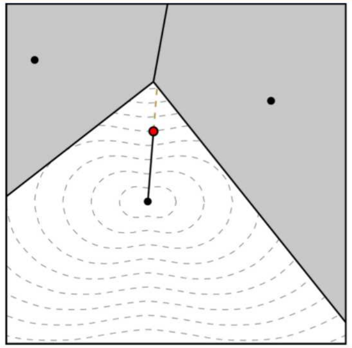
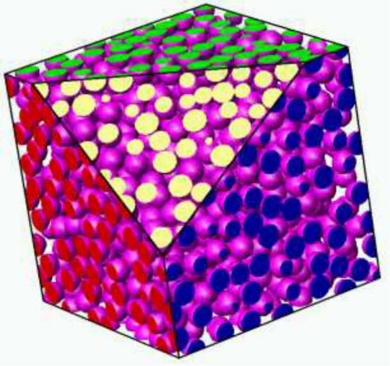
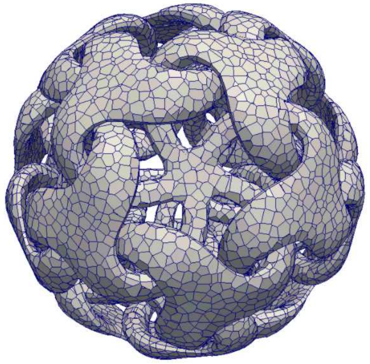
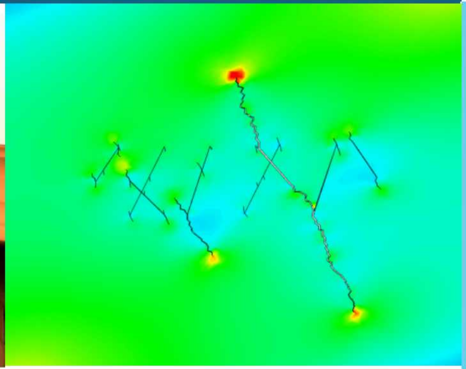
SAND2019-8350C

VoroSpokes Sampling for Bayesian Inference









Mohamed S. Ebeida

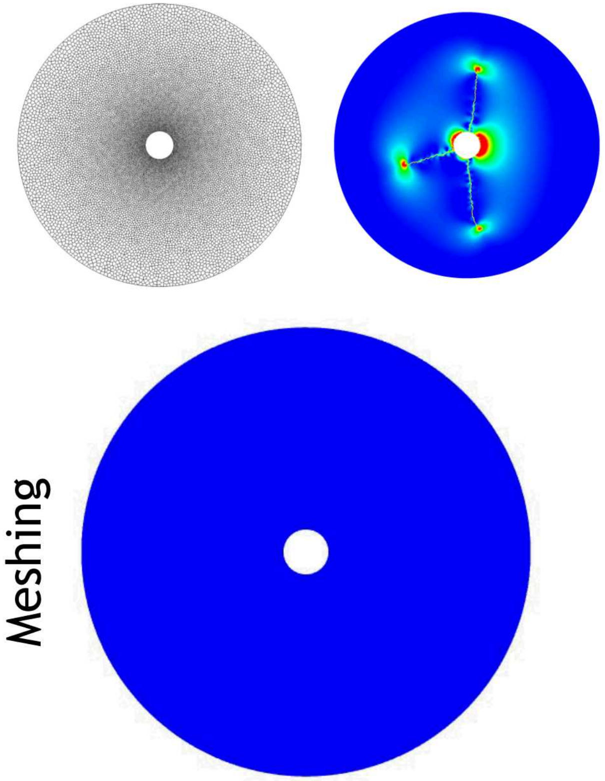
Discrete Math and Optimization, Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

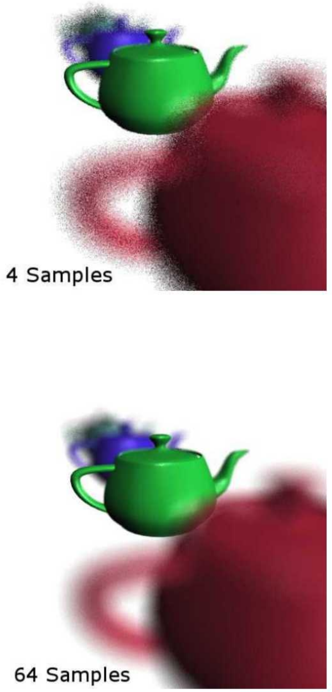
Background: Random Well-spaced Point sets

Random Well-spaced Point sets are desired for a wide range of applications:

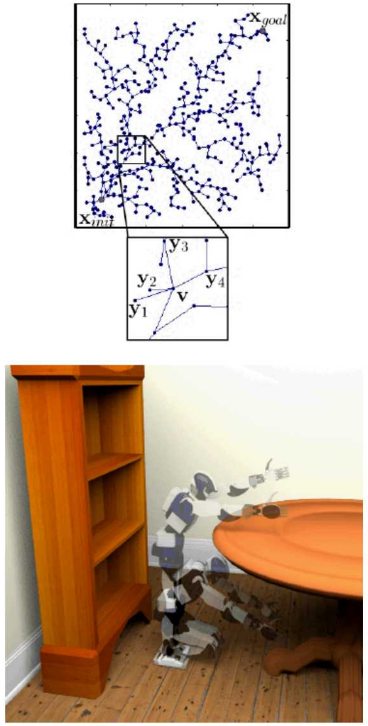
Meshing, Rendering, Motion planning, Uncertainty Quantification, Global Optimization, AI, …



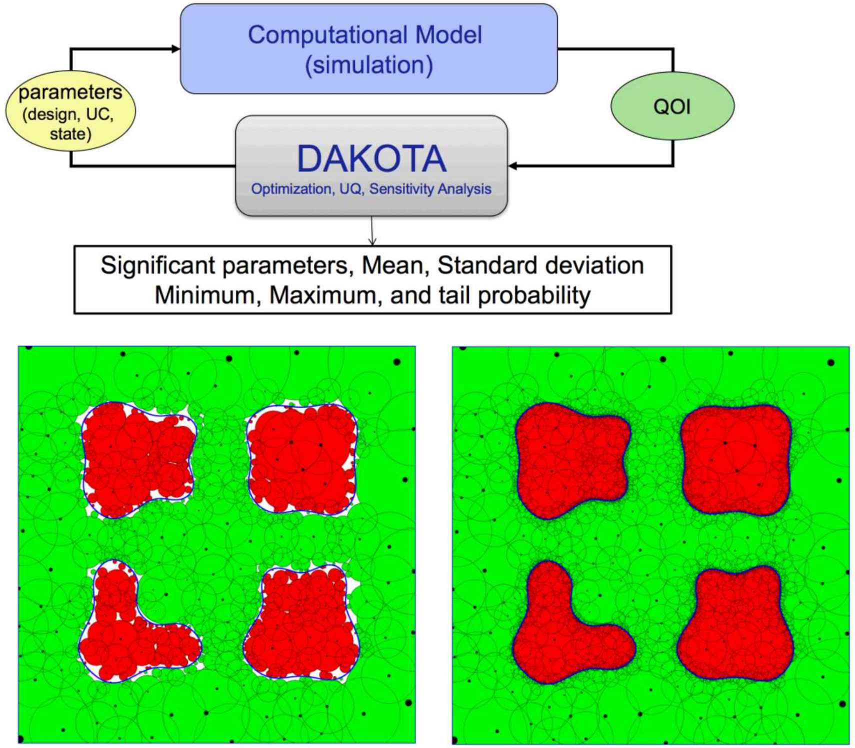
Meshing



Rendering



Motion Planning



Estimating Tail Probability in UQ

Background: Random Well-spaced Point sets

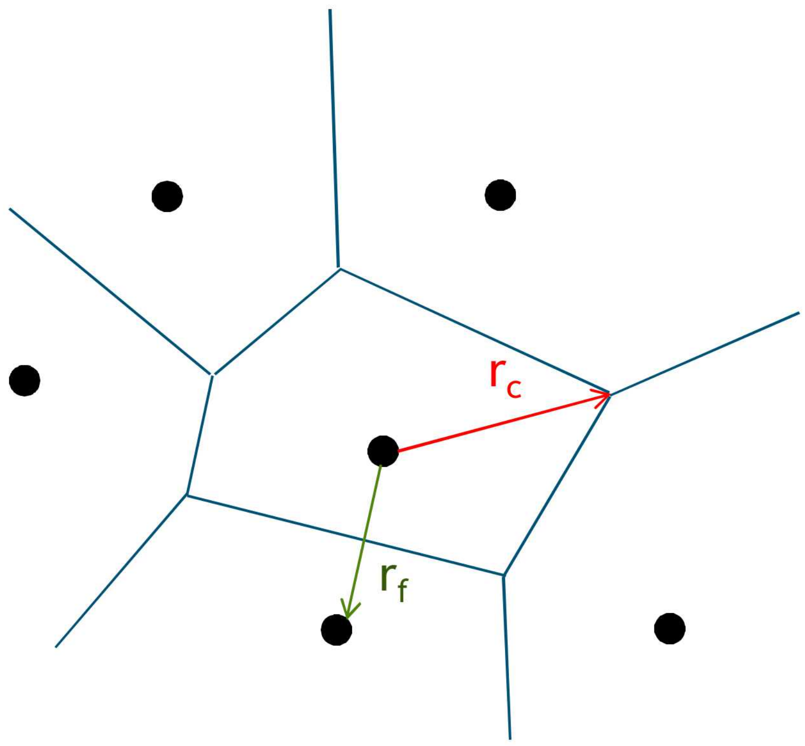
Well Spaced Point-Sets

Consider the Voronoi cell around any given point in the set:

Well-spaced means that Furthest Voronoi vertex is not much farther from closest neighbor.

r_c : distance between a given Voronoi seed and its furthest Voronoi corner

r_f: distance between a given Voronoi seed and its closest neighbor



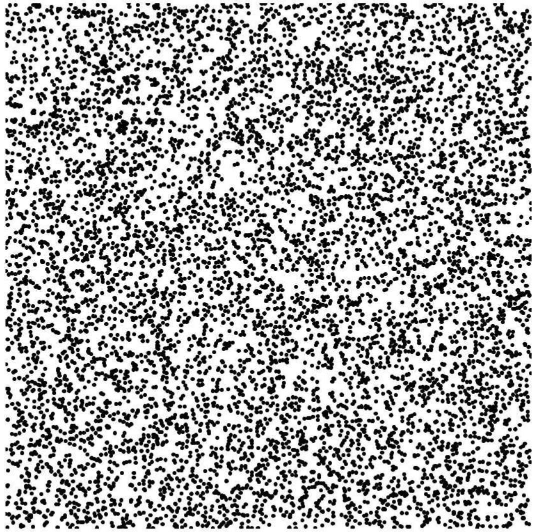
Voronoi Tessellations

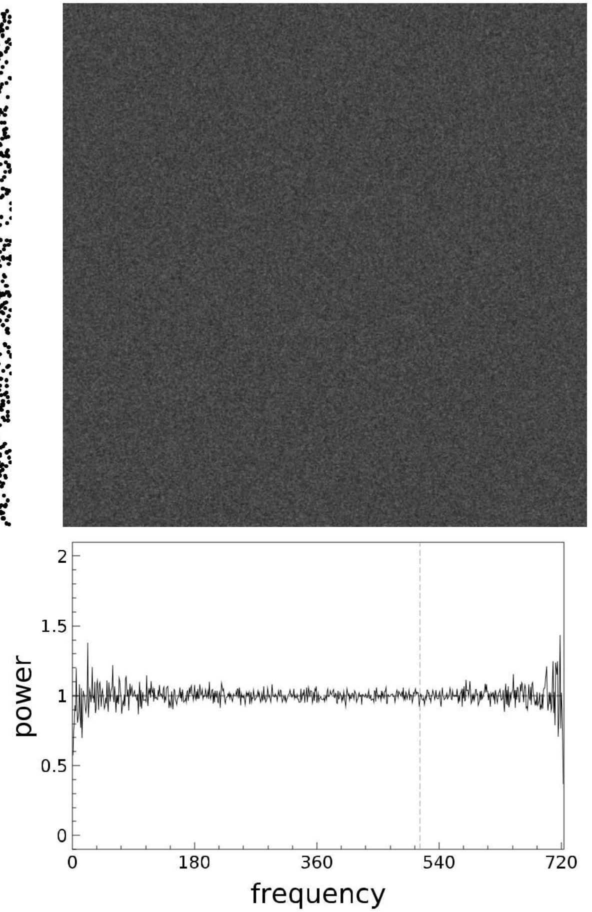
Voronoi partitions are uniquely defined by the seed location with a number of cells that matches the number of seeds regardless of dimensions, each cell is convex and bounded by planar convex facets.

Background: Random Well-spaced Point sets

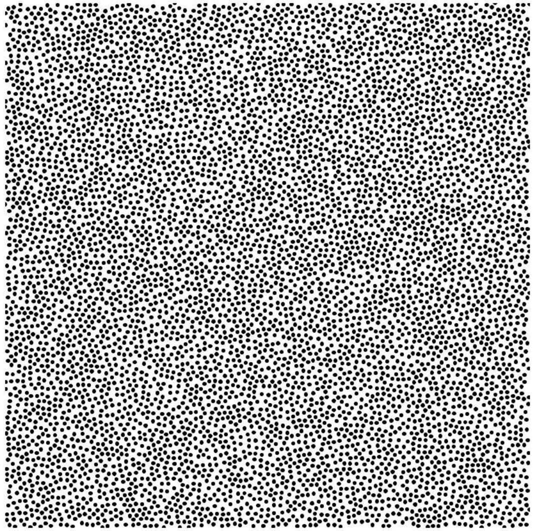
Random Point-Sets

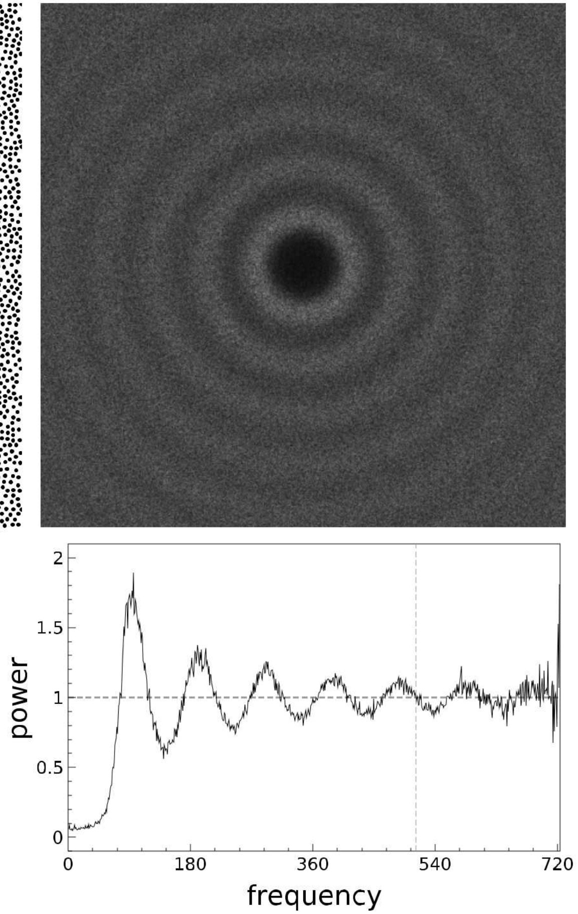
Random Point sets makes the exploration more efficient in higher dimensions (e.g. MC sampling).





MC sampling (white noise)





Poisson-disk sampling (Blue Noise)

Background: Poisson-disk Sampling

Disk-free condition

$$\forall x_i, x_j \in X, x_i \neq x_j : ||x_i - x_j|| \geq r$$

Bias-free:

Bias-free condition

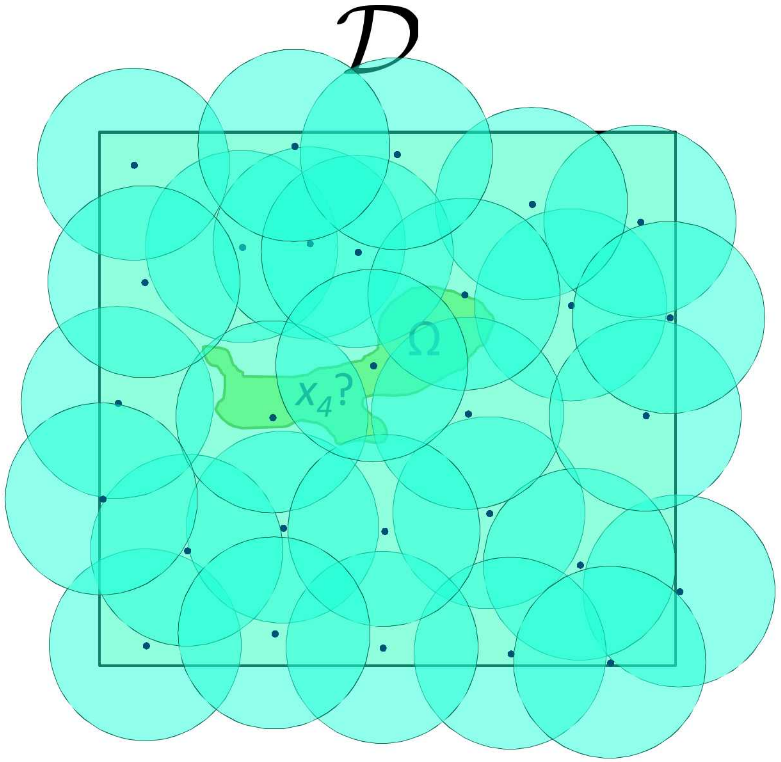
$$\forall x_i \in X, \forall \Omega \subset \mathcal{D}_{i-1} :$$

$$P(x_i \in \Omega) = \frac{\text{Area}(\Omega)}{\text{Area}(\mathcal{D}_{i-1})}$$

Maximal:

Maximal condition

$$\forall x \in \mathcal{D}, \exists x_i \in X : ||x - x_i|| < r$$



Simple Problem?!

Background: Poisson-disk Sampling

First E(n log n) algorithm with provably correct output

- Efficient Maximal Poisson-Disk Sampling,
  SIGGRAPH 2011

Simpler, less memory, , provably bias-free, faster in practice but no run-time proof, maximal up to round-off error

- A Simple Algorithm for Maximal Poison-Disk Sampling in High Dimensions,
  Eurographics 2012

Voronoi Meshes

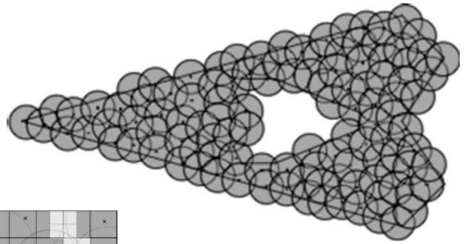
- VoroCrust: First method to mesh a non-convex domains without clipping
  SoCG 2018, TOG 2019 (Conditionally accepted), SIGGRAPH 2020!
- Uniform Random Voronoi Meshes
  IMR 2011

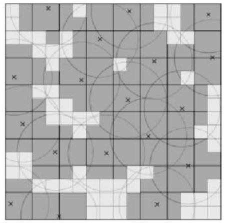
Random Delaunay Meshes

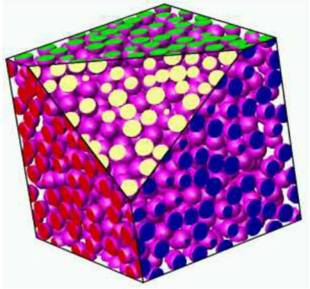
- Efficient and Good Delaunay Meshes from Random Points
  SIAM GD/SPM 2011 → Computer Aided Design
- Delaunay quadrangulation by two-coloring vertices
  IMR 2014

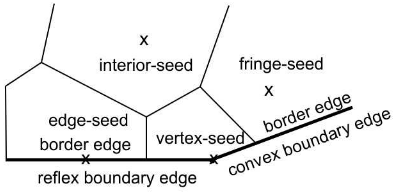
MPS with varying radii

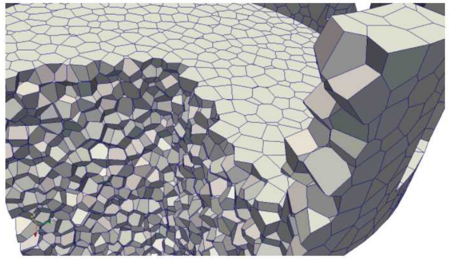
- Variable Radii Poisson-disk sampling
  CCCG 2012

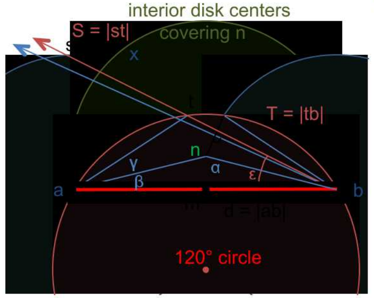












Background: Poisson-disk Sampling

Simulation of Propagating fractures

- Mesh Generation for modeling and simulation of carbon sequestration processes **SciDAC 2011**

Hyperplanes for integration, MPS and UQ

- K-d darts, **TOG 2014 (SIGGRAPH 2014)**
- Spoke Darts for high-dimensional blue noise sampling, **TOG 2018 (SIGGRAPH 2018)**

Rendering using line darts

- High quality parallel depth of field using line samples, **HPG 2012**

Reducing Sample size while respecting sizing function

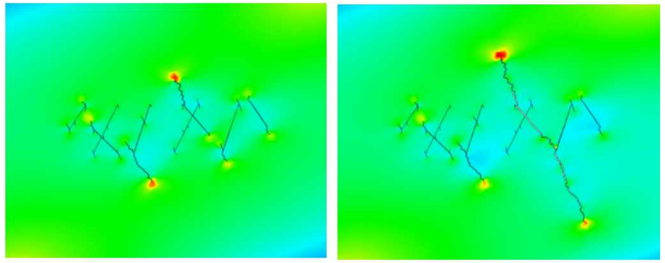
- Sifted Disks, **Eurographics 2013**
- Disk density tuning of a maximal random packing, **Eurographics 2016**

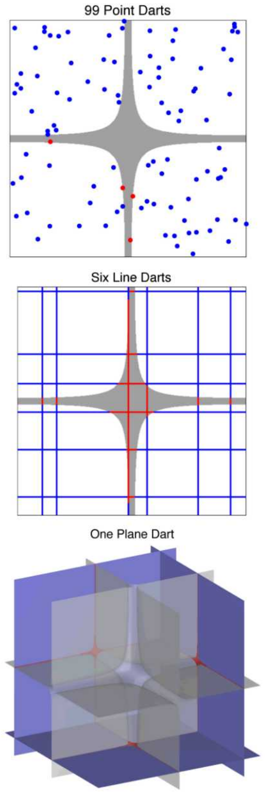
MPS with improved Coverage

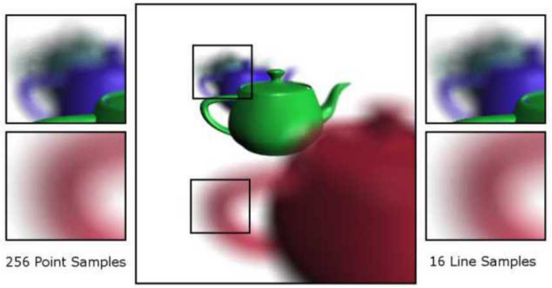
- Improving spatial coverage while preserving blue noise, **Computer Aided Design 2014**

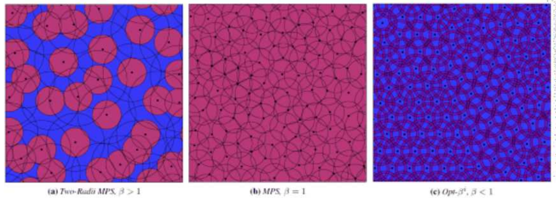
Uncertainty Quantification

- VPS: Voronoi piecewise surrogate models for high-dimensional data fitting, **IJ4UQ 2017**
- Pof-darts: Geometric adaptive sampling for probability of failure, **RESS 2016**

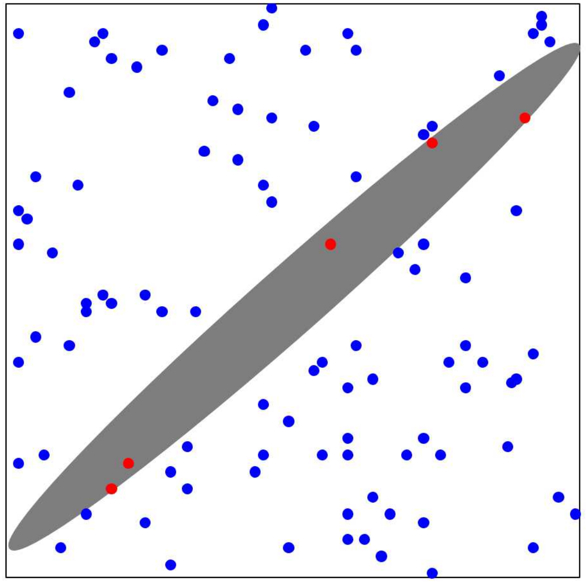




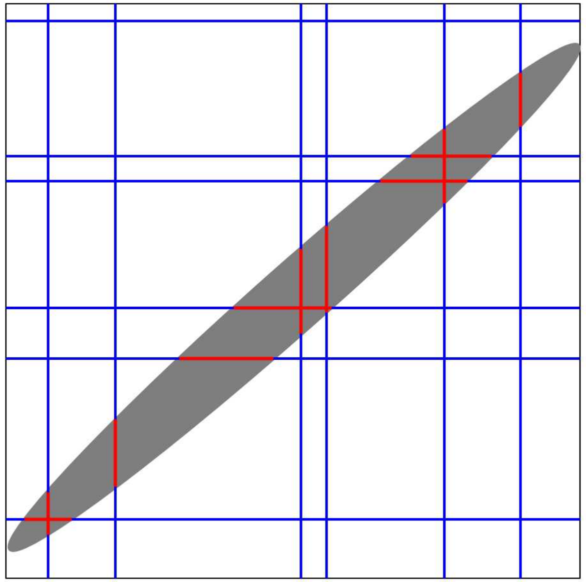




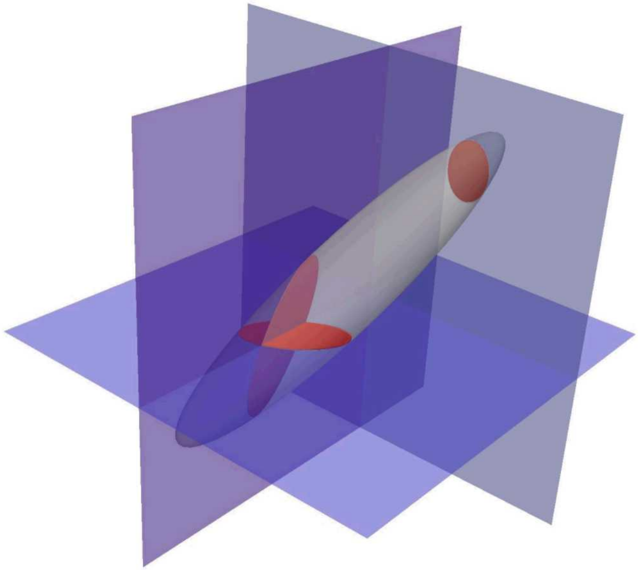
Background: kd-darts (SIGGRAPH 2014)



99 points



6 line darts

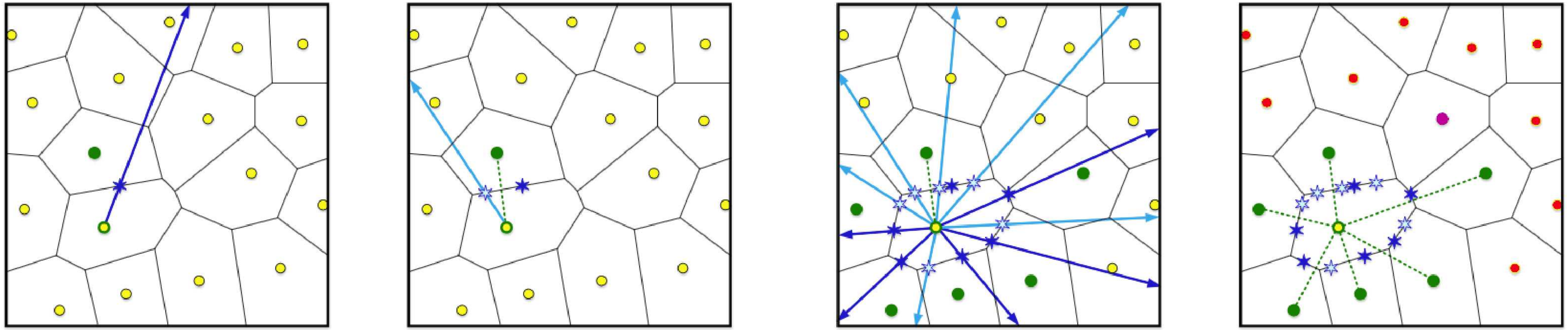


1 plane dart

Lines, hyperplanes, are more likely to intersect these regions, and they give more information

But they are more expensive. Is it worth it?

Background: Spoke darts (SIGGRAPH 2018)



(a) A spoke samples a point from a Voronoi facet.

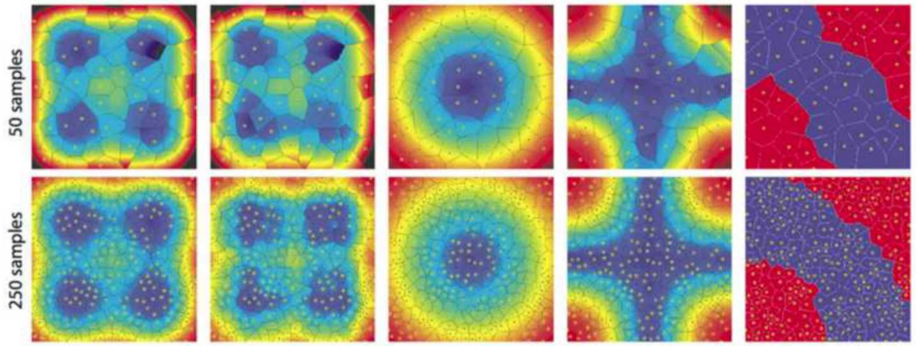
(b) A spoke hitting same facet again.

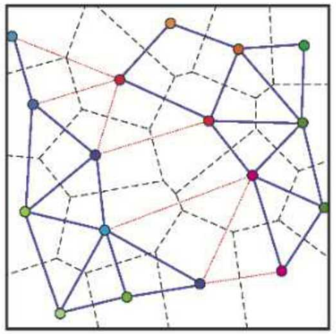
(c) Local MC line sampling to find Delaunay neighbors.

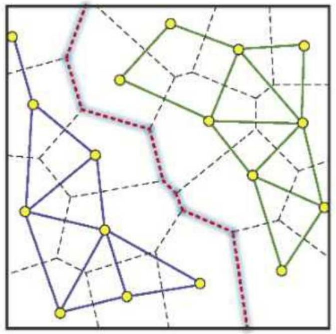
(d) Neighbors with relatively small angles missed.

Collecting Voronoi Neighbors from all directions

A direct application: Voronoi Piecewise Surrogates (VPS)





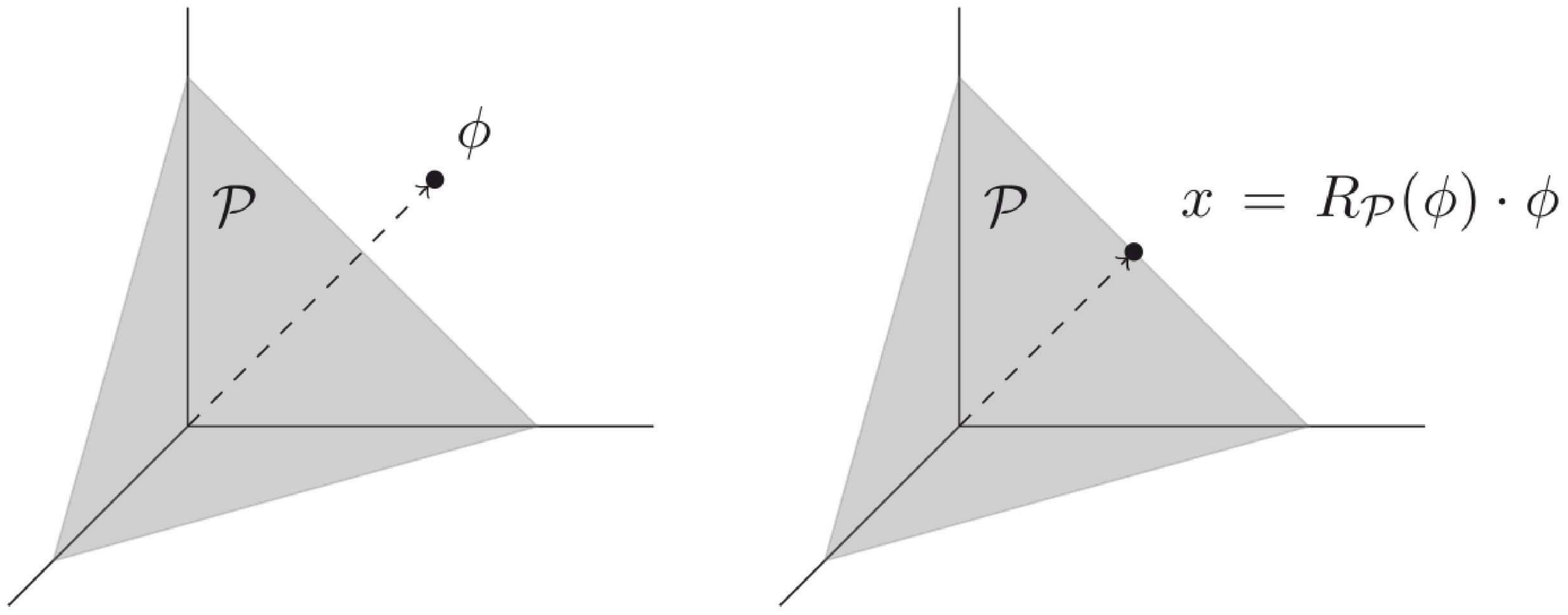


Next: VoroSpokes for Bayesian Inference

VoroSpokes for Bayesian Inference

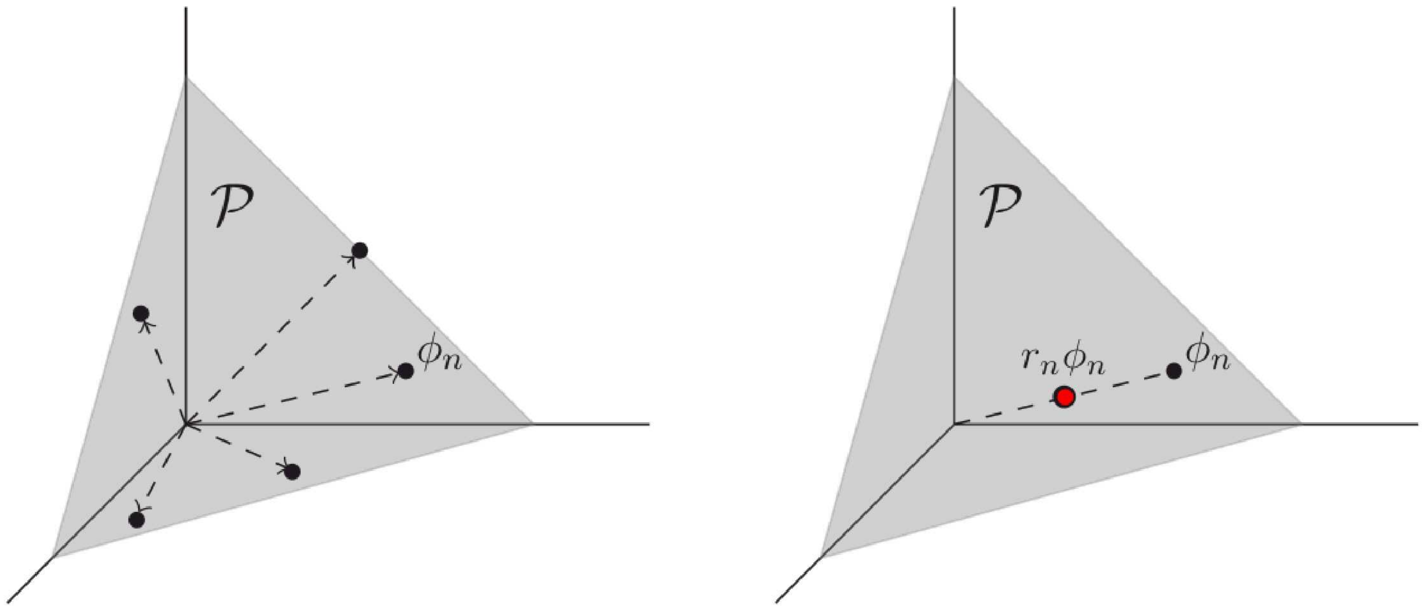
Nick Winovich, Ahmad A. Rushdi , Eric T. Phipps , Laura P. Swiler, Jaideep Ray, Guang Lin, & Mohamed S. Ebeida

VoroSpokes: Approximation with Spokes



- Random unit-vectors $\phi \sim \mathrm{Unif}(S^{d-1})$ can be used to identify the boundaries of convex polytopes, e.g. Voronoi cells.
- The maximum radial distance $R_{\mathcal{P}}(\phi)$ inside of $\mathcal{P}$ in the direction of the unit-vector ϕ can be computed using trimming operations.

VoroSpokes: Approximation with Spokes



- By considering a sufficient number of random directions $\{\phi_n\}_{n=1}^{N}$, the volume ot the polytope $\mathcal{P}$ (or integral of a function defined therein) can be approximated to an arbitrary level of accuracy.
- To probe the interior of the polytope, radial distances $\{r_n\}$ in the directions $\{\phi_n\}$ can be randomly sampled from appropriate distributions placed over the intervals $\{[0, R_{\mathcal{P}}(\phi_n)]\}$, respectively.

Spoke Quadrature on Convex Polytopes

Given a bounded function defined inside of a convex polytope, we can express the integral in terms of polar coordinates via:

$$
\begin{aligned}
\int_{\mathcal{P}} f(x)\,dx &= \int_{\mathbb{R}^d} \mathbf{1}_{\mathcal{P}}(x) \cdot f(x)\,dx \\
&= \int_0^{\infty} \int_{S^{d-1}} \mathbf{1}_{\mathcal{P}}(r \cdot \phi) \cdot f(r \cdot \phi) \cdot r^{d-1}\,d\sigma(\phi)\,dr \\
&= \int_{S^{d-1}} \int_0^{\infty} \mathbf{1}_{\mathcal{P}}(r \cdot \phi) \cdot f(r \cdot \phi) \cdot r^{d-1}\,dr\,d\sigma(\phi) \\
&= \int_{S^{d-1}} \int_0^{R_{\mathcal{P}}(\phi)} f(r \cdot \phi) \cdot r^{d-1}\,dr\,d\sigma(\phi)
\end{aligned}
$$

Spoke Quadrature on Convex Polytopes

$$
\begin{aligned}
\int_{\mathcal{P}} f(x)\,dx &= \frac{1}{\sigma_d} \int_{S^{d-1}} \int_0^{R_{\mathcal{P}}(\phi)} \sigma_d \cdot f(r \cdot \phi) \cdot r^{d-1}\,dr\,d\sigma(\phi) \\
&= \mathbb{E}_{\theta \sim \mathrm{Unif}(S^{d-1})} \left[\int_0^{R_{\mathcal{P}}(\theta)} \sigma_d \cdot f(r \cdot \theta) \cdot r^{d-1}\,dr \right] \\
&= \mathbb{E}_{\theta \sim \mathrm{Unif}(S^{d-1})} \left[\sigma_d \cdot I_{\mathcal{P}}(\theta) \right]
\end{aligned}
$$

where we define $I_{\mathcal{P}}(\theta) = \int_0^{R_{\mathcal{P}}(\theta)} f(r \cdot \theta) \cdot r^{d-1}\,dr$ -dimensional

"radial integral" inside of the polytope in the direction of theta.

Spoke Quadrature on Convex Polytopes

Applications of the strong law of large numbers and central limit theorem yield:

$$\frac{1}{N}\sum_{n=1}^{N}\sigma_d \cdot I_{\mathcal{P}}(\theta^{(n)}) \xrightarrow{a.s.} \int_{\mathcal{P}} f(x)\,dx \quad \text{as} \quad N \longrightarrow \infty$$

$$\frac{1}{\sqrt{N}}\left(\sum_{n=1}^{N}\sigma_d \cdot I_{\mathcal{P}}(\theta^{(n)}) - N \cdot \int_{\mathcal{P}} f(x)\,dx\right) \xrightarrow{distribution} \mathcal{N}(0,\sigma^2) \quad \text{where } \sigma^2 = \mathrm{Var}\left[\sigma_d \cdot I_{\mathcal{P}}(\theta)\right]$$

where $\theta^{(n)}$ correspond to independent identically distributed (i.i.d.) samples drawn from the uniform distribution $p_\theta(\phi) = \frac{1}{\sigma_d}\mathbf{1}_{S^{d-1}}(\phi)$ on the sphere.

Spoke Sampling on Convex Polytopes

We propose the following hierarchical sampling procedure on polytopes:

$$p_\theta(\phi) \;=\; \frac{1}{C}\int_0^{R_{\mathcal{P}}(\phi)} f(r\cdot\phi)\cdot r^{d-1}\,dr \;=\; \frac{1}{C}\,I_{\mathcal{P}}(\phi) \qquad \forall\,\phi\,\in\,S^{d-1}$$

$$p_\rho\big(\,r\mid\theta=\phi\big) \;=\; \frac{1}{I_{\mathcal{P}}(\phi)}\cdot f(r\cdot\phi)\cdot r^{d-1} \qquad \forall\;r\,\in\,\big[0,R_{\mathcal{P}}(\phi)\big]\;,\;\;\phi\,\in\,S^{d-1}$$

where the normalization constants C and $I_{\mathcal{P}}(\phi)$ for the densities are given by:

$$C \;=\; \int_{S^{d-1}}\int_0^{R_{\mathcal{P}}(\phi)} f(r\cdot\phi)\cdot r^{d-1}\;dr\,d\sigma(\phi) \;=\; \int_{\mathcal{P}} f(x)\,dx$$

$$I_{\mathcal{P}}(\phi) \;=\; \int_0^{R_{\mathcal{P}}(\phi)} f(r\cdot\phi)\cdot r^{d-1}\;dr \;=\; C\,\cdot\,p_\theta(\phi) \qquad \forall\,\phi\,\in\,S^{d-1}$$

Spoke Sampling on Convex Polytopes

The joint density for (ρ, θ) is defined for each pair (r, ϕ) such that $r \cdot \phi \in \mathcal{P}$ as follows:

$$\begin{aligned} p_{\rho,\theta}(r,\phi) &= p_\theta(\phi) \cdot p_\rho\big(r \mid \theta = \phi\big) \\ &= p_\theta(\phi) \cdot \left[\frac{1}{I_{\mathcal{P}}(\phi)} \cdot f(r \cdot \phi) \cdot r^{d-1}\right] = \frac{p_\theta(\phi)}{C \cdot p_\theta(\phi)} \, f(r \cdot \phi) \cdot r^{d-1} \\ &= \frac{1}{C} \, f(r \cdot \phi) \cdot r^{d-1} \end{aligned}$$

Spoke Sampling on Convex Polytopes

The measure μ_X induced by the random variable $X = \rho \cdot \theta$ is thus defined by:

$$
\begin{aligned}
\mu_X(E) &= \mathbb{P}\big[X \in E\big] = \mathbb{P}\big[\rho \cdot \theta \in E\big] \\
&= \int_{S^{d-1}} \int_0^\infty \mathbf{1}_E(r \cdot \phi) \cdot p_{\rho,\theta}(r, \phi) \; dr \, d\sigma(\phi) \\
&= \int_{S^{d-1}} \int_0^\infty \mathbf{1}_E(r \cdot \phi) \cdot \left[\frac{1}{C} \mathbf{1}_{\mathcal{P}}(r \cdot \phi) \cdot f(r \cdot \phi) \cdot r^{d-1}\right] \; dr \, d\sigma(\phi) \\
&= \frac{1}{C} \int_0^\infty \int_{S^{d-1}} \mathbf{1}_{E \cap \mathcal{P}}(r \cdot \phi) \cdot f(r \cdot \phi) \cdot r^{d-1} \; d\sigma(\phi) \, dr \\
&= \frac{1}{C} \int_{\mathbb{R}^d} \mathbf{1}_{E \cap \mathcal{P}}(x) \cdot f(x) \; dx \\
&= \frac{1}{C} \int_{E \cap \mathcal{P}} f(x) \; dx
\end{aligned}
$$

Spoke Sampling on Convex Polytopes

The main difficulty with this approach is sampling from the density:

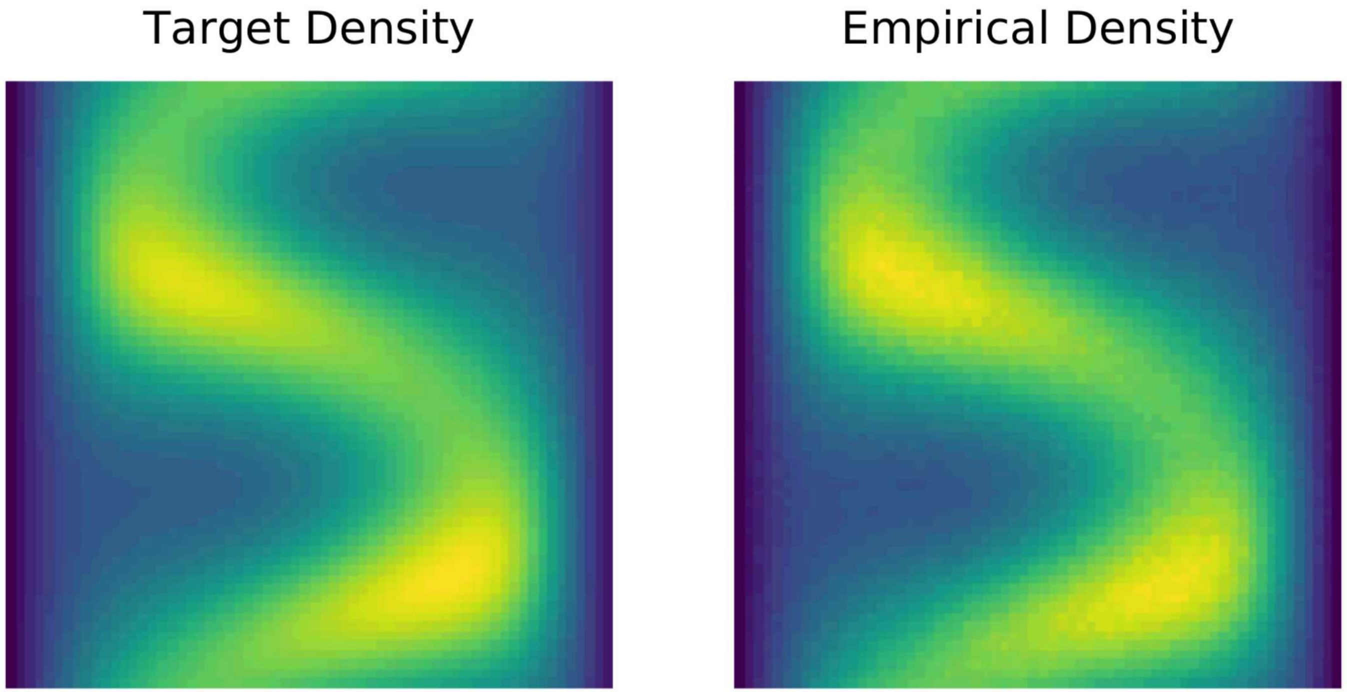
$$p_\theta(\phi) \;=\; \frac{1}{C}\, I_{\mathcal{P}}(\phi) \;=\; \frac{1}{C}\int_0^{R_{\mathcal{P}}(\phi)} f(r\cdot\phi)\cdot r^{d-1}\,dr \qquad \text{for} \qquad \phi \in S^{d-1}$$

To avoid this issue, we introduce the following approximate sampling procedure which is shown to converge in distribution to the target density on the sphere:

1. $\theta^{(n)} \sim \mathrm{Unif}(S^{d-1}) \quad \text{for} \quad 1 \le n \le N$
2. $I_{\mathcal{P}}\left(\theta^{(n)}\right) = \int_0^{R_{\mathcal{P}}(\phi)} f\left(r\cdot\theta^{(n)}\right)\cdot r^{d-1}\,dr$
3. $\widehat{C}_N = \frac{1}{N}\sum_{n=1}^{N} \sigma_d \cdot I_{\mathcal{P}}\left(\theta^{(n)}\right)$
4. $\theta^*_N \sim \widehat{p}_{\theta^*}(\phi) = \frac{\sigma_d}{N\cdot \widehat{C}_N}\sum_{n=1}^{N} \delta_{\theta^{(n)}}(\phi)\cdot I_{\mathcal{P}}\left(\theta^{(n)}\right)$

Spoke Sampling on Convex Polytopes

Empirical results show that this theta-sampling procedure converges quite quickly to the target density on the sphere (e.g. using only 100 spokes/sample):



[200M Samples w/ 100 Spokes]

Moreover, the overall sampling procedure is shown to converge in distribution:

For all Borel measurable sets $E \in \mathcal{B}(\mathbb{R}^d)$ *we have:* $\lim_{N\to\infty} \mathbb{P}\left[X_N^* \in E\right] = \frac{1}{C}\int_{E\cap\mathcal{P}} f(x)\,dx$

Spoke Quadrature and Sampling

The procedures outlined in the previous slides thus provide numerical methods for approximating integrals and generating approximate samples in polytopes.

These generalize to global procedures in the context of Voronoi tessellations:

$$\mathcal{V}(i) = \{x \in \Omega : \|x - \overline{x}_i\| < \|x - \overline{x}_j\| \; \forall j \in \mathcal{I} \setminus \{i\}\} \quad \text{with} \quad \bigcup_{i \in \mathcal{I}} \overline{\mathcal{V}(i)} = \Omega$$

Quadrature is extended via:

$$\int_{\Omega} f(x)\, dx = \sum_{i \in \mathcal{I}} \int_{\mathcal{V}(i)} f(x)\, dx \approx \sum_{i \in \mathcal{I}} w_i$$

Sampling is extended by defining approximate/target cell weights:

$$\widehat{\pi}(\mathcal{V}(i)) = \frac{w_i}{\sum_{j \in \mathcal{I}} w_j} \quad , \quad \pi(\mathcal{V}(i)) = \frac{1}{C} \int_{\mathcal{V}(i)} f(x)\, dx \qquad \forall i \in \mathcal{I}$$

and using the resulting discrete p.d.f. on cells to select a single Voronoi cell.

Once the cell is selected, the local sampling procedure on polytopes is applied.

VoroSpokes: Overview

The VoroSpokes framework consists of two fundamental phases:

Phase I: Adaptive VPS Approximation

- The posterior density is adaptively approximated by a VPS model.
- Geometric properties of the posterior, such as multi-modality, are identified and refined prior to sampling.

Phase II: VoroSpokes Sampling Procedure

- A regional sampling procedure is used to select which Voronoi cell a given sample will be drawn from.
- Once a cell is selected, a hierarchical sampling procedure is employed within the cell to select a direction θ and radial length ρ specifying the final sample: $x^*_{seed} + \rho \cdot \theta$

VoroSpokes Phase I: Adaptive VPS

The VPS approximation to the posterior density is designed to construct local surrogate models within each cell of an associated Voronoi tessellation. A local surrogate can, for example, take the form of a polynomial of a specified degree D:

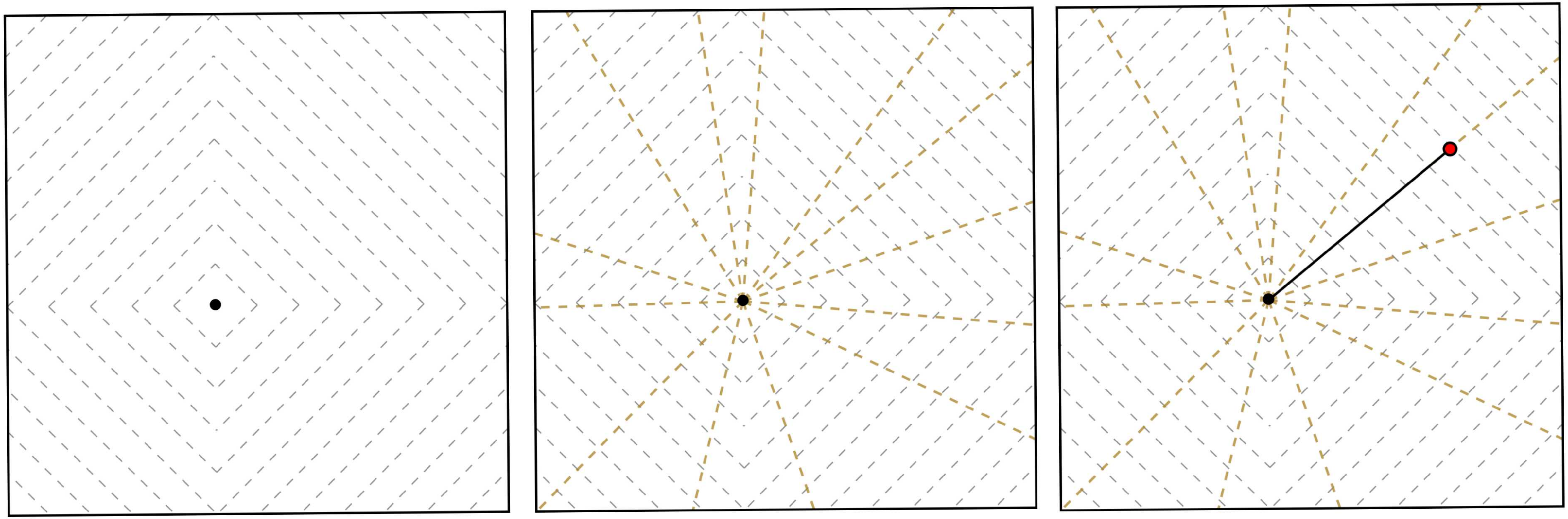
$$f(x) \approx f_{VPS}(x) = \sum_{|\alpha| \leq D} c_\alpha \cdot (x - x_{seed})^\alpha \quad \forall\, x \in \mathcal{V}(x_{seed})$$

The *adaptive* component of the VPS approximation in Phase I is derived from the local error estimates:

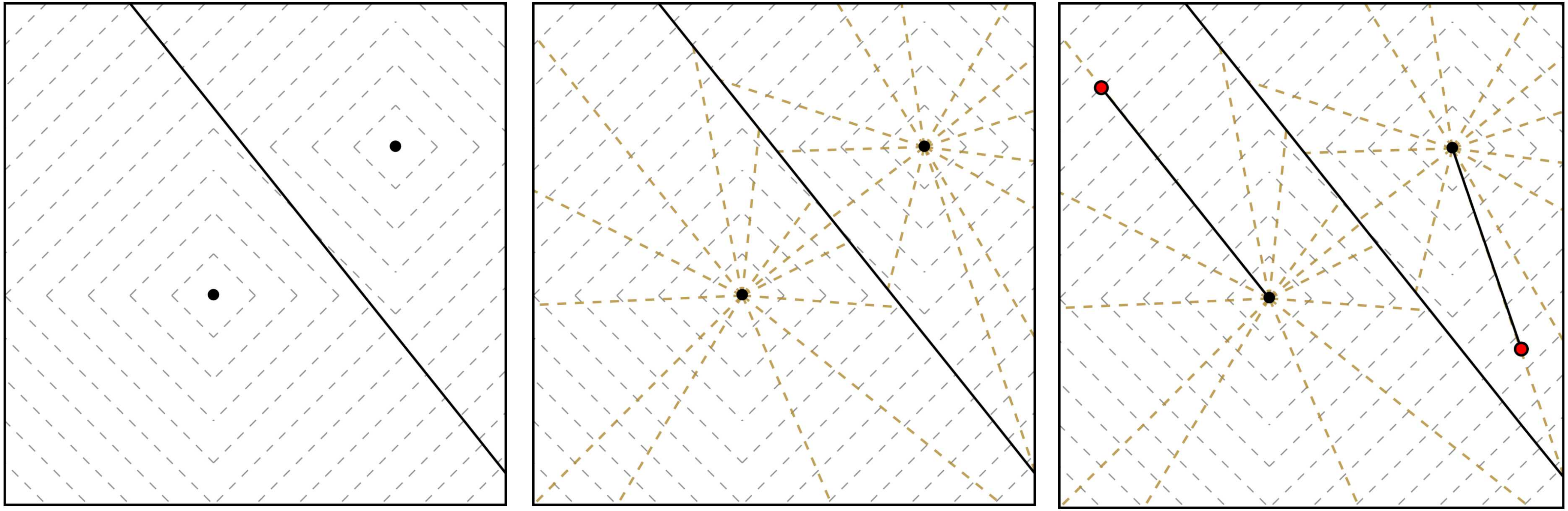
$$\mathcal{E}_{VPS}(x) = \sum_{|\alpha| = D} |c_\alpha| \cdot |x - x_{seed}|^\alpha \quad \forall\, x \in \mathcal{V}(x_{seed})$$

In particular, new Voronoi seeds are added using a sampling procedure which is designed to target regions with high uncertainty/error.

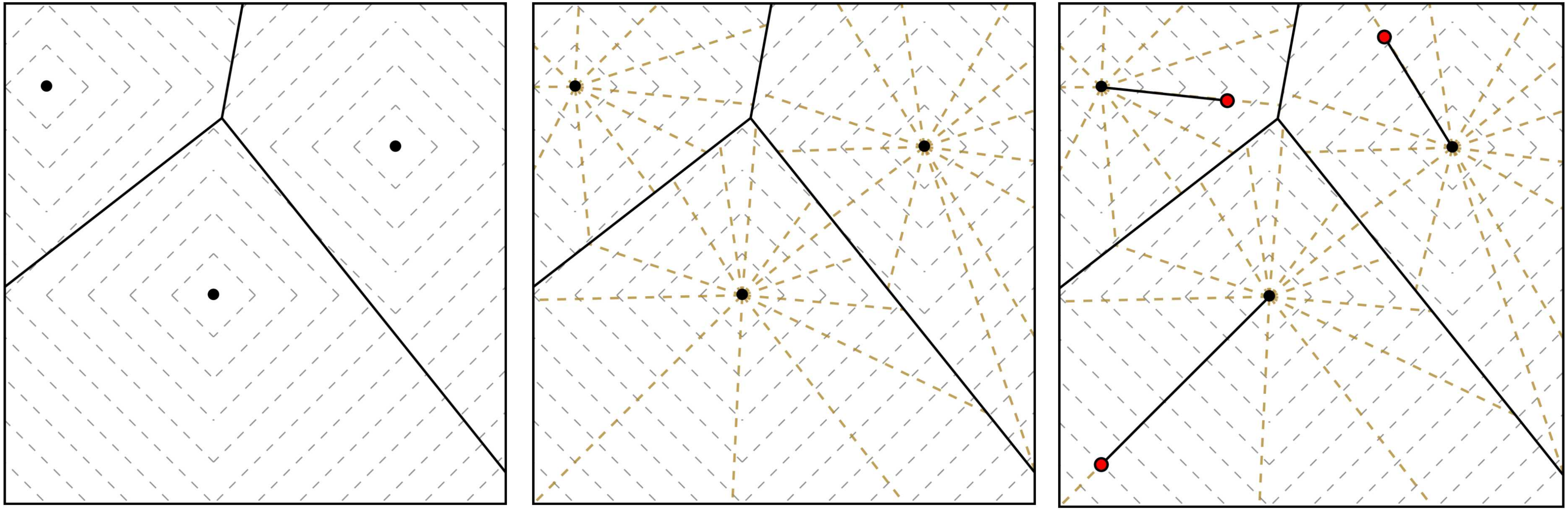
VoroSpokes Phase I: Adaptive VPS



VoroSpokes Phase I: Adaptive VPS



VoroSpokes Phase I: Adaptive VPS



VoroSpokes Phase II: Sampling Procedure

Once the VPS approximation from Phase I is complete, the cell probability weights are re-computed using the approximate posterior $f_{VPS}(x)$ in place of the adaptive error estimates $\mathcal{E}_{VPS}(x)$:

$$\widehat{\pi}(\mathcal{V}(i)) \;\approx\; \pi(\mathcal{V}(i)) \;=\; \frac{1}{C}\int_{\mathcal{V}(i)} f_{VPS}(x)\,dx \qquad \forall\, i \in \mathcal{I}$$

The discrete probability weights are approximated via:

$$\widehat{\pi}(\mathcal{V}(i)) \;=\; \frac{w_i}{\sum\limits_{j\in\mathcal{I}} w_j} \quad \text{where} \quad w_i \;\approx\; \int_{\mathcal{V}(i)} f_{VPS}(x)\,dx$$

[Spoke Quadrature]

Note: The sum $\sum\limits_{j\in\mathcal{I}} w_j$ approximates the constant $C = \int f_{VPS}(x)\,dx$

VoroSpokes Phase II: Sampling Procedure

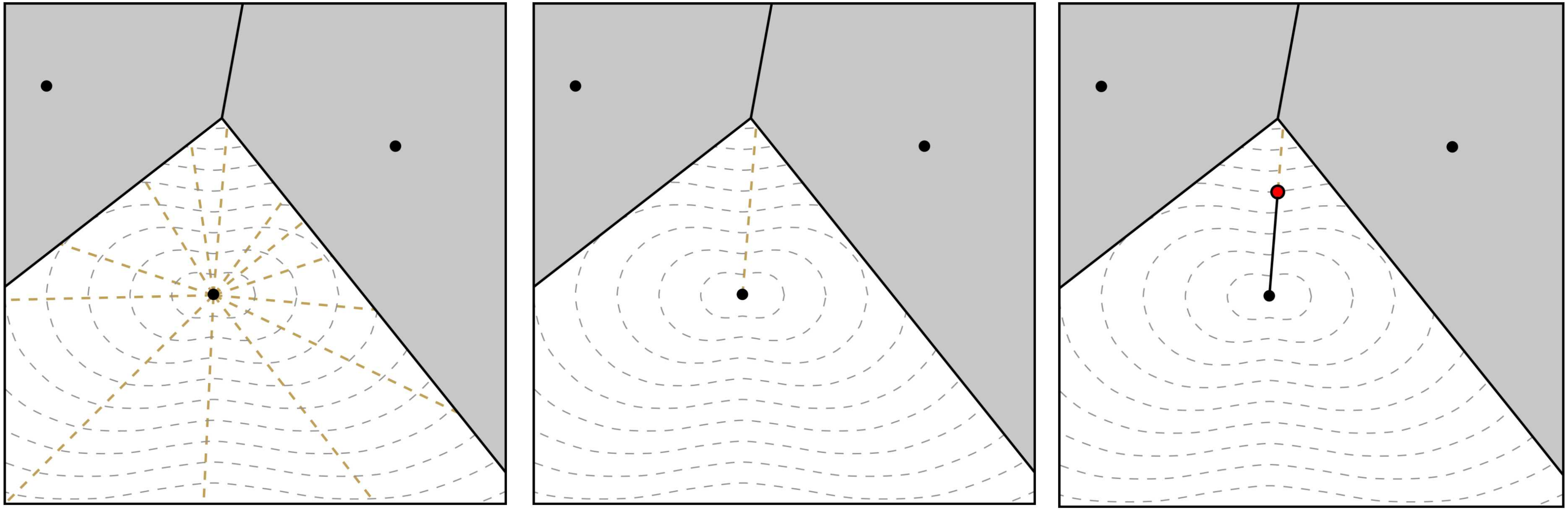
Once a cell has been selected, a unit-vector/direction θ is sampled from a carefully specified marginal density.

A radial length/distance ρ is then sampled from an associated conditional density, and the end-point of the associated spoke is returned as an approximate sample from the posterior.

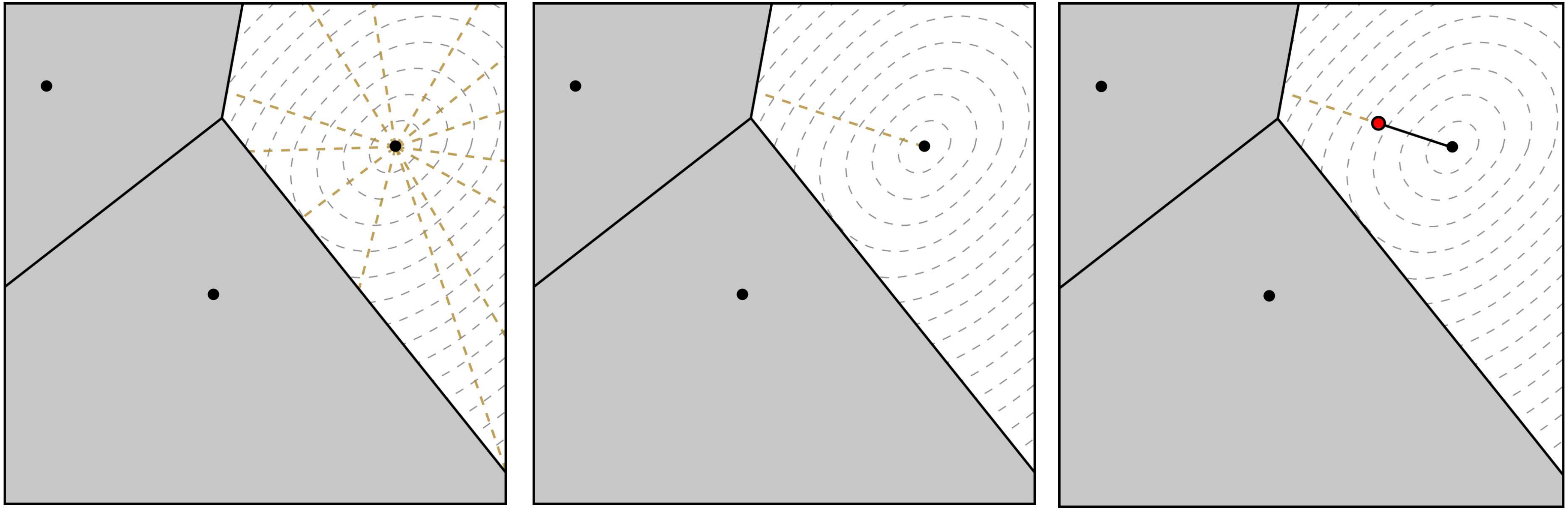
Overview of Sampling Procedure

1. Select a Voronoi cell: $\mathcal{V}(i^*)$
2. Sample a direction: $\theta \in \mathcal{S}^{d-1}$
3. Sample a distance: $\rho \in \left[0,\ R_{\mathcal{V}(i^*)}(\theta)\right]$
4. Return the sample: $X = x^*_{seed} + \rho \cdot \theta$

VoroSpokes Phase II: Sampling



VoroSpokes Phase II: Sampling



Parameter Estimation for Normal Distribution

$$Y = \{y_m\}_{m=1}^{M} \sim \mathcal{N}(\mu, \sigma)$$

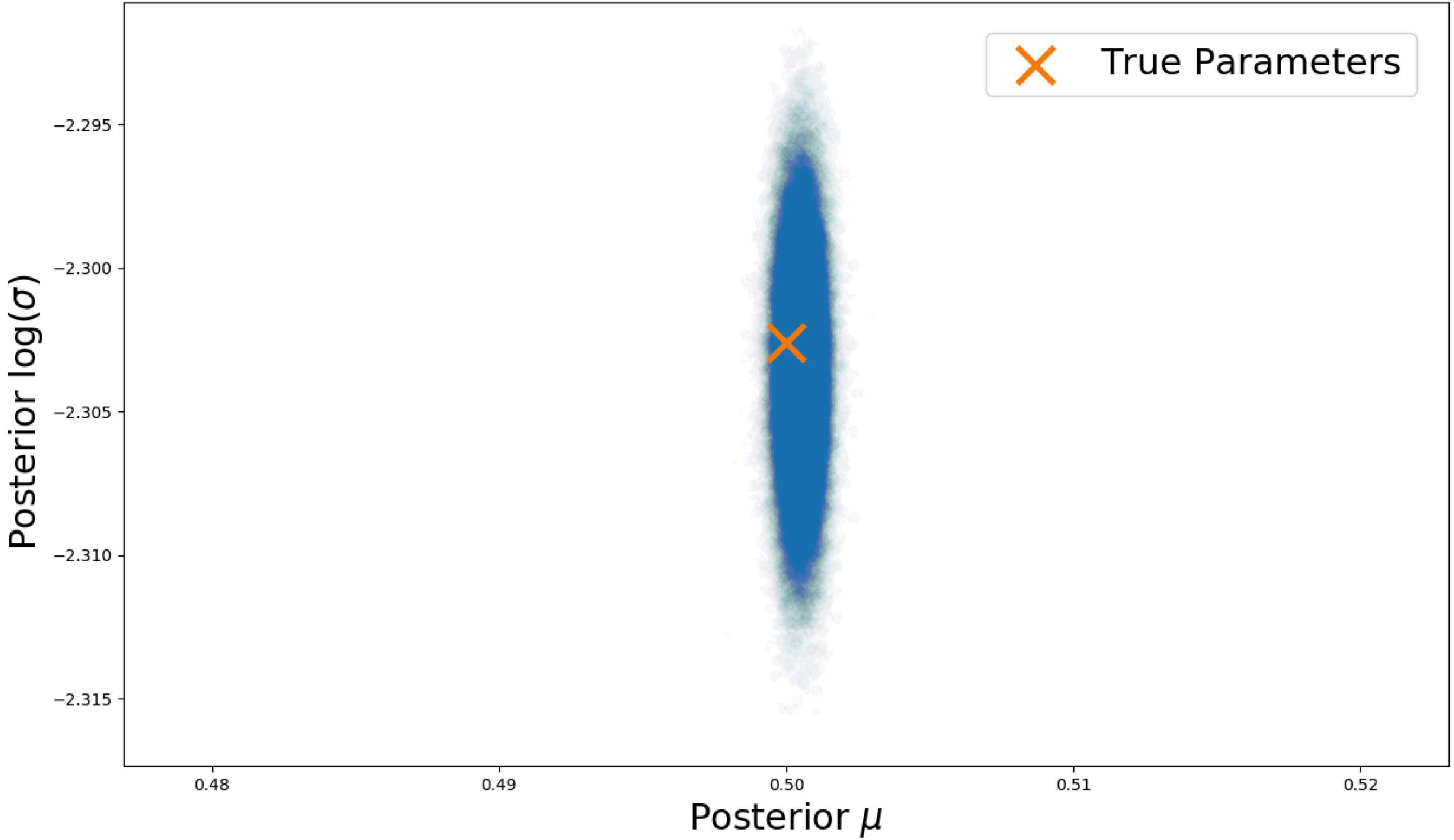
i.i.d. observations

$$\mu \sim [-1, 1]$$

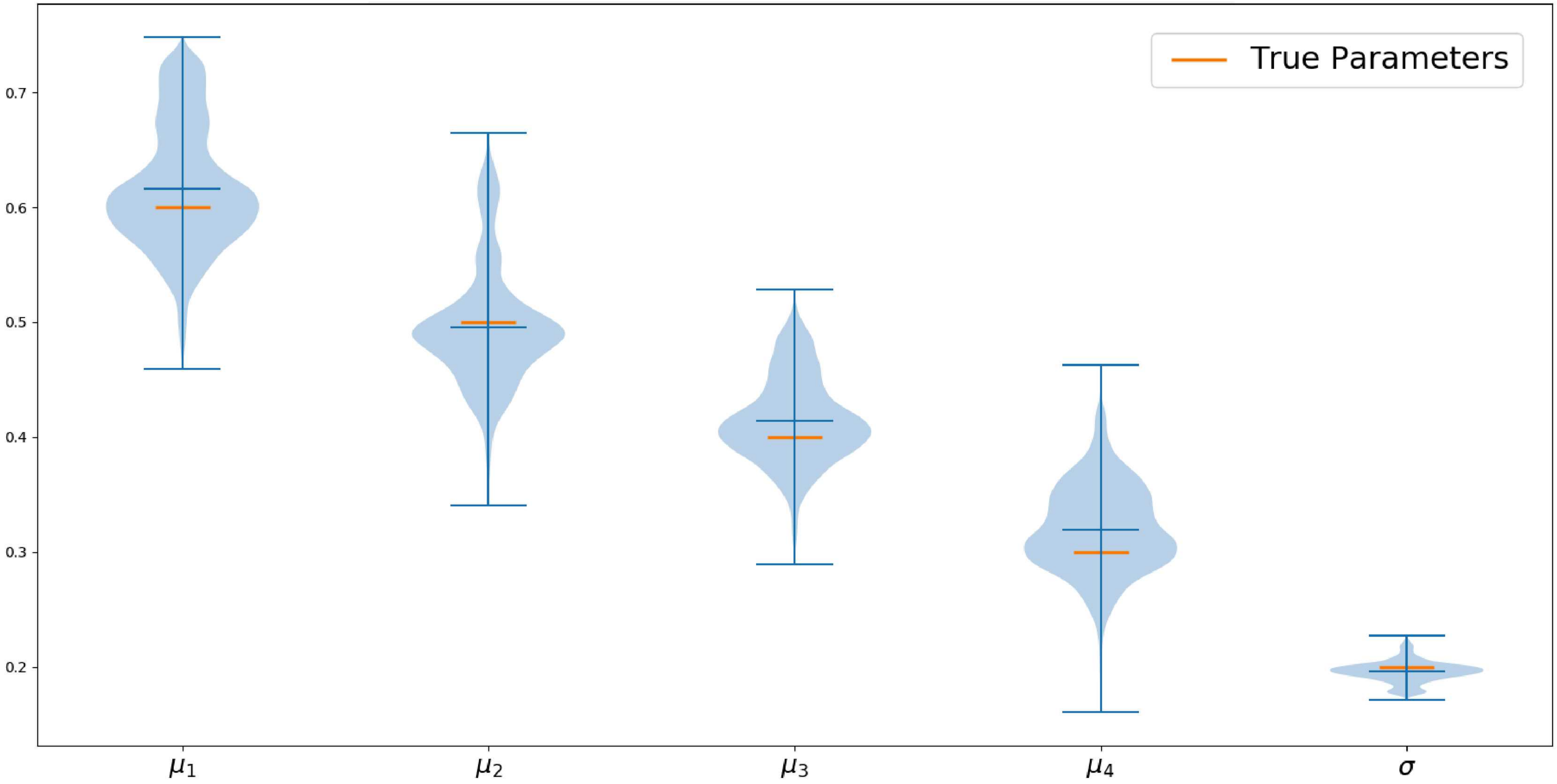
$$\sigma \sim [e^{-5}, 1]$$

$$f(\mu, \sigma \,|\, Y) \;\propto\; \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- |y_m - \mu|^2 \,/\, 2\sigma^2 \Big)$$

Parameter Estimation for Normal Distribution



Parameter Estimation for Multivariate Normal



Parameter Estimation for Mixture Model

$$\mathcal{N}(\mu^{(i)}, \Sigma^{(i)}) \quad \text{with} \quad \mu^{(i)} = [\mu_x^{(i)}, \mu_y^{(i)}] \ , \quad \Sigma^{(i)} = \begin{bmatrix} 1 & \rho^{(i)} \\ \rho^{(i)} & 1 \end{bmatrix} \quad \text{for} \quad i = 1, 2$$

$$p^{(i)}(y \,|\, \theta^{(i)}) = \frac{1}{2\pi} \frac{1}{\sqrt{\det(\Sigma^{(i)})}} \exp\Big(-\frac{1}{2} (y - \mu^{(i)})^T \left[\Sigma^{(i)}\right]^{-1} (y - \mu^{(i)}) \Big) \quad \text{where} \quad \theta^{(i)} = [\mu^{(i)}, \rho^{(i)}]$$

$$f\big(\theta^{(1)}, \theta^{(2)}, w \,|\, Y\big) \propto \prod_{m=1}^{M} \Big[w \cdot p^{(1)}(y_m \,|\, \theta^{(1)}) + (1 - w) \cdot p^{(2)}(y_m \,|\, \theta^{(2)}) \Big]$$

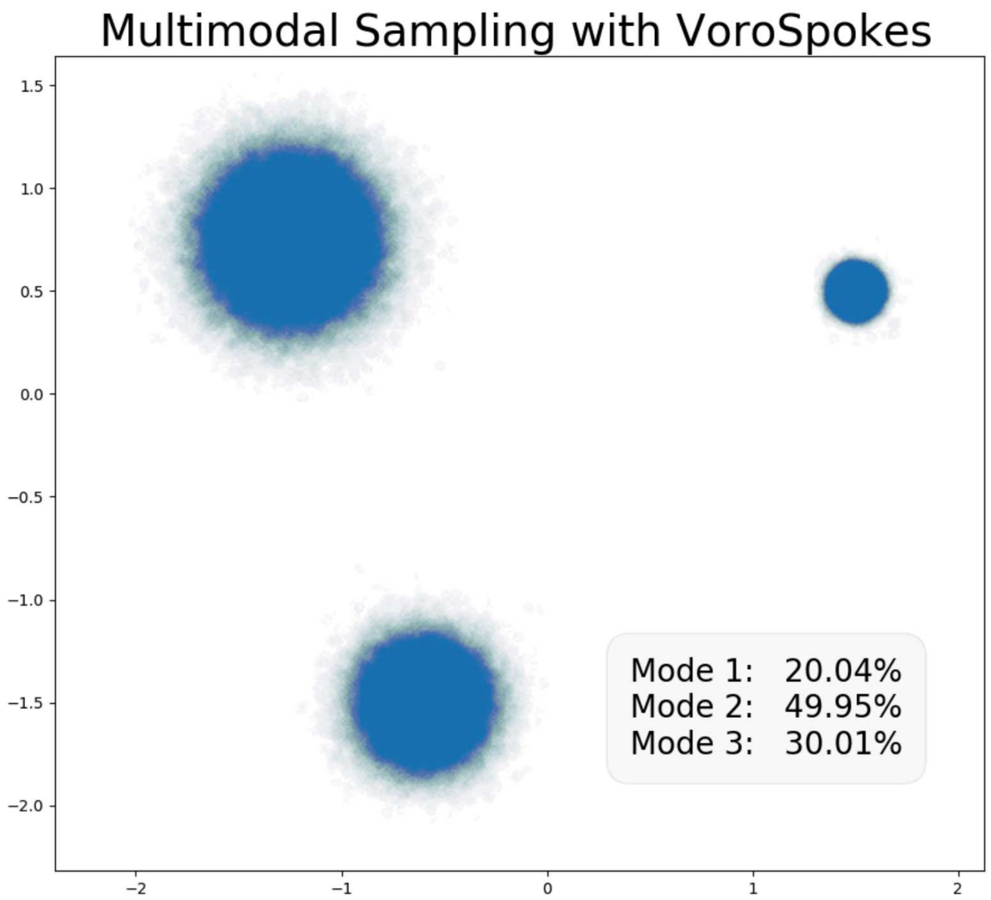
Parameter	$\mu_x^{(1)}$	$\mu_y^{(1)}$	$\rho^{(1)}$	$\mu_x^{(2)}$	$\mu_y^{(2)}$	$\rho^{(2)}$	w
True Value	2.5000	0.7500	0.5000	-2.5000	0.7500	-0.5000	0.7500
Estimate	2.4758	0.7824	0.4526	-2.5276	0.7599	-0.5356	0.7551

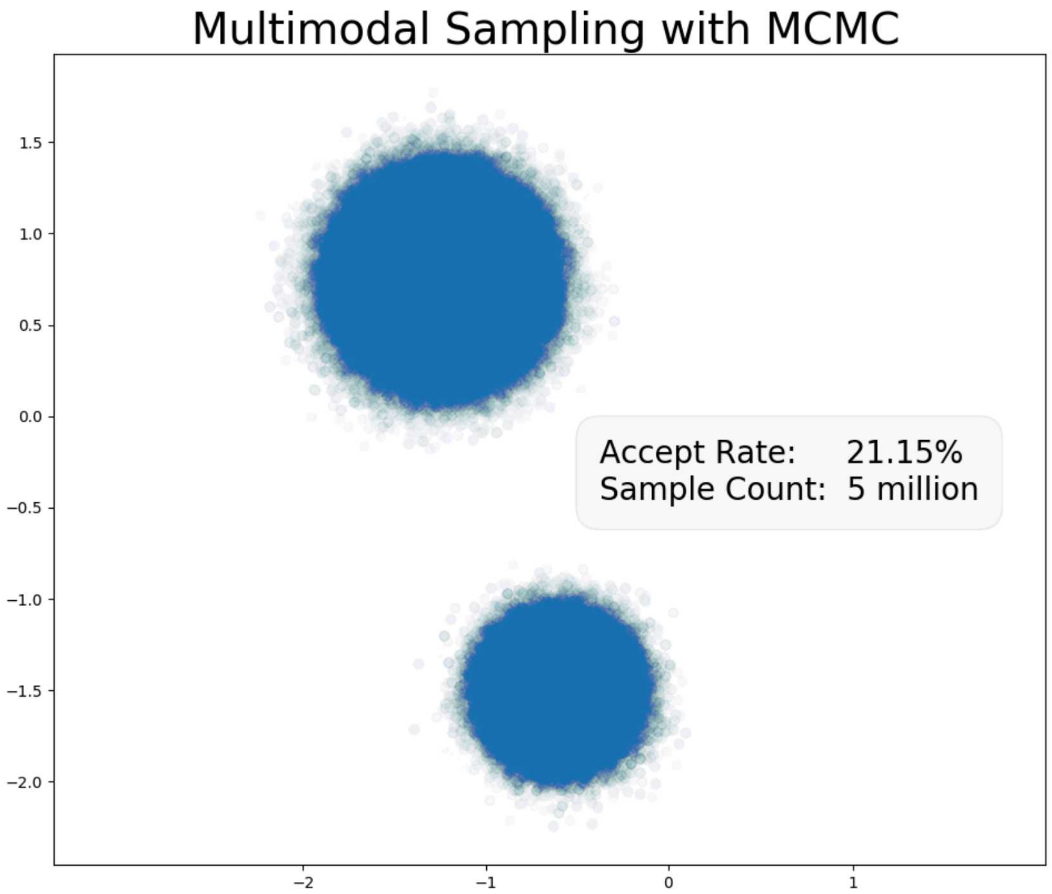
Sampling from Multimodal Distributions

$$f(x) \ \propto \ \sum_{m=1}^{3} \frac{w_m}{2\pi\sigma_m} \exp\left(- \frac{1}{2\sigma_m^2} \|x - \mu_m\|^2 \right)$$

$$\begin{aligned} w_1 &= 0.2 \ , & \mu_1 &= [1.5, 0.5] \ , & \sigma_1 &= 0.05 \\ w_2 &= 0.5 \ , & \mu_2 &= [-1.25, 0.75] \ , & \sigma_2 &= 0.20 \\ w_3 &= 0.3 \ , & \mu_3 &= [-0.6, -1.5] \ , & \sigma_3 &= 0.15 \end{aligned}$$

Sampling from Multimodal Distributions





Summary

- Local calculations within polytopes can be performed using randomly directed spokes; in particular, integral estimates and approximate samples can be generated.
- Voronoi tessellations provide a natural extension of the local polytope results to more general domains (since Voronoi cells partition a domain into convex polytopes).
- The global sampling procedure can be used to adaptively place function evaluations in regions with the largest estimated error; in particular, this can be used to adaptively construct surrogate models with a limited budget of function evaluations.
- Adaptive VPS surrogates can be used to approximate unnormalized probability density functions, and accurate samples can be generated independently and in parallel.

Questions?