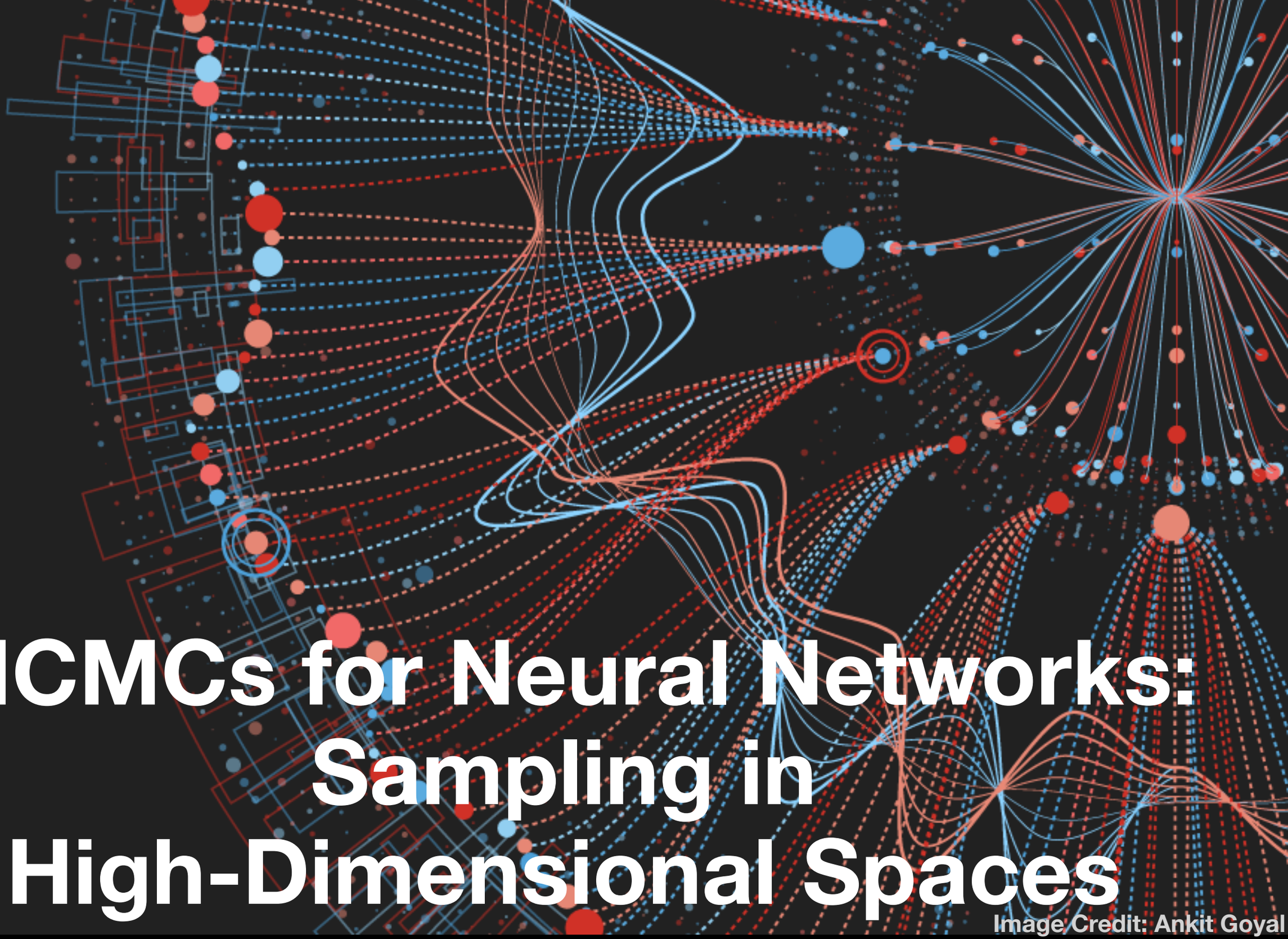


MCMCs for Neural Networks: Sampling in High-Dimensional Spaces

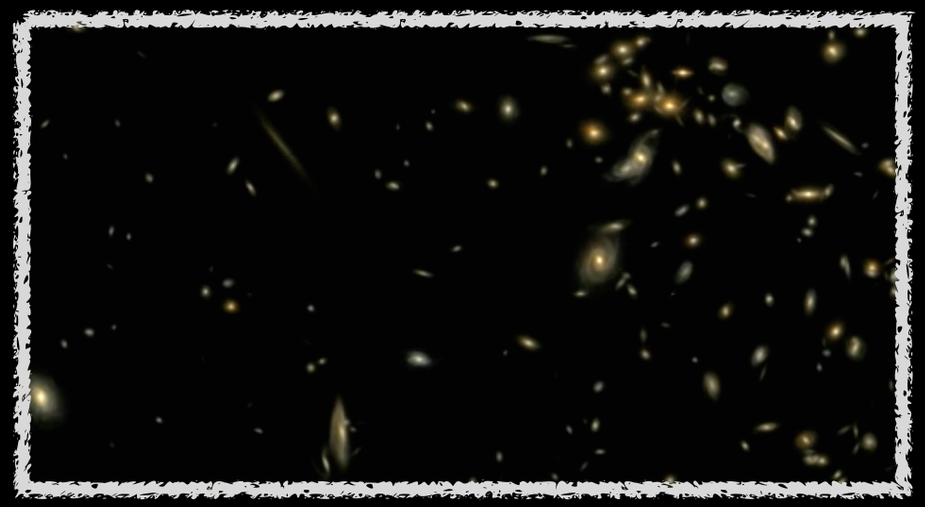
Image Credit: Ankit Goyal

Peter Behroozi (he/him)
U. Arizona + NAOJ

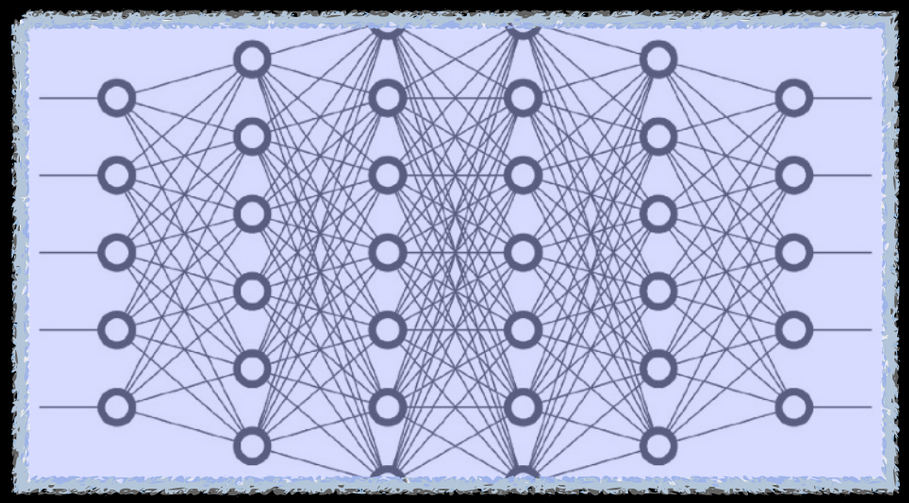
KITP-Galevo23

This talk will be very informal.

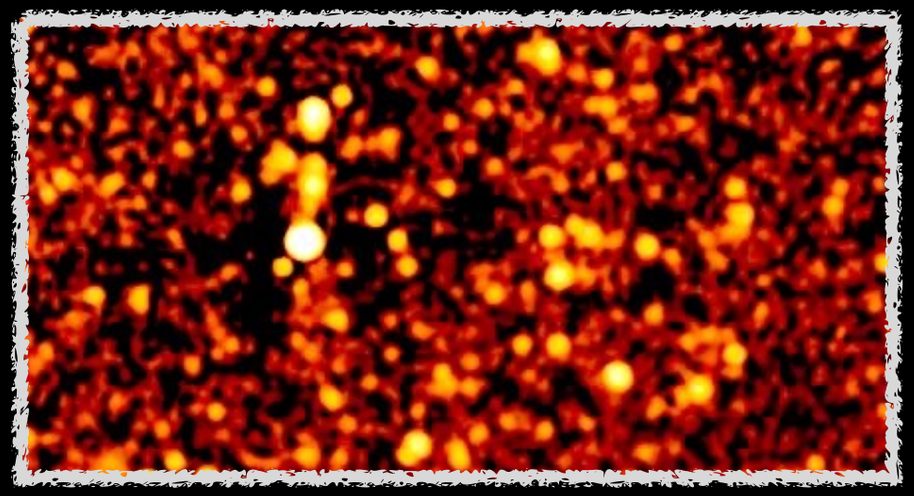
Motivation



Empirical Galaxy Modeling
50-100 dimensions



Deep Neural Networks
100-10^{14} dimensions



Confused Source Recovery
100 - 10^9 dimensions

Constrained Realizations
> 10^9 dimensions

Peter Behroozi

Image Credits: SDSS, Radu Raicea, Jim Condon, CLUES

MCMC Review

Observations

$$\mathcal{L}(x) = \exp(-\chi^2/2)$$

Parameter space → **MCMC**

Range of allowed parameters

MCMC Review

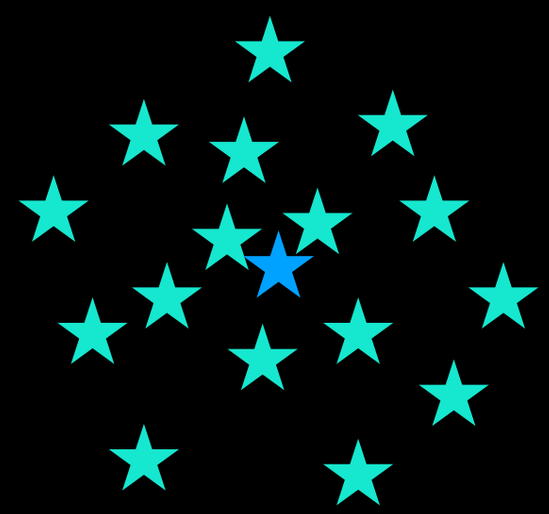
$$\pi(\mathbf{x} :\rightarrow \mathbf{y})$$

Transition Probability Density

MCMC Review

$$\pi(\mathbf{x} :\rightarrow \mathbf{y})$$

Transition Probability Density

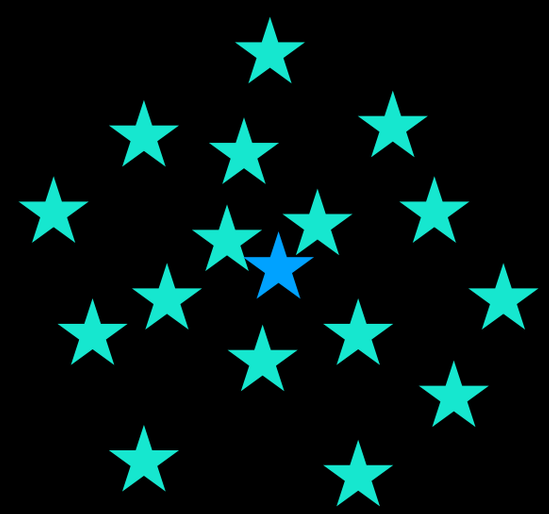


Gaussian Random Walk
(Metropolis-Hastings)

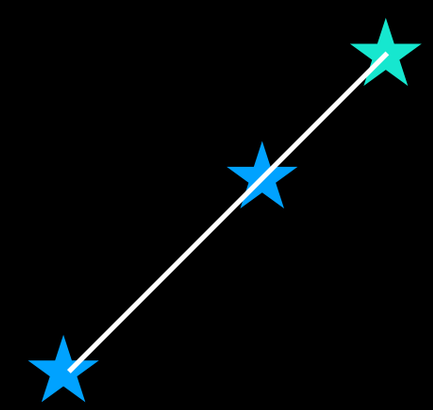
MCMC Review

$$\pi(\mathbf{x} :\rightarrow \mathbf{y})$$

Transition Probability Density



Gaussian Random Walk
(Metropolis-Hastings)

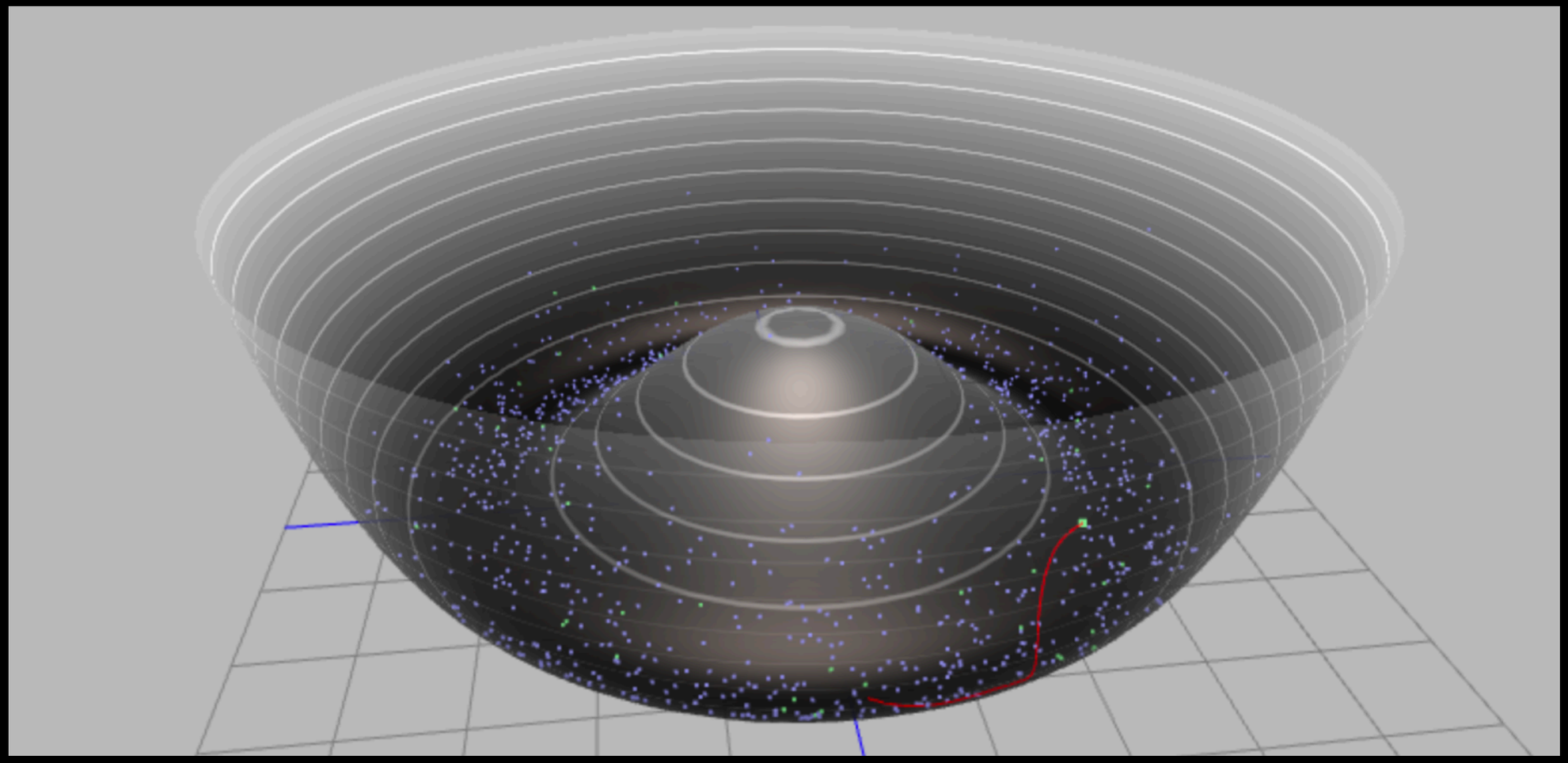


Stretch Step
(Goodman & Weare)

MCMC Review

$$\pi(\mathbf{x} :\rightarrow \mathbf{y})$$

Transition Probability Density



Hamiltonian Monte Carlo

https://arogozhnikov.github.io/2016/12/19/markov_chain_monte_carlo.html

High-D Spaces (D>10)

High-D Spaces (D>10)

$$P(\vec{x}) \propto \exp(-x^2)$$

Gaussian Distribution

High-D Spaces (D>10)

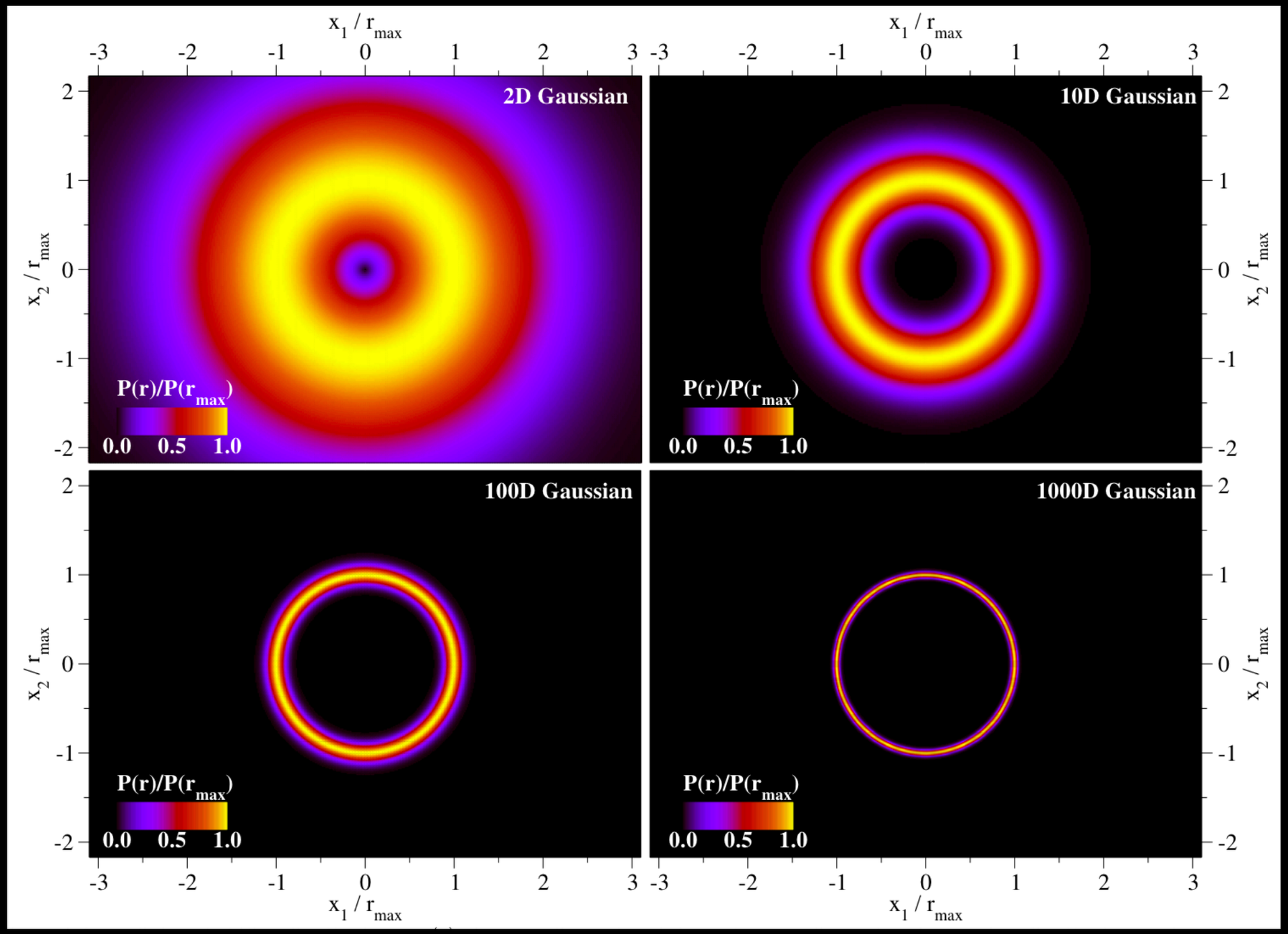
$$P(\vec{x}) \propto \exp(-x^2)$$

Gaussian Distribution

$$P(r) \propto r^{D-1} \exp(-r^2)$$

Distribution of Radii

High-D Spaces (D>10)



High-D Spaces (D>10)

Coffee

Lemon

Pizza

Credits: Julius Schorzman; https://food.ndtv.com/ingredient/lemon-701101; Valerio Capello

High-D Spaces (D>10)

Coffee

$\cdot$

Pizza

$=$

Credits: Julius Schorzman; https://food.ndtv.com/ingredient/lemon-701101; Valerio Capello

High-D Spaces (D>10)

Coffee

$\cdot$

Pizza

$=$

$$|\text{Coffee}||\text{Pizza}|\cos\theta$$

Credits: Julius Schorzman; https://food.ndtv.com/ingredient/lemon-701101; Valerio Capello

High-D Spaces (D>10)

Coffee

$\cdot$

Pizza

$=$

$$|\text{Coffee}||\text{Pizza}|\cos\theta \sim 0$$

Credits: Julius Schorzman; https://food.ndtv.com/ingredient/lemon-701101; Valerio Capello

High-D Spaces (D>10)

Coffee

$\cdot$

Pizza

$=$

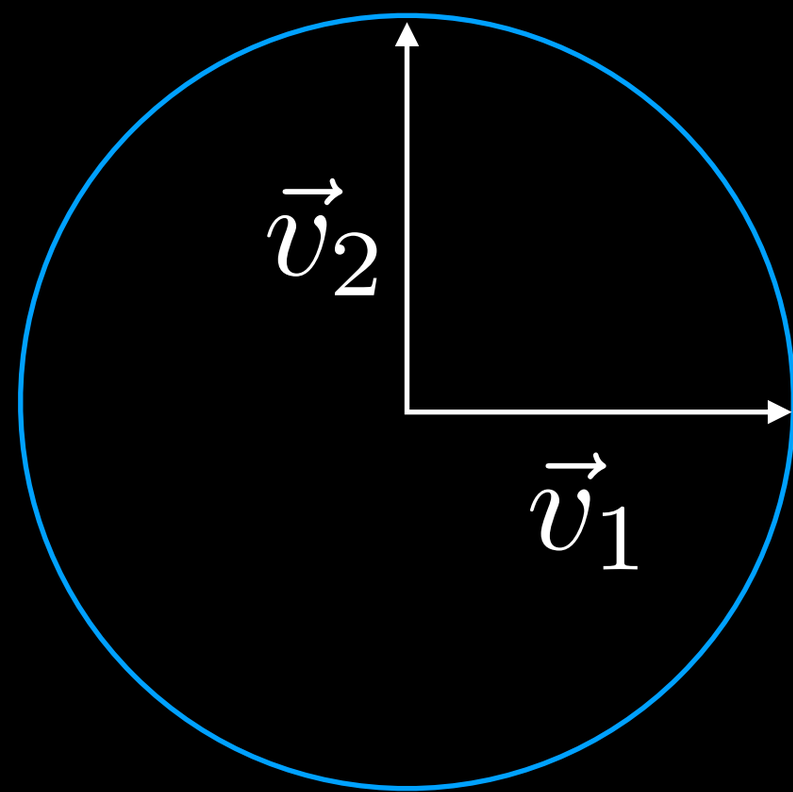
$$|\text{Coffee}||\text{Pizza}|\cos\theta \sim 0$$

$$\langle(\hat{v}_1 \cdot \hat{v}_2)^2\rangle = \frac{1}{D}$$

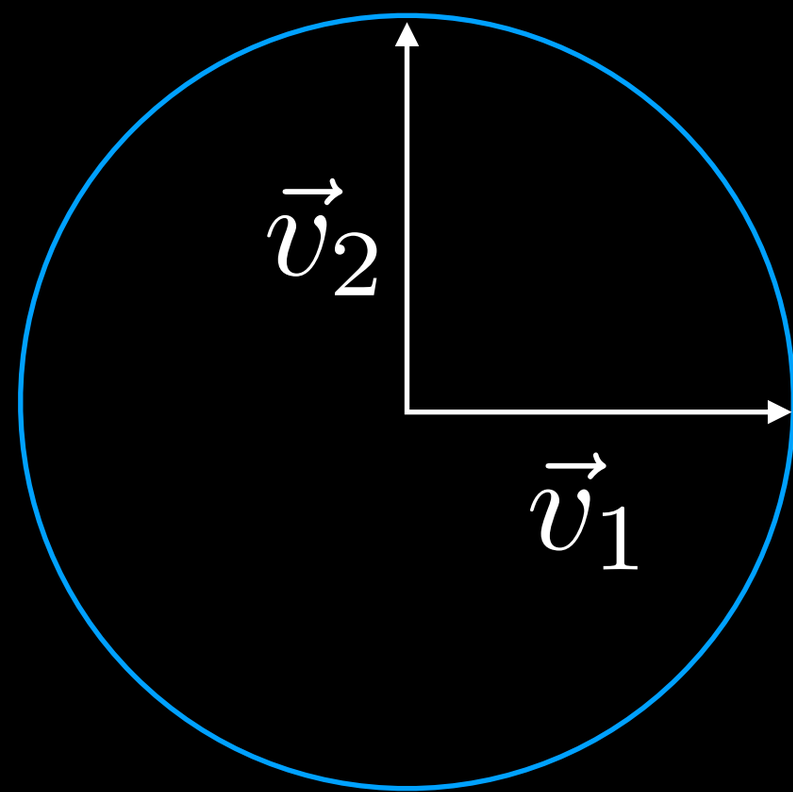
Random vectors in high dimensional spaces are orthogonal!

Credits: Julius Schorzman; https://food.ndtv.com/ingredient/lemon-701101; Valerio Capello

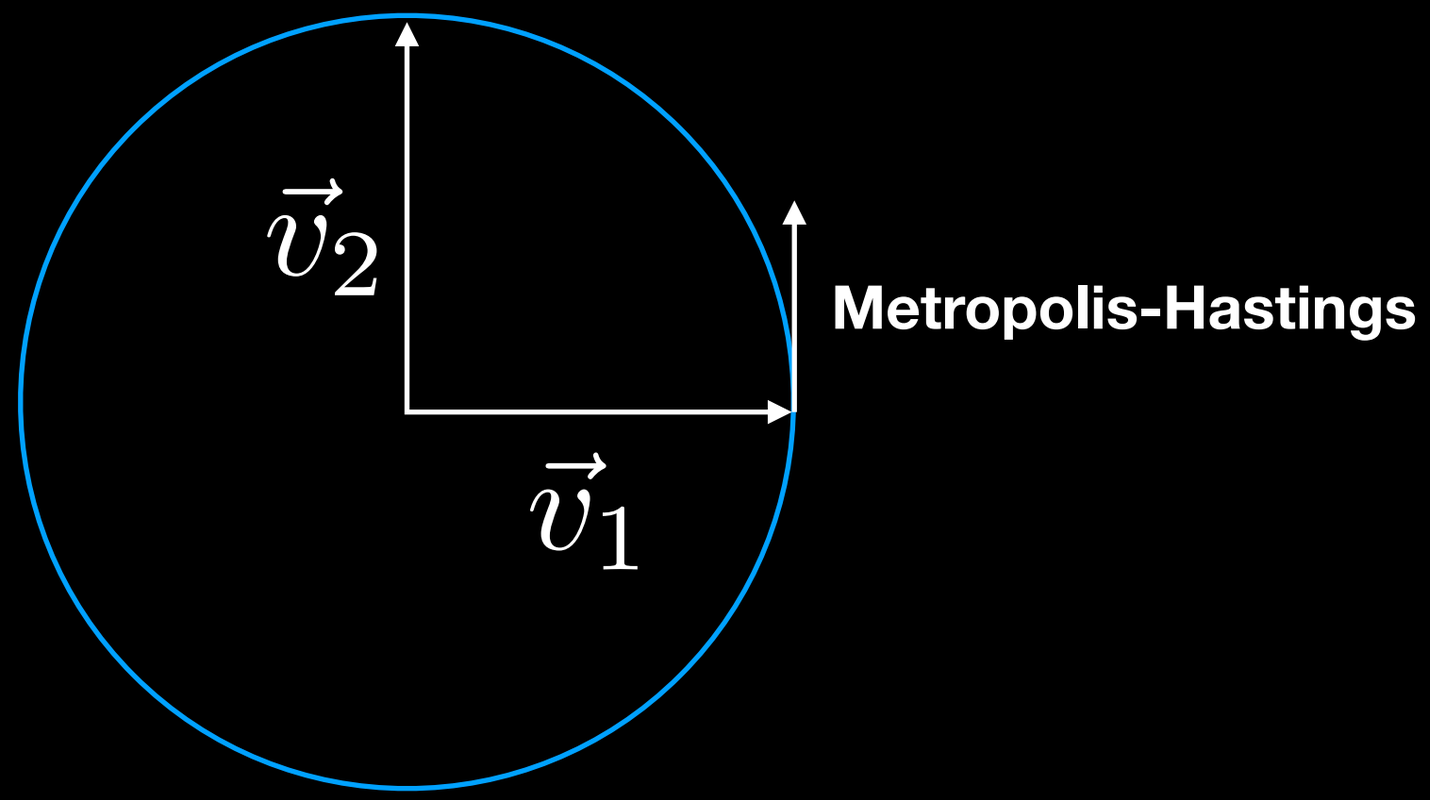
High-D Spaces



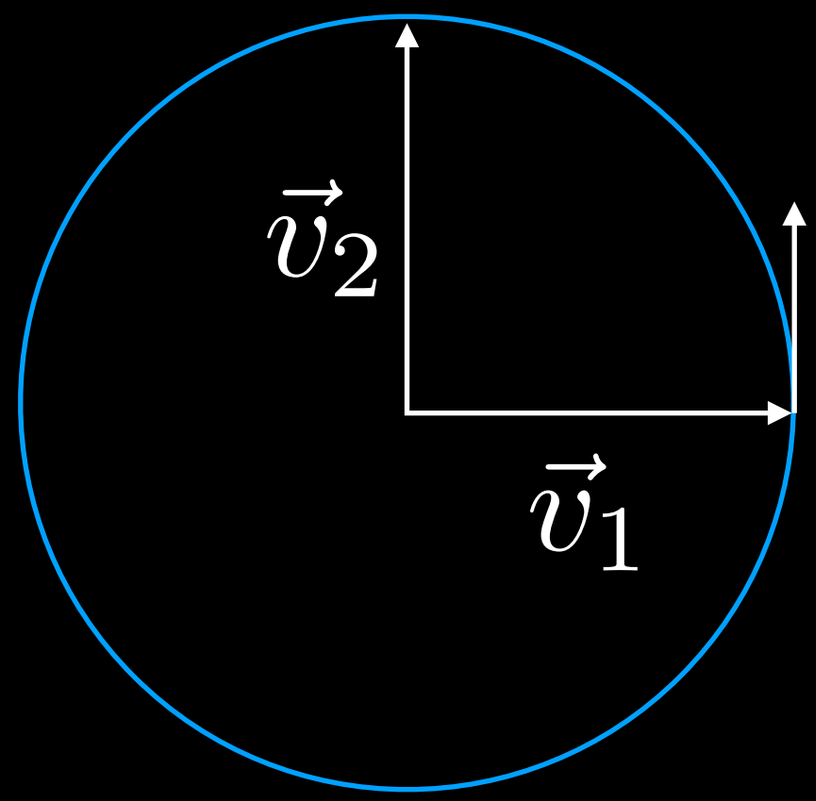
Why is it so hard?



Why is it so hard?



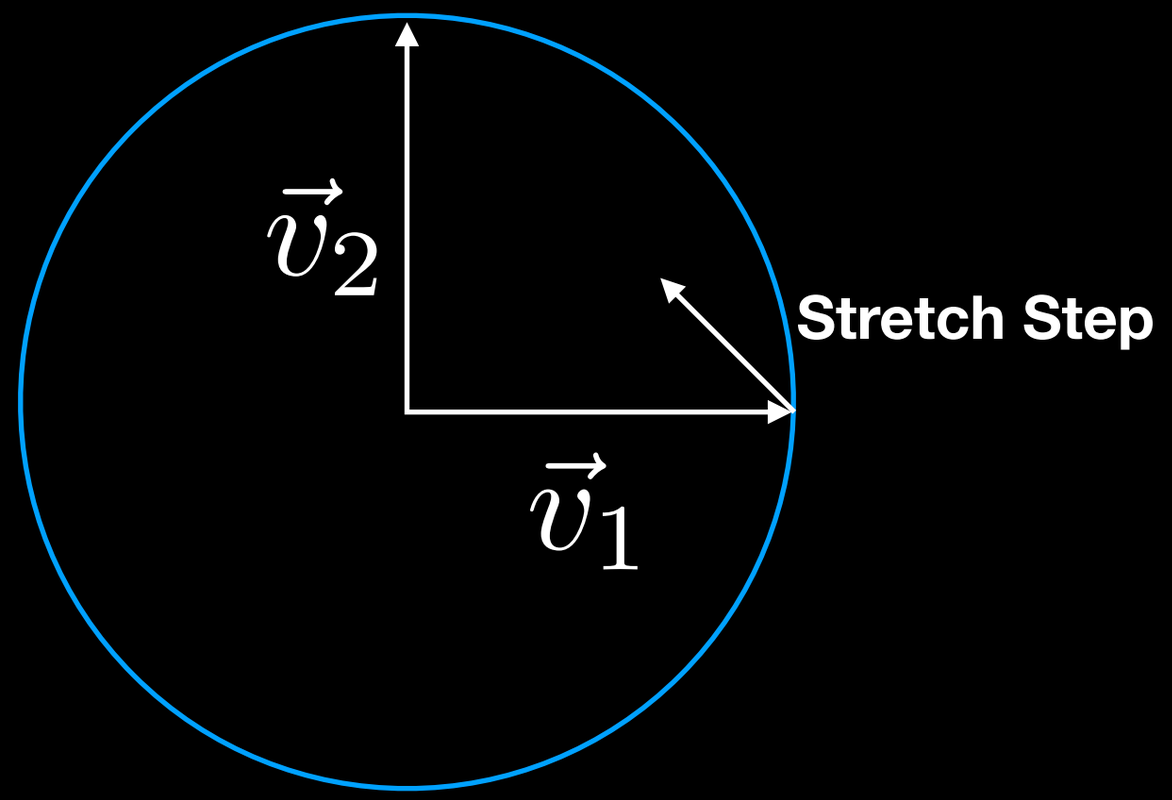
Why is it so hard?



Metropolis-Hastings

$\mathcal{O}(D)$

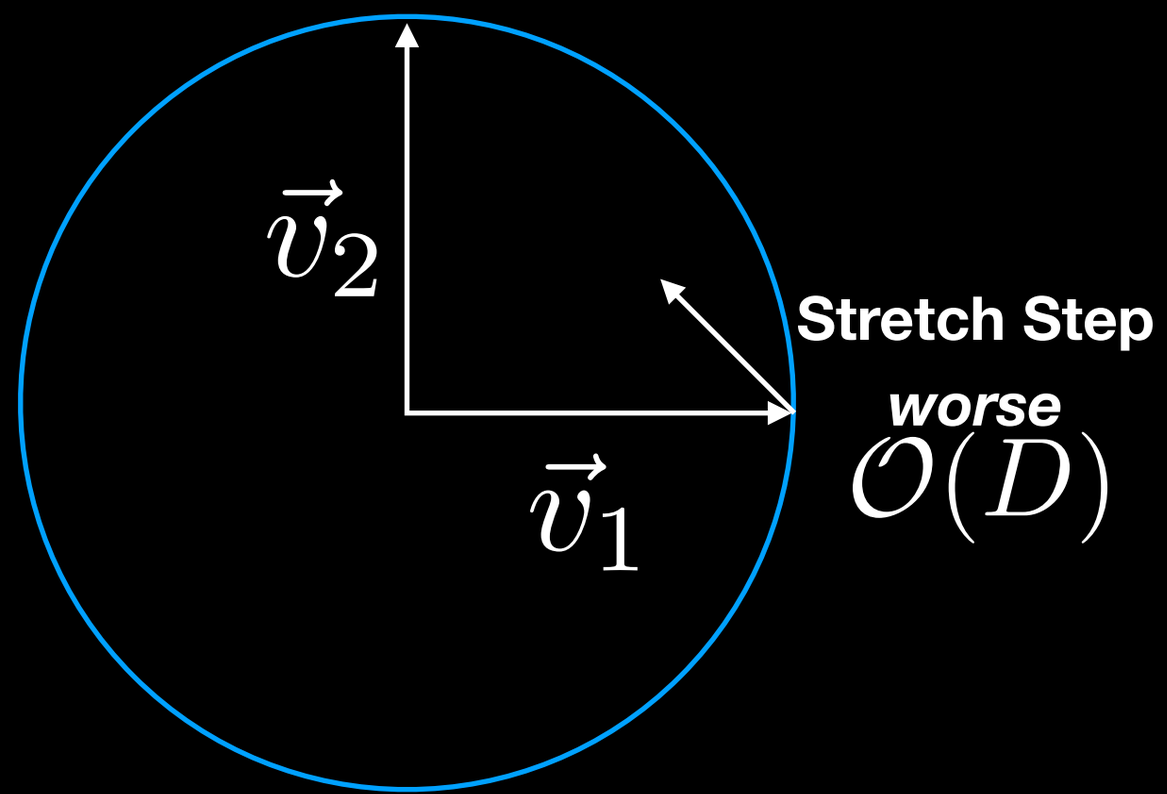
Why is it so hard?



Metropolis-Hastings

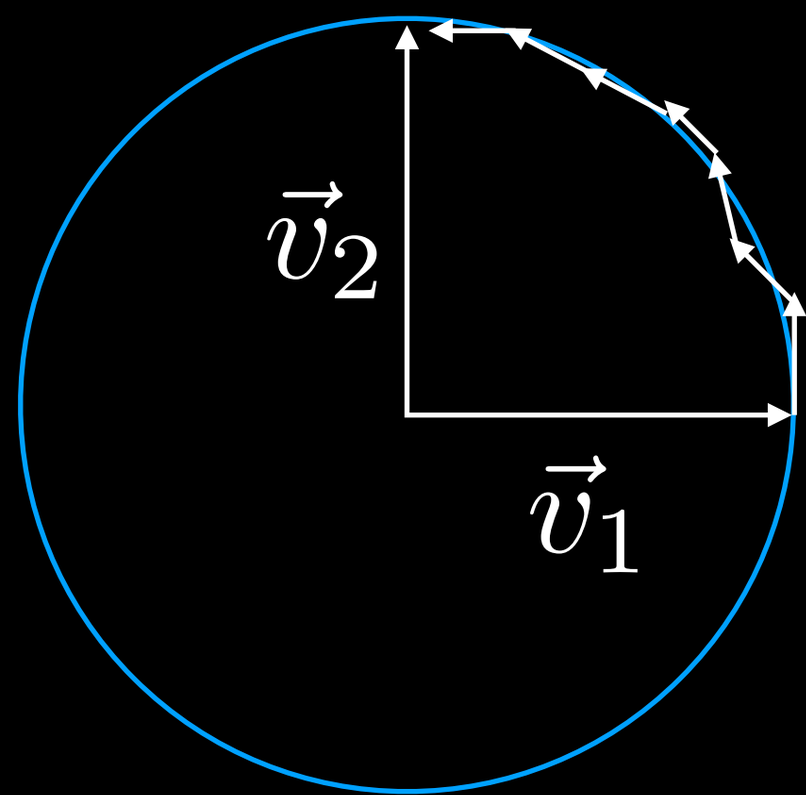
$$\mathcal{O}(D)$$

Why is it so hard?



Metropolis-Hastings
$\mathcal{O}(D)$

Why is it so hard?



Hamiltonian MC

Metropolis-Hastings

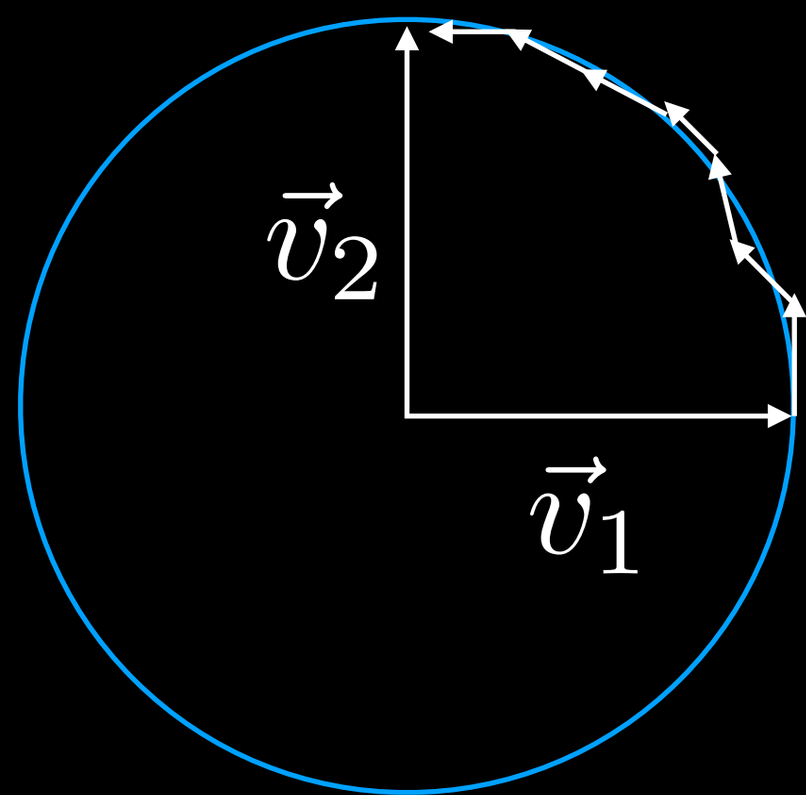
$$\mathcal{O}(D)$$

Stretch Step

worse

$$\mathcal{O}(D)$$

Why is it so hard?



Hamiltonian MC

$$\mathcal{O}(D^{1/4}) \cdot T(\vec{\nabla})$$

Metropolis-Hastings

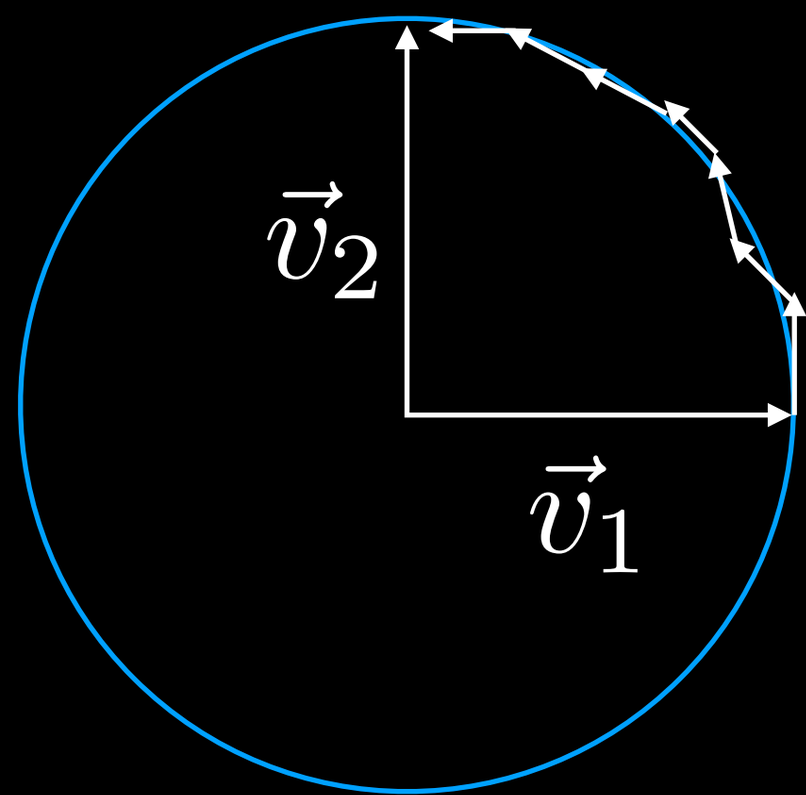
$$\mathcal{O}(D)$$

Stretch Step

worse

$$\mathcal{O}(D)$$

Why is it so hard?



Metropolis-Hastings

$$\mathcal{O}(D)$$

Stretch Step

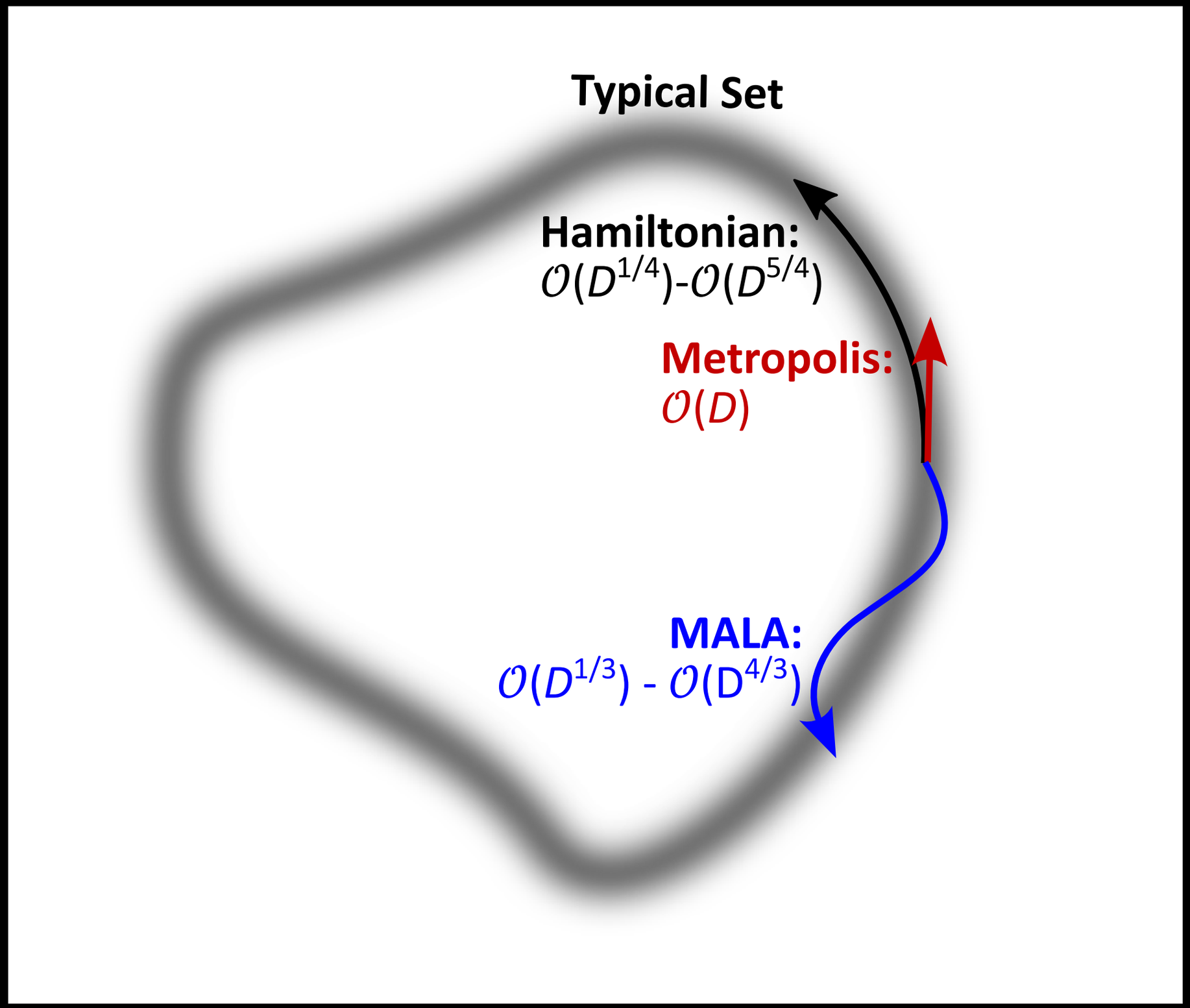
worse

$$\mathcal{O}(D)$$

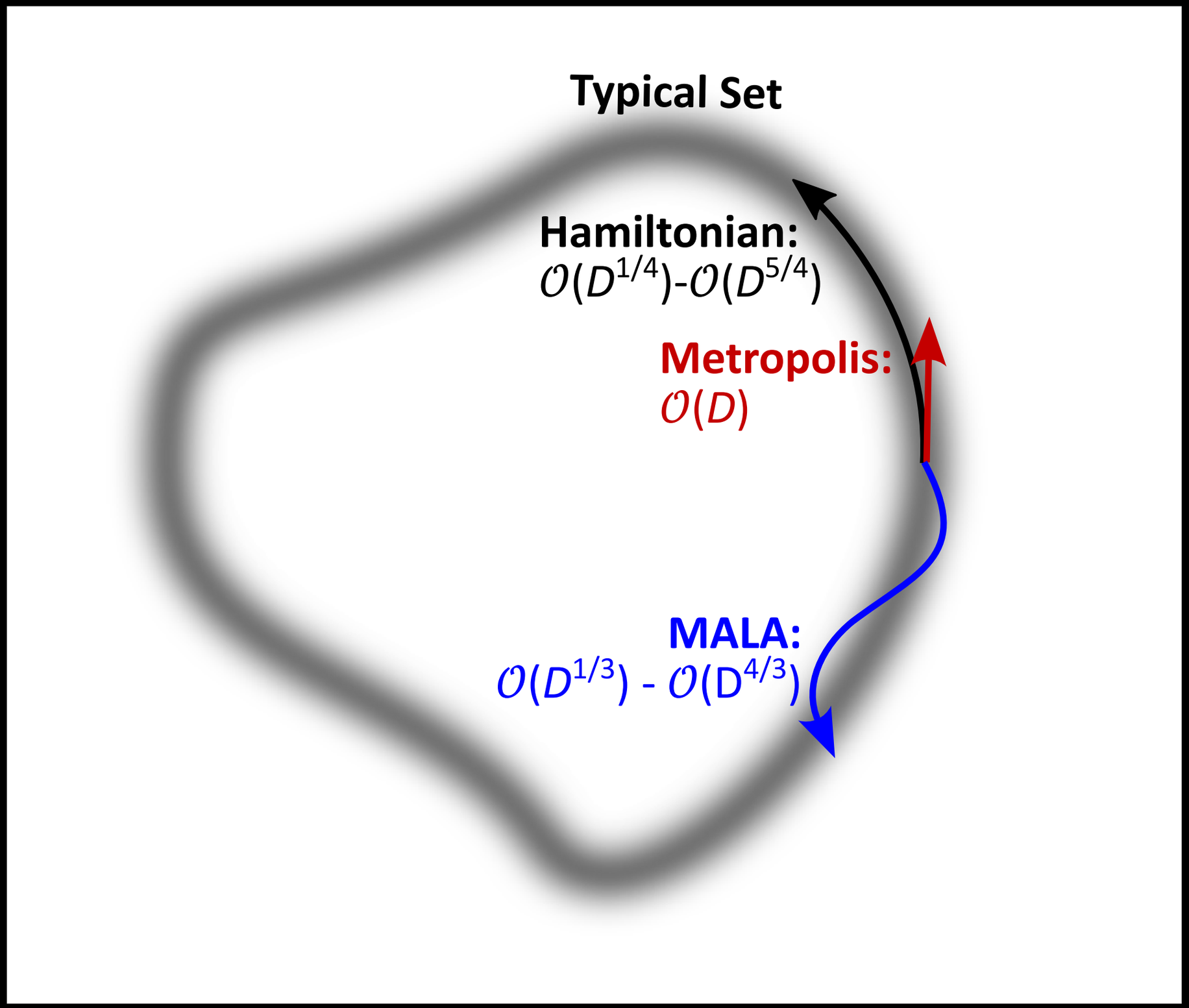
Hamiltonian MC

$$\mathcal{O}(D^{1/4}) - \mathcal{O}(D^{5/4})$$

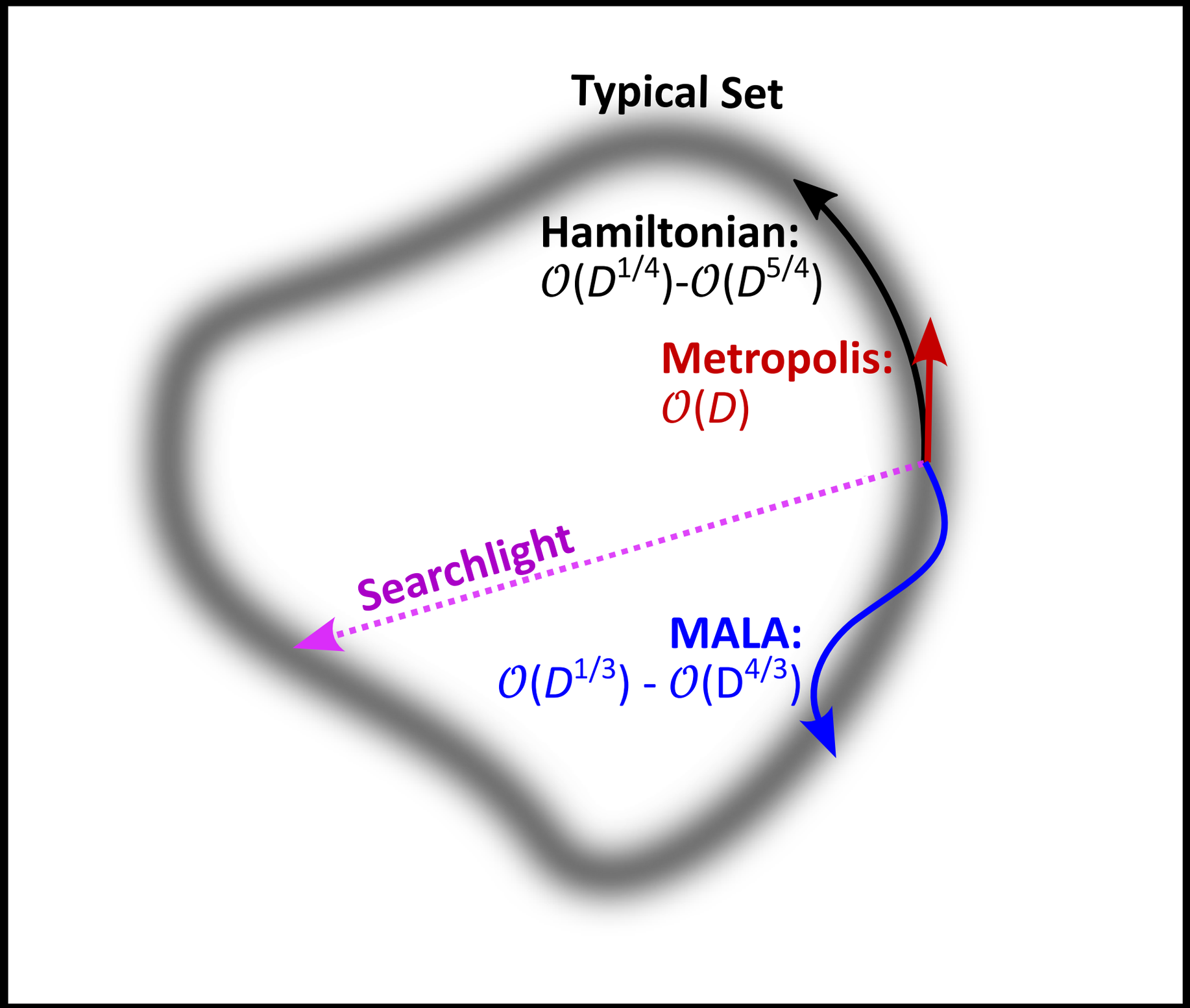
Why is it so hard?



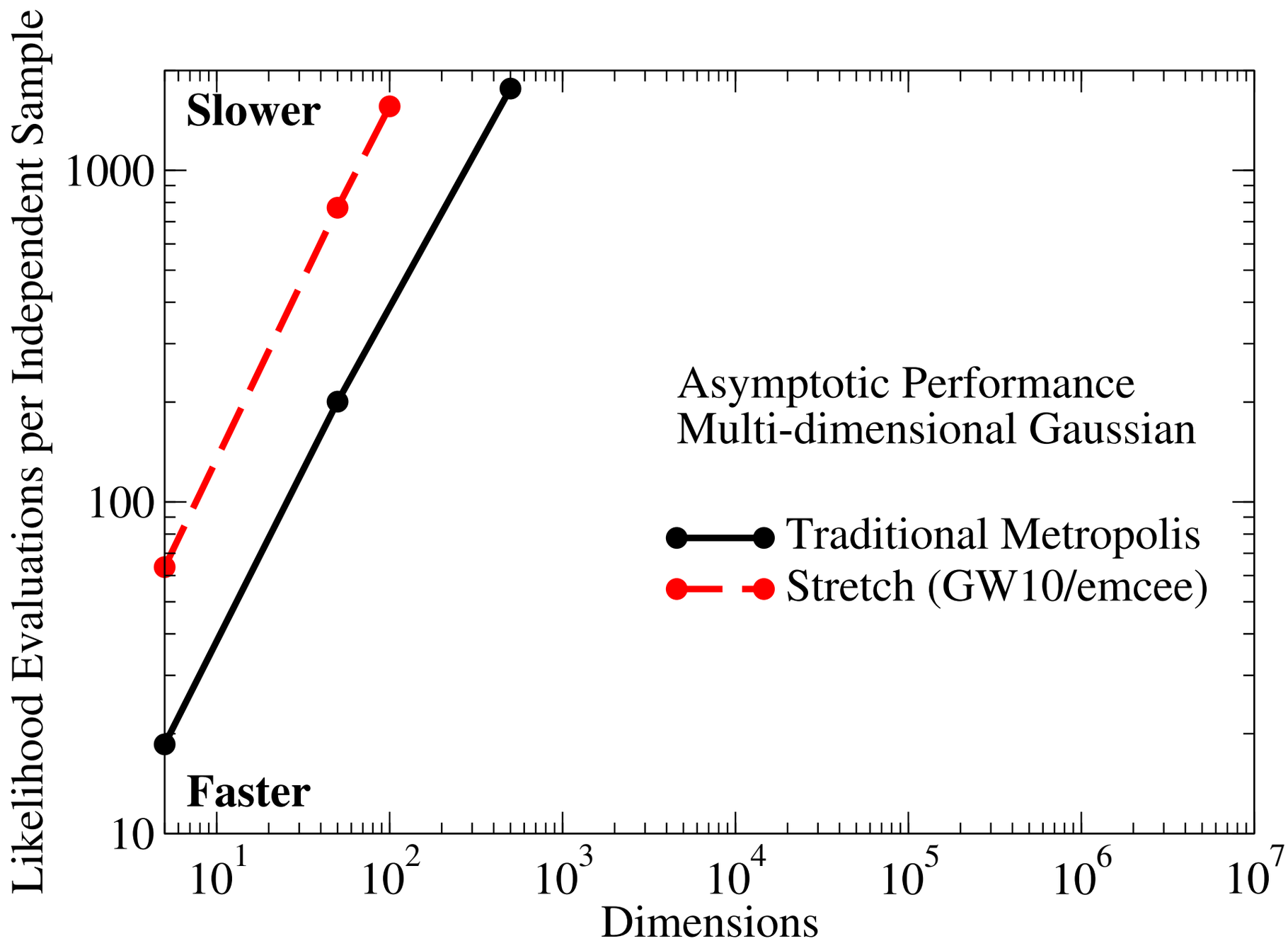
Should it be so hard?



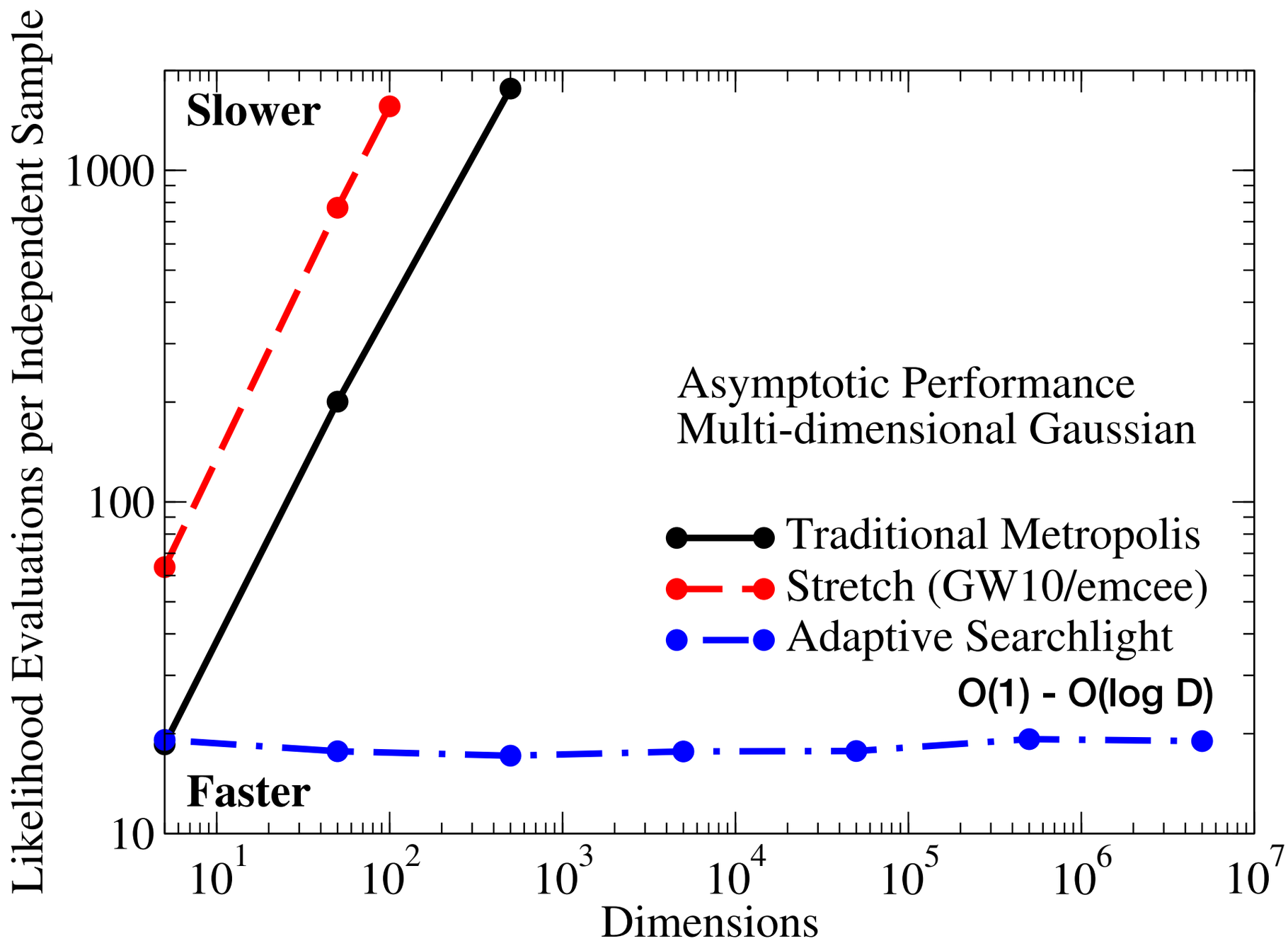
Should it be so hard?



Should it be so hard?

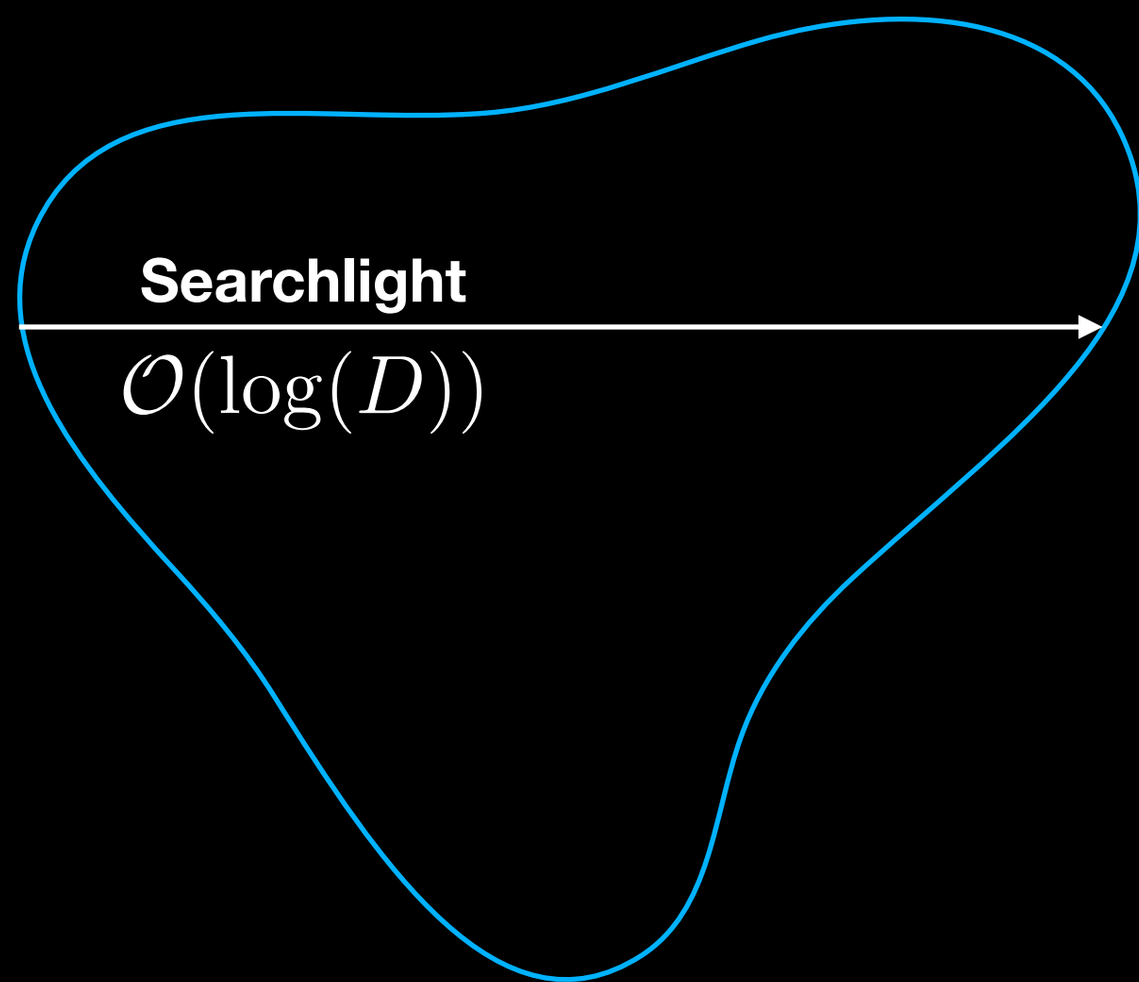


Should it be so hard?



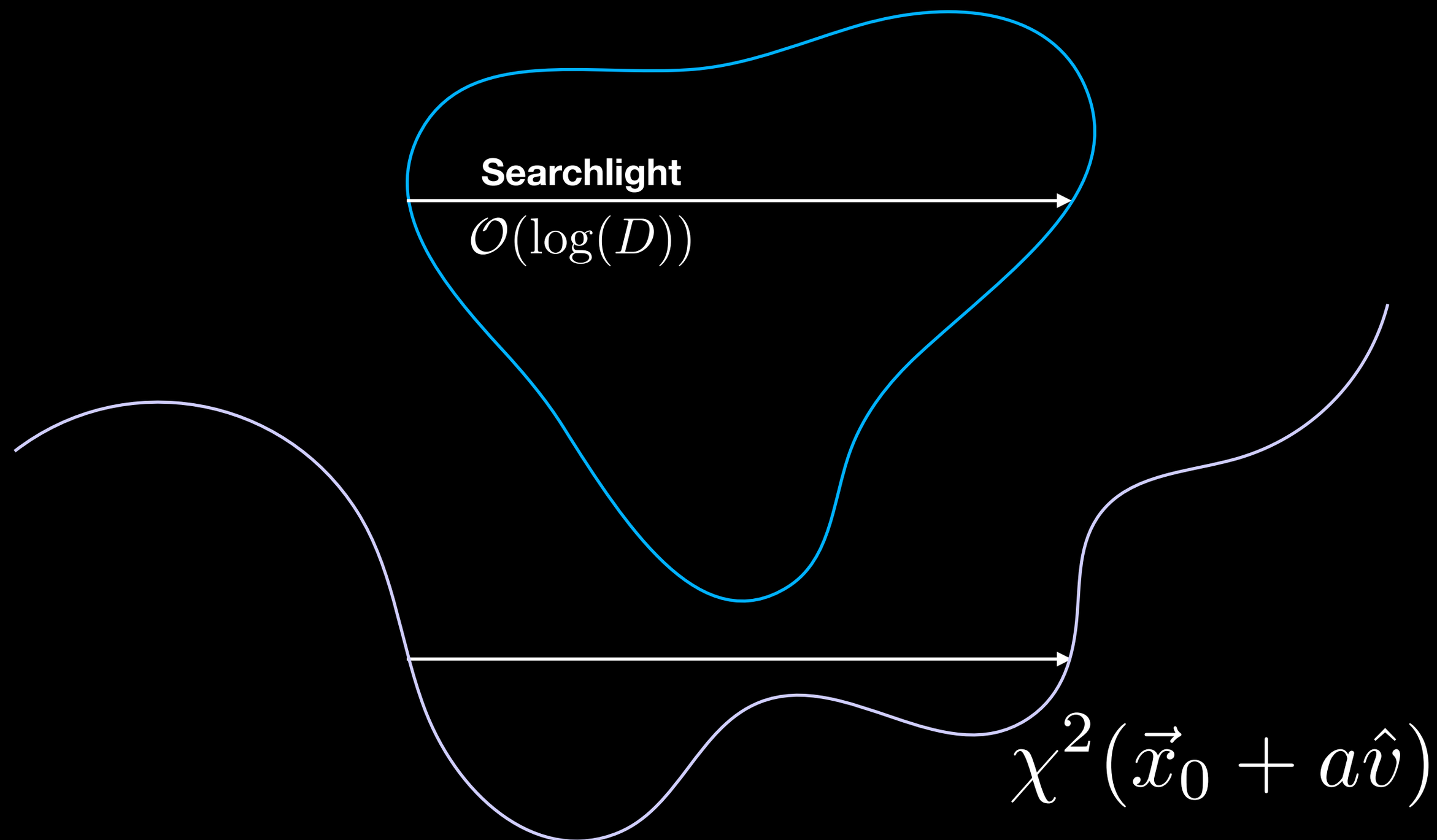
Should it be so hard?

Why not combine MCMCs with search algorithms?



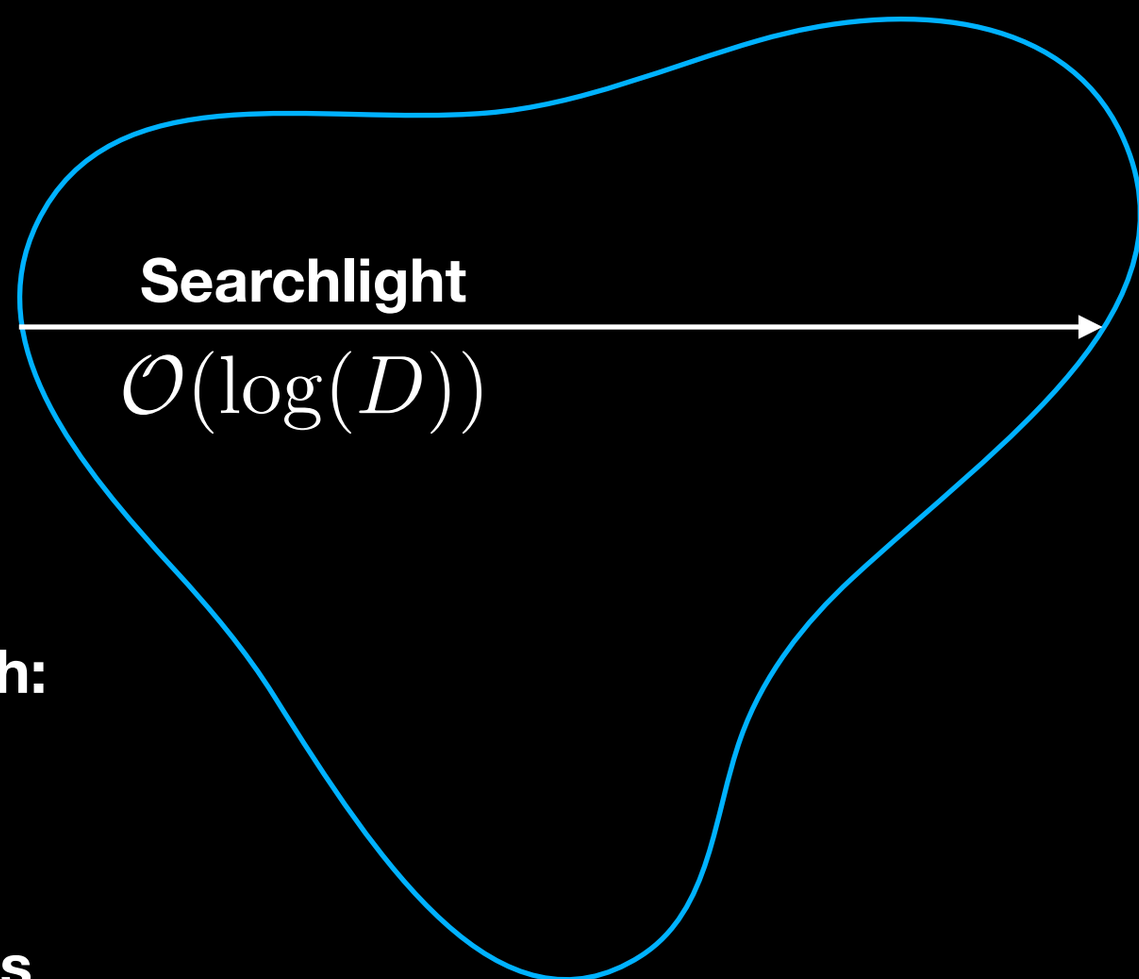
Should it be so hard?

Why not combine MCMCs with search algorithms?



Should it be so hard?

Why not combine MCMCs with search algorithms?



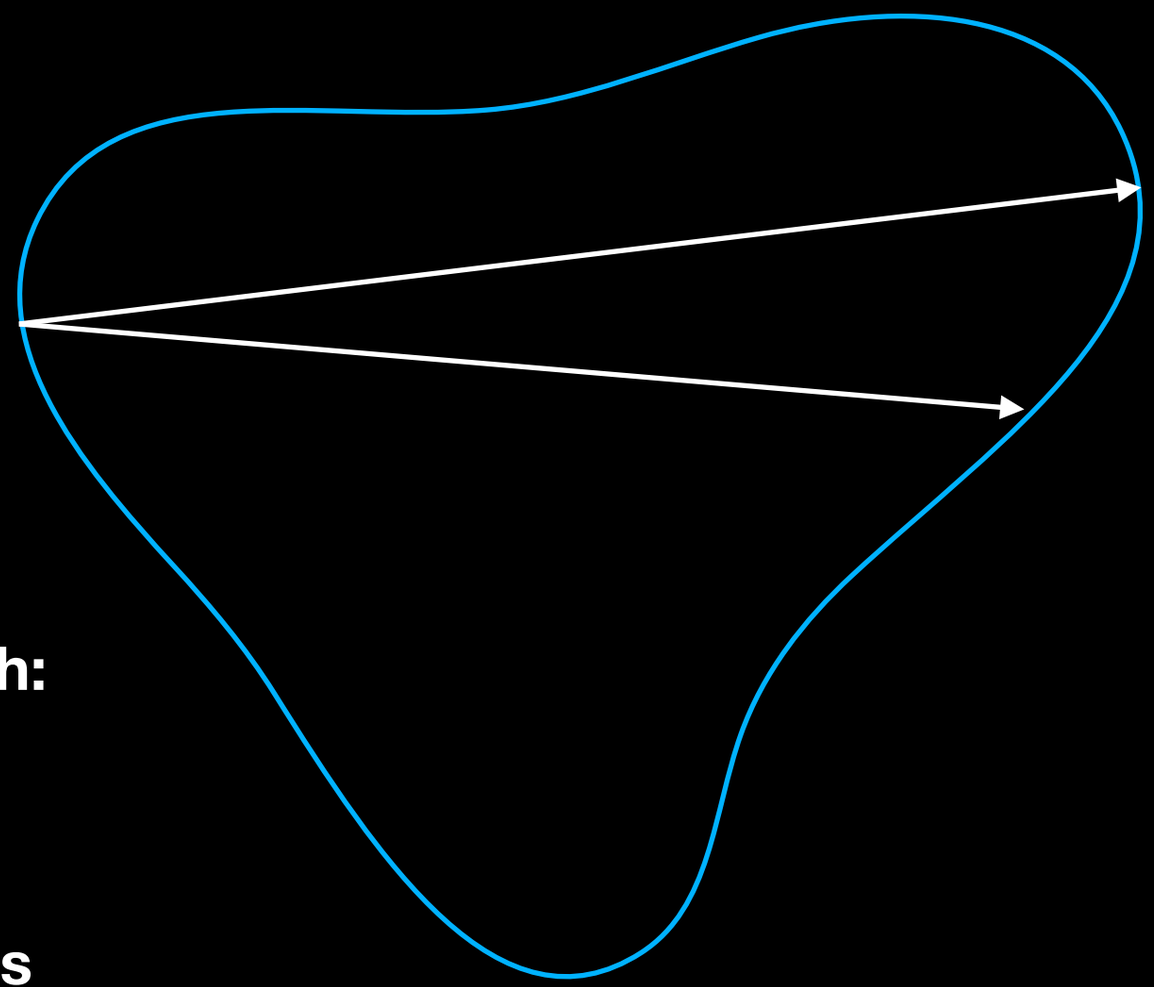
Metropolis / Stretch:
Diffusion

Hamiltonian:
Gravitational Orbits

Searchlight:
Optical Paths

Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



Metropolis / Stretch: Diffusion

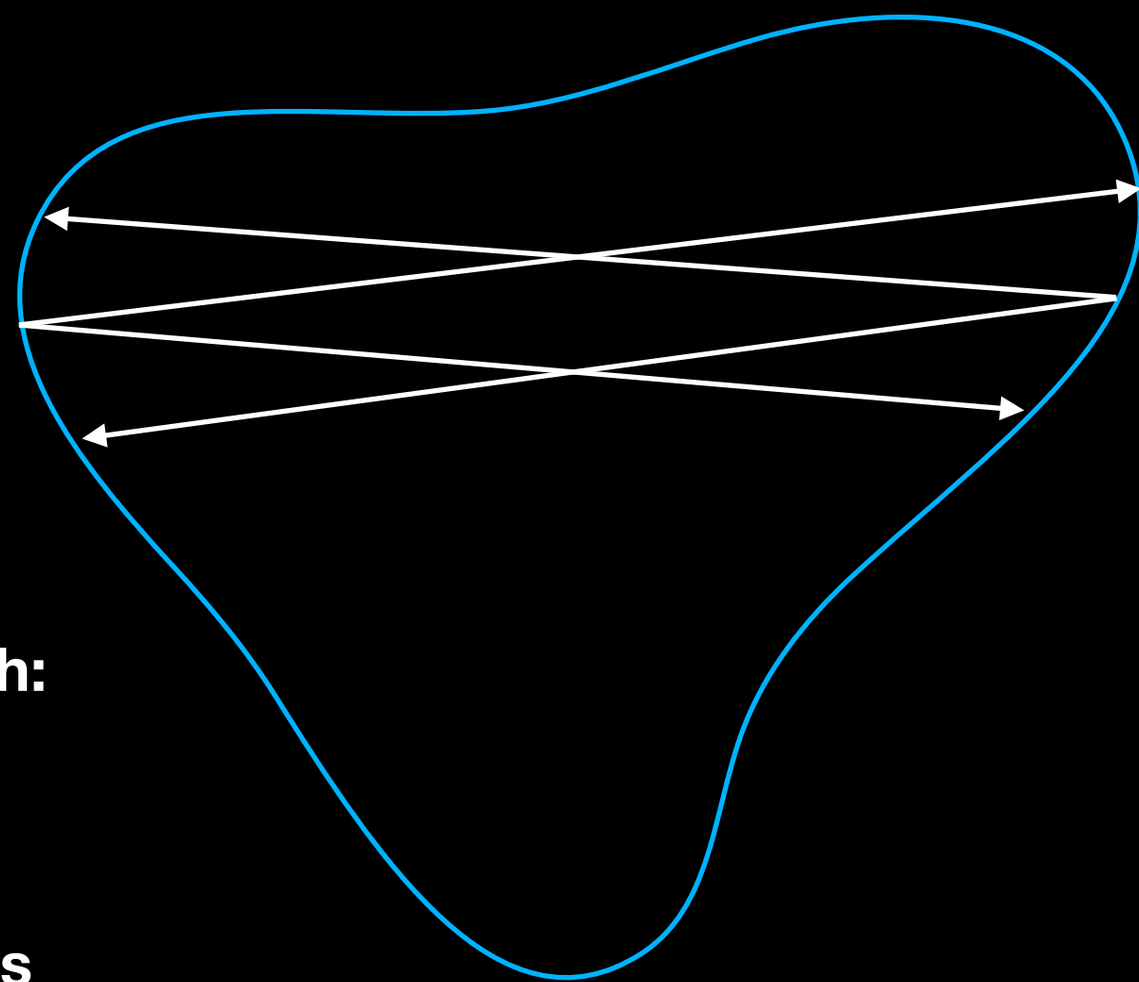
Hamiltonian: Gravitational Orbits

Searchlight: Optical Paths

$$L(r) \propto r^{1-D}$$

Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



Metropolis / Stretch: Diffusion

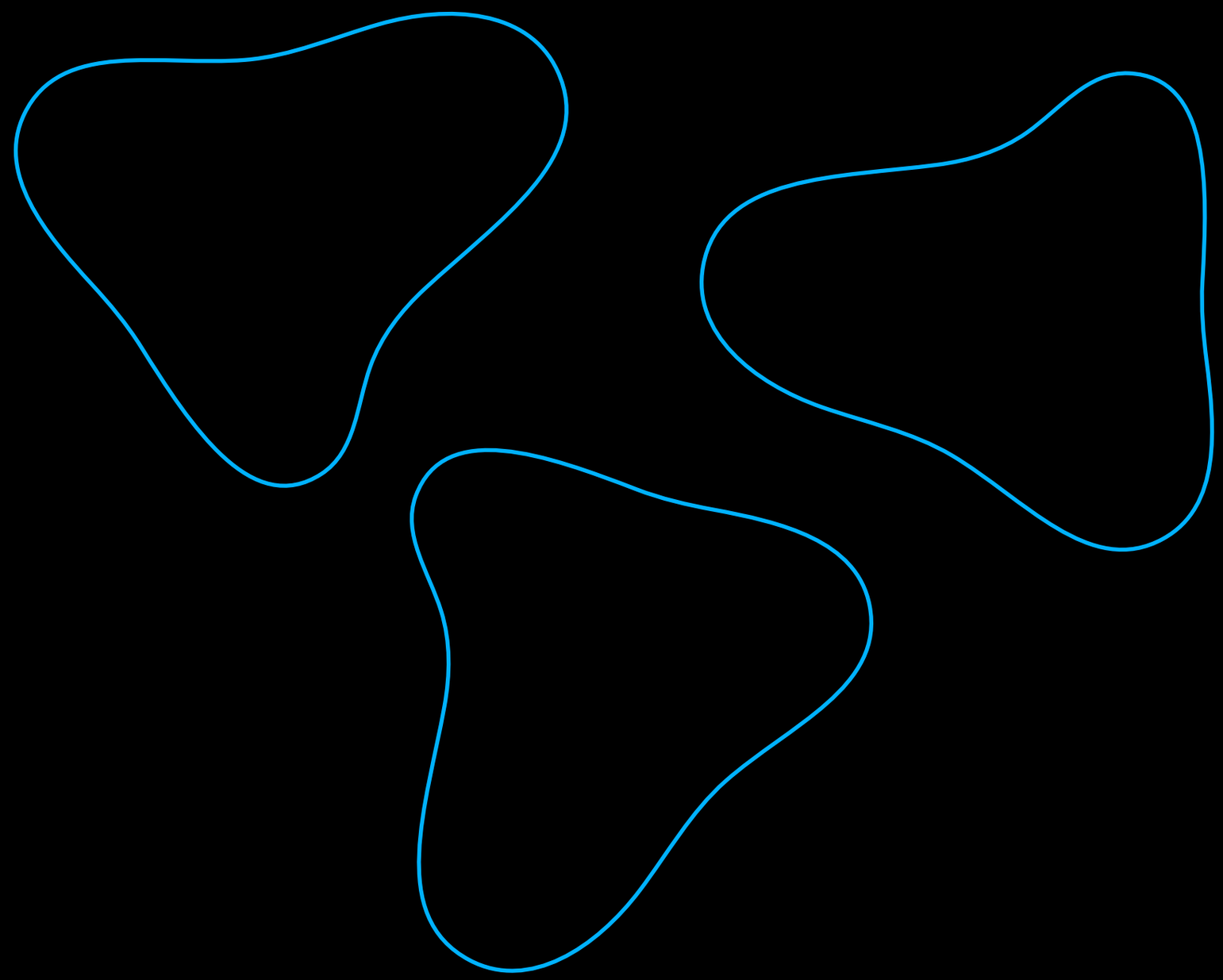
Hamiltonian: Gravitational Orbits

Searchlight: Optical Paths

$$L(r) \propto r^{1-D}$$

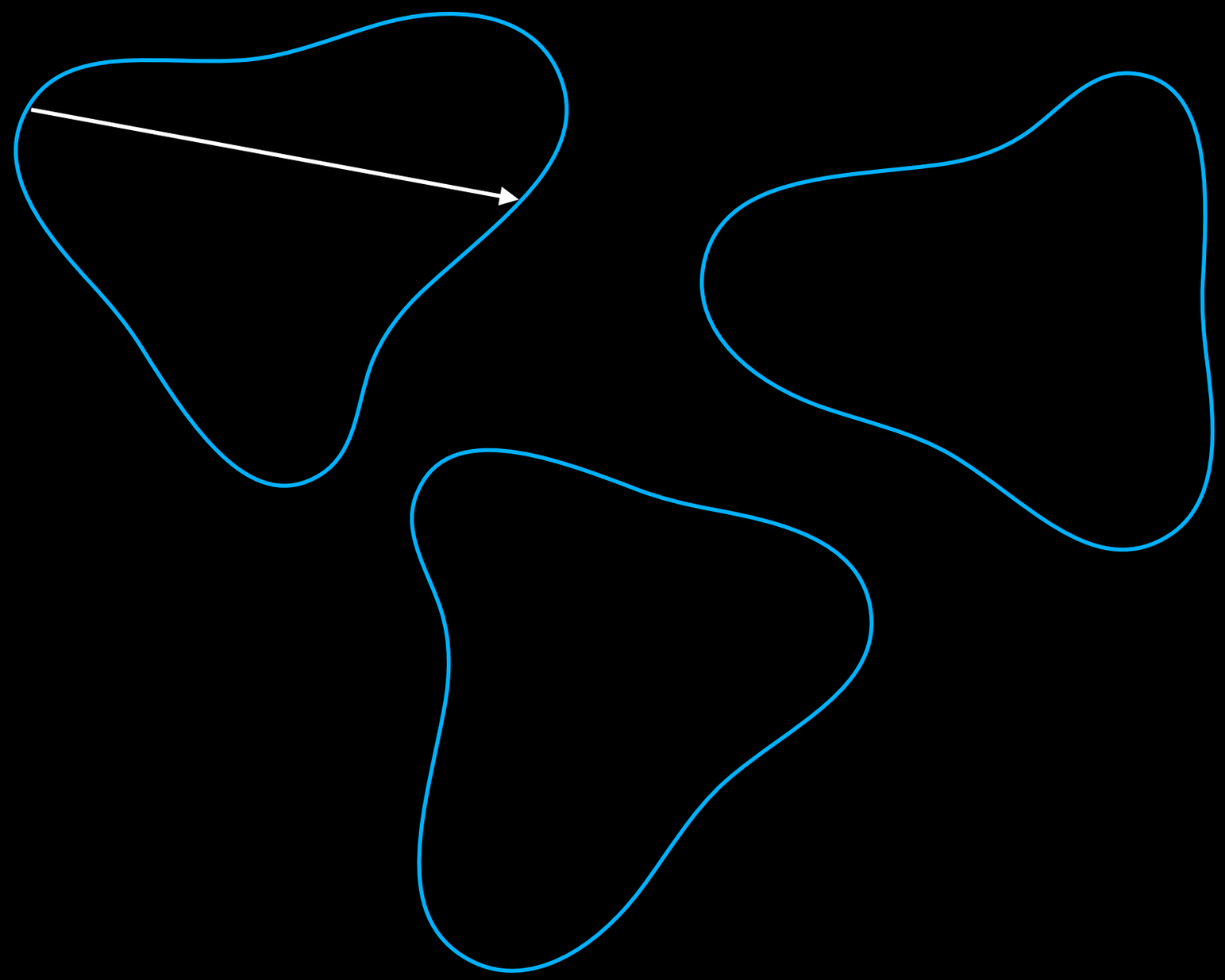
Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



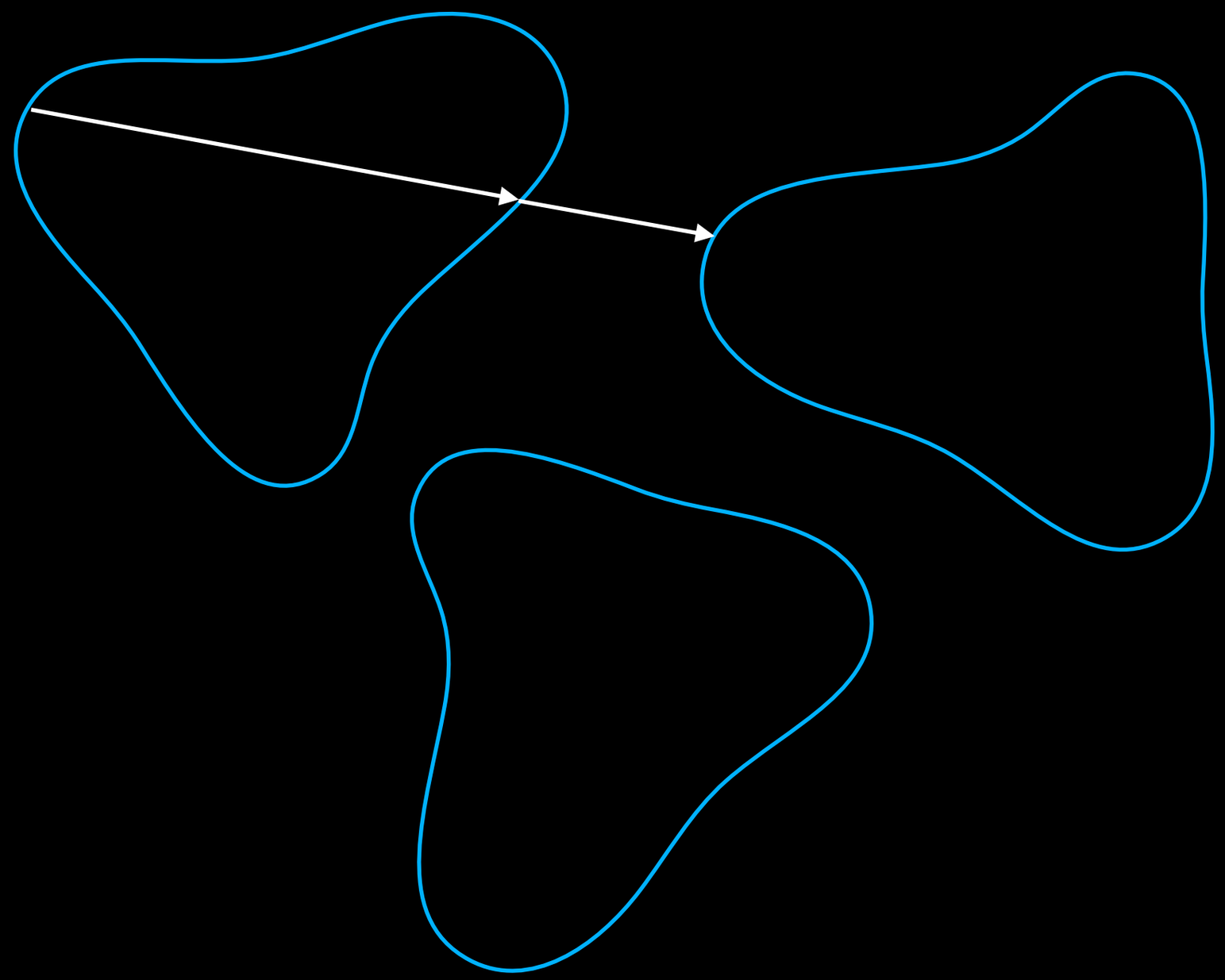
Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



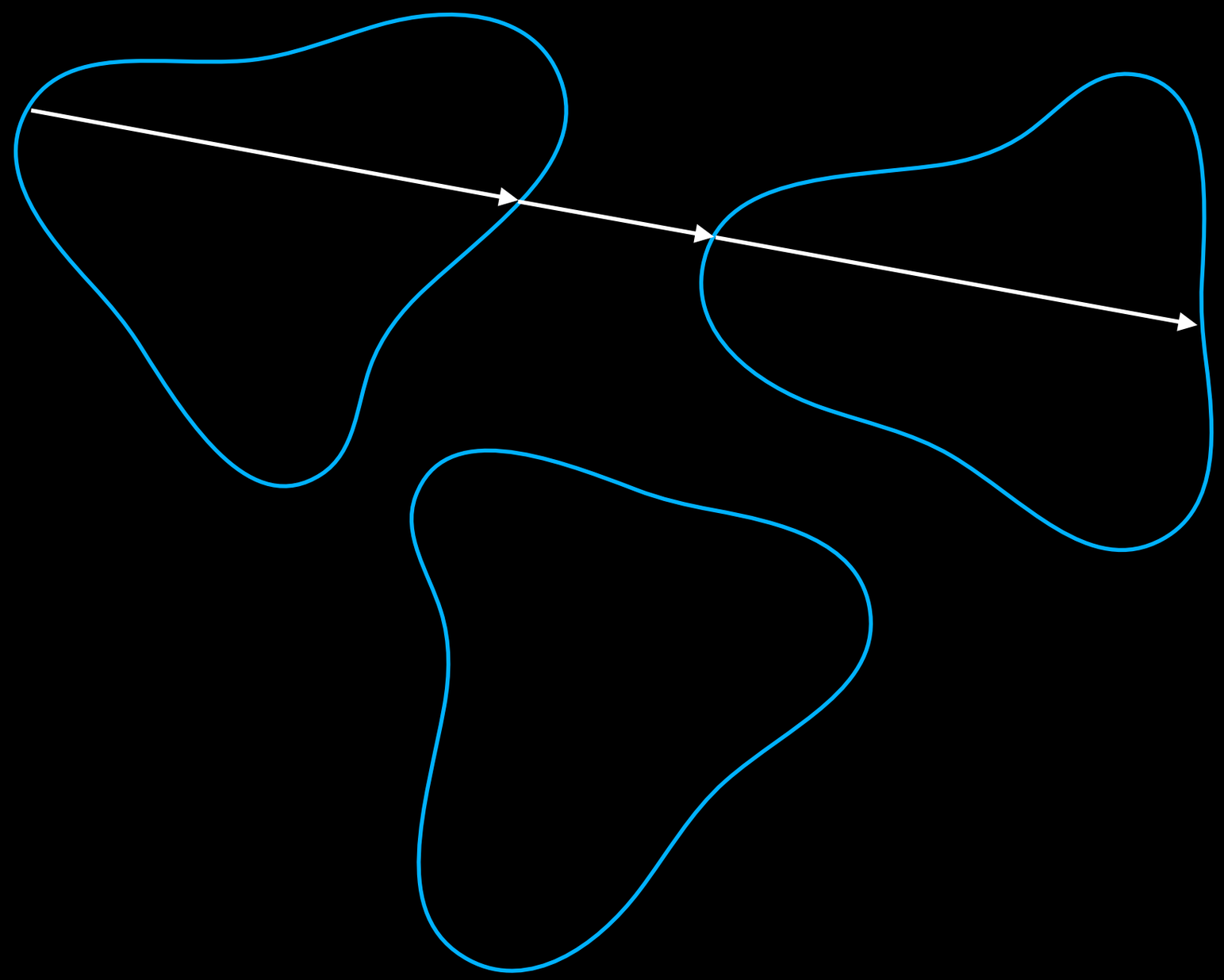
Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



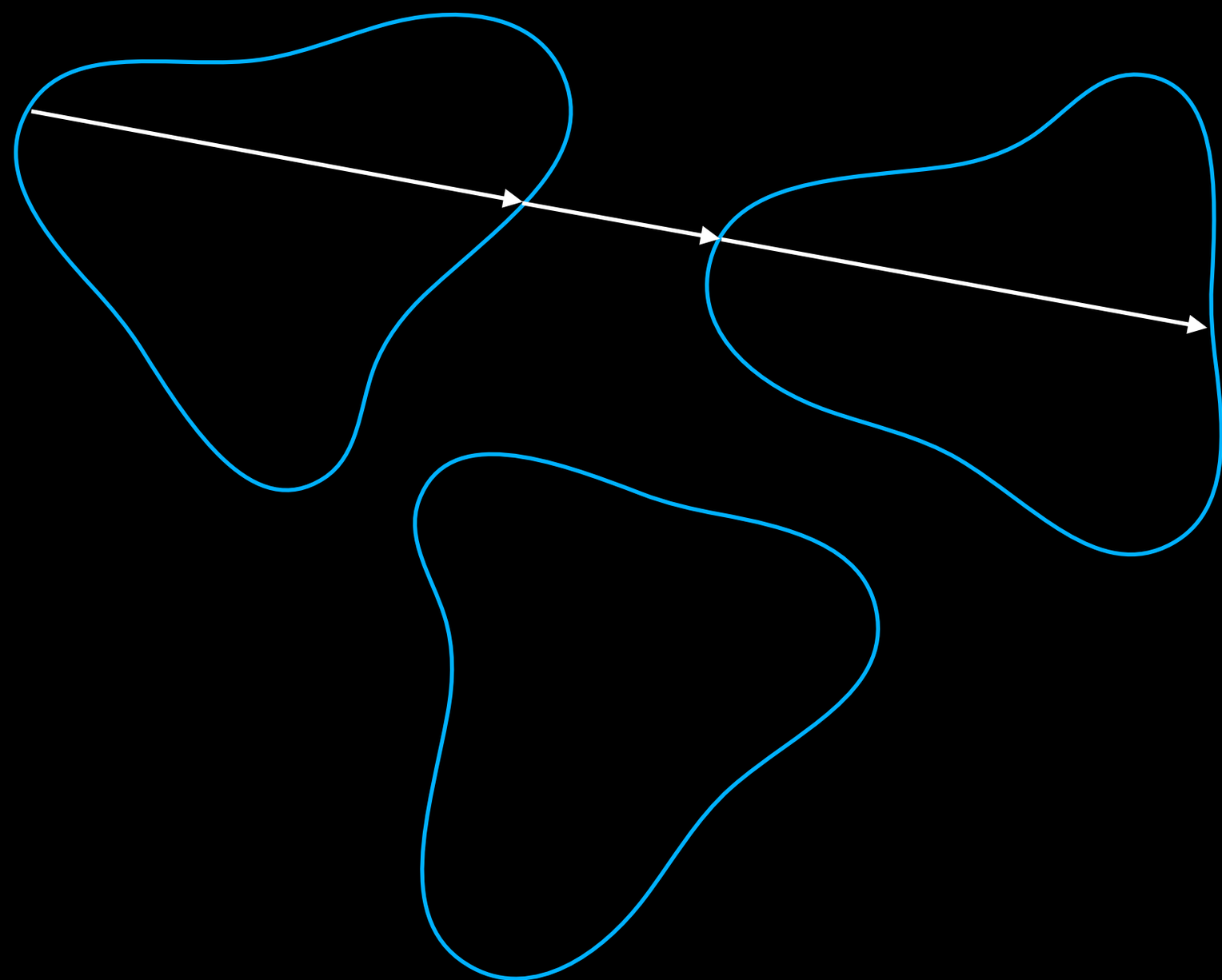
Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



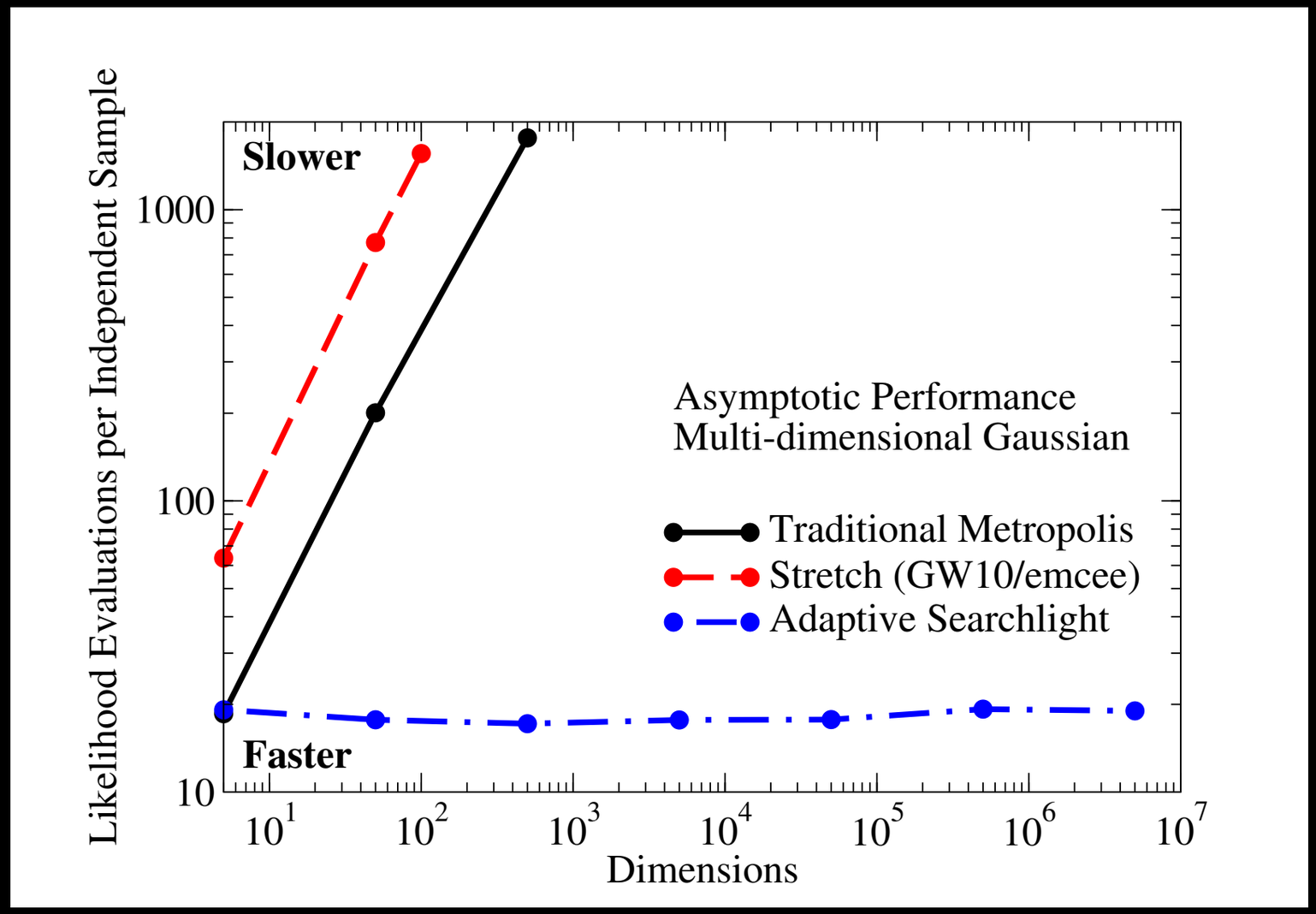
Should it be so hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



Naturally handles disjoint isoprobability surfaces(!)

Summary



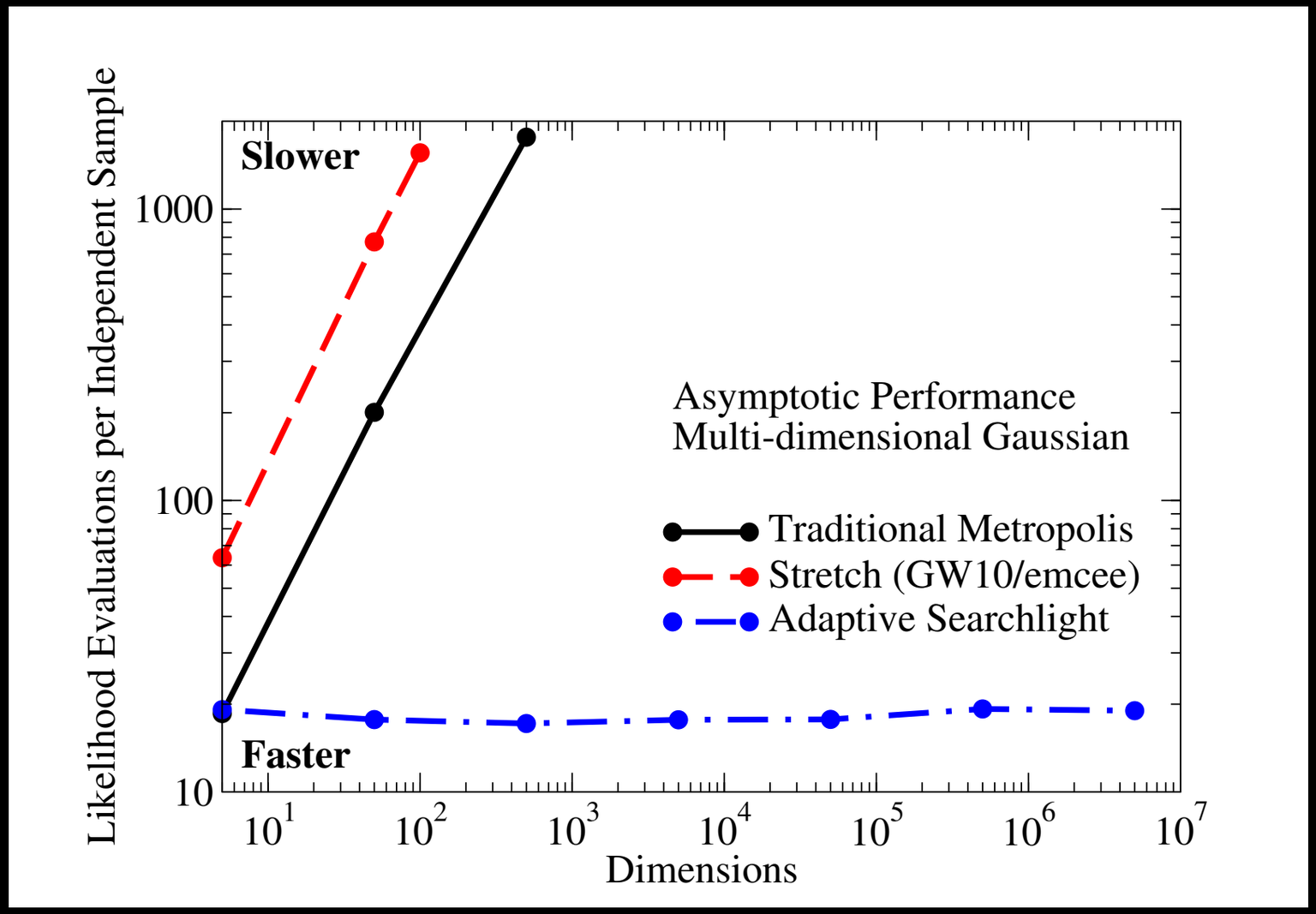
Searchlight: blackbody radiation from isolikelihood surfaces.

Connects MCMCs to Hamiltonian optics.

Naturally handles disjoint isoprobability surfaces(!)

O(1) - O(Log D) scaling with dimensionality

Summary



Searchlight: blackbody radiation from isolikelihood surfaces.

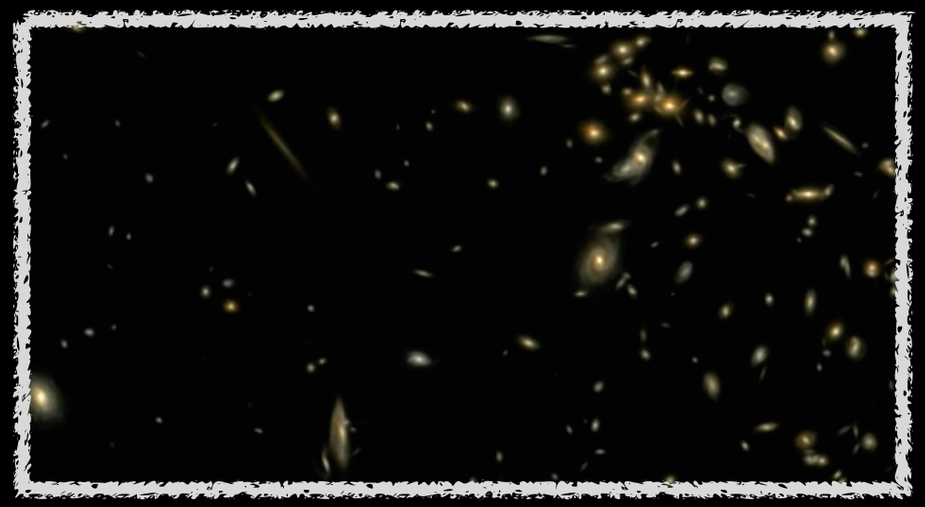
Connects MCMCs to Hamiltonian optics.

Naturally handles disjoint isoprobability surfaces(!)

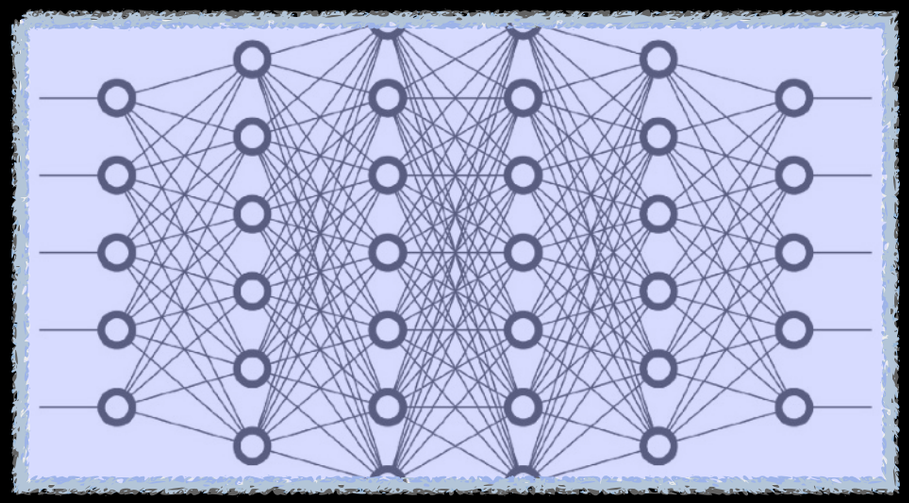
O(1) - O(Log D) scaling with dimensionality

Combining with normalizing flows would yield even better performance(!)

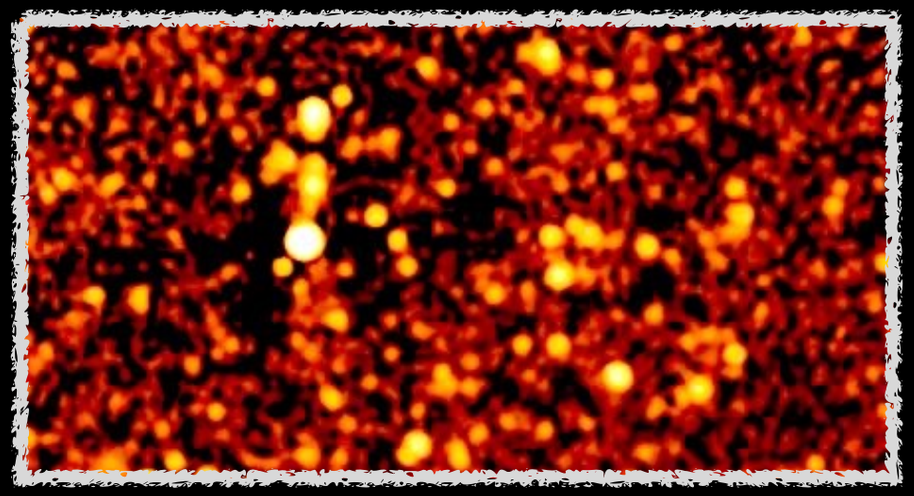
Motivation



Empirical Galaxy Modeling
50-100 dimensions



Deep Neural Networks
100-10^{14} dimensions



Confused Source Recovery
100 - 10^9 dimensions

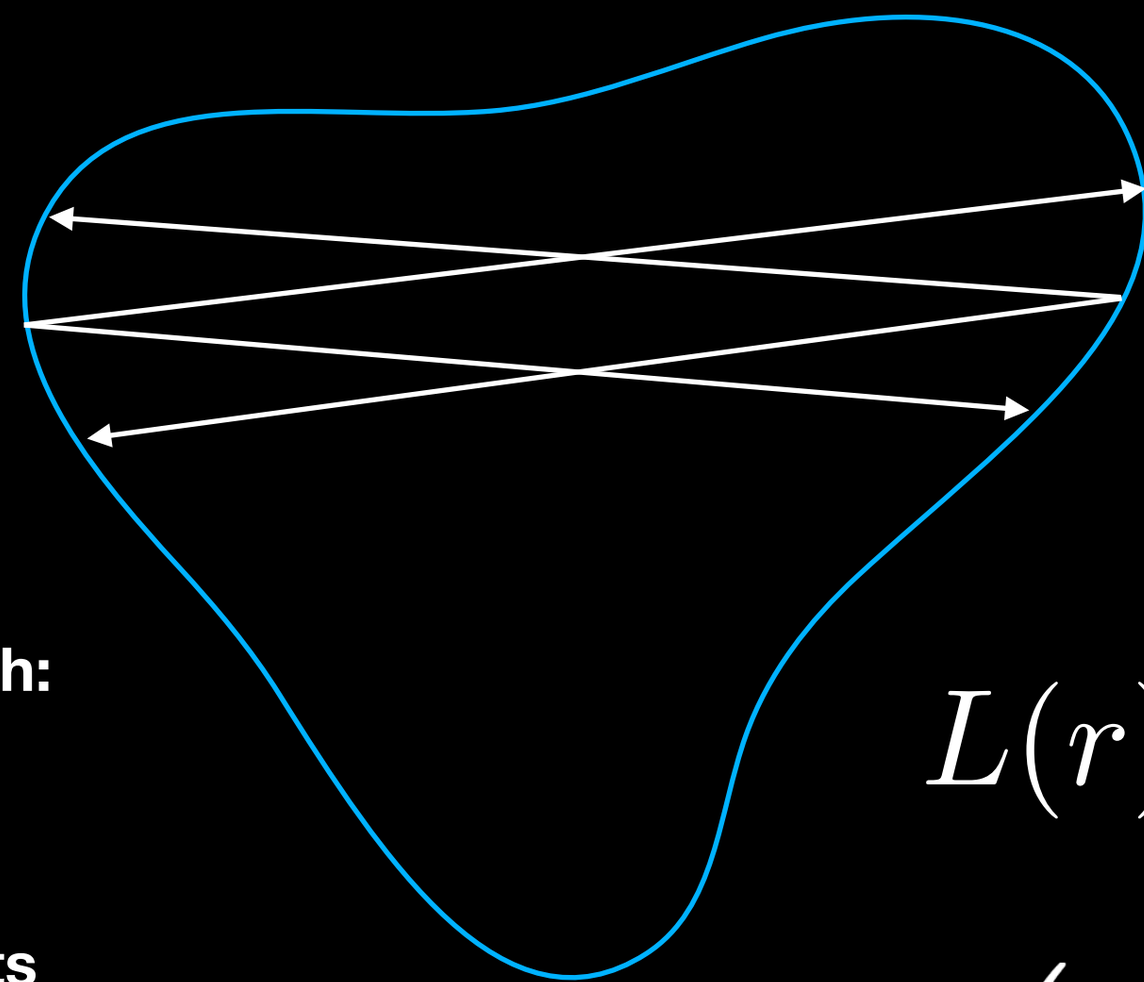
Constrained Realizations
> 10^9 dimensions

Peter Behroozi

Image Credits: SDSS, Radu Raicea, Jim Condon, CLUES

Should it be hard?

Searchlight: blackbody radiation from isolikelihood surfaces.



Metropolis / Stretch: Diffusion

Hamiltonian: Gravitational Orbits

$$L(r) \propto r^{1-D}$$

Searchlight: Optical Paths

$$P(x \to y) = \min\left(1, \frac{\hat{v} \cdot \vec{\nabla}\mathcal{L}(x)}{\hat{v} \cdot \vec{\nabla}\mathcal{L}(y)}\right)$$