in K_t \end{cases}$$



Goal: $st(\Omega_t) \in K_t$

$\Omega_t = \{x \mid w_t(x) \leq r\}$

$st(\Omega_t) = \lim_{\gamma \to \infty} cg(\Omega_t + \gamma B)$