VANDERBILT UNIVERSITY

High-dimensional node generation with variable density

Alex Vlasiuk

Fast Algorithms for
Generating Static and Dynamically Changing Point Configurations
ICERM
March 2018

Joint work with
N. Flyer, B. Fornberg, D. Hardin, T. Michaels, E. B. Saff

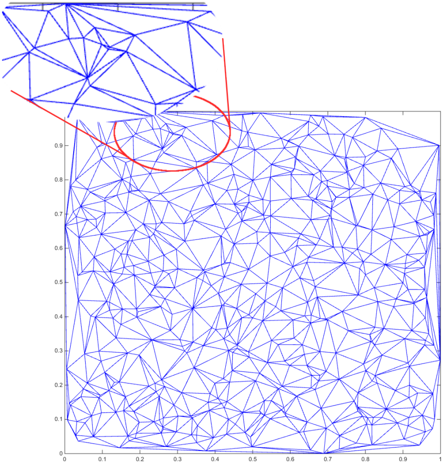
Why discretize?

- Model network and sensor deployment.
- Data: storing smooth manifold as a discrete configuration.
- PDE solvers, radial basis functions interpolation etc. need well-distributed nodes.
- Additionally, the meshless methods often require non-uniform nodes.

VANDERBILT UNIVERSITY

A bad example

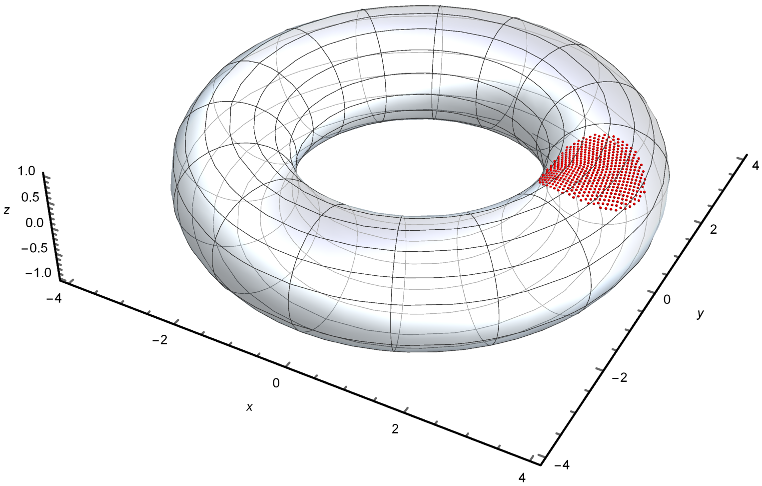
- (Uniform) random points exhibit clustering!



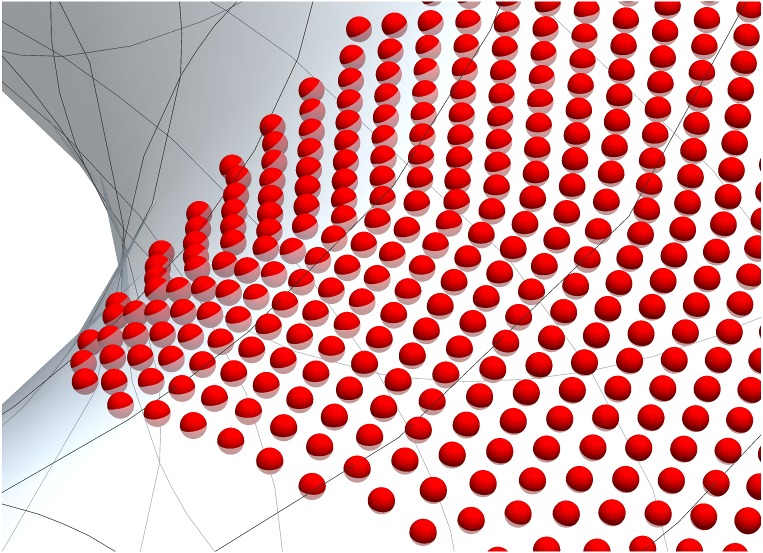
Delaunay triangulation of 500 uniformly random nodes in $[0,1]^2$.

A good example

Figure: 500 points, $s = 7,\ d = 2,\ q(x) = |x - x_0|^2$, where $x_0 := (3, 0, 1)^T$



Zoomed view



How to improve?

- sample from a distribution/process with repulsive properties
- apply thinning (to a Monte Carlo Markov chain)

  (still, neither of the above guarantees deterministic separation)

- Quasi-Monte Carlo methods
- Our suggestion: apply the gradient flow of a suitable functional to a suitable starting distribution

Weighted Riesz kernel

- $\mathcal{H}_d^A$ – normalized restriction of the d-dimensional Hausdorff measure
- $\omega_N = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset A$; $\tilde{\omega}_N$ – minimizers
- $E_s(\omega_N) := \sum_{\boldsymbol{x}\neq\boldsymbol{y}\in\omega_N} |\boldsymbol{x}-\boldsymbol{y}|^{-s}$, Riesz s-energy
- if $w(\cdot,\cdot)$ is $\mathcal{H}_d^A$-a.e. continuous on $A\times A$,

$$E_s(\omega_N; w) := \sum_{\substack{\boldsymbol{x}\neq\boldsymbol{y}\\ \boldsymbol{x},\boldsymbol{y}\in\omega_N}} \frac{w(\boldsymbol{x},\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|^s};$$

Theorem (Borodachov - Hardin - Saff, '14)

$s > d$ and $A \subset \mathbb{R}^p$ is d-rectifiable, compact, $\mathcal{H}_d(A) > 0$.
Every (asymptotically) minimal sequence $\{\tilde{\omega}_N\}_{N\geq 2}$ converges weak:*

$$\frac{1}{N}\sum_{\boldsymbol{x}\in\tilde{\omega}_N} \delta_{\boldsymbol{x}} \xrightarrow{*} \frac{1}{\mathcal{H}_d^{s,w}(A)} \cdot w^{-d/s}(\boldsymbol{x},\boldsymbol{x})\mathrm{d}\mathcal{H}_d^A \quad \text{as } N\to\infty.$$

($\mathcal{H}_d^{s,w}(A)$ is a normalizing constant) In particular, the asymptotics of the energy exists:

$$\lim_{N\to\infty} \frac{E_s(\tilde{\omega}_N; w)}{N^{1+s/d}} = \frac{C_{s,d}}{\mathcal{H}_d^{s,w}(A)^{s/d}}.$$

External field

- $q : A \to (\infty, \infty]$ – lower semi-continuous function
- For $s > d$, (s, d, q)-energy is

$$E_s(\omega_N; q) := \sum_{\substack{\boldsymbol{x} \neq \boldsymbol{y} \\ \boldsymbol{x}, \boldsymbol{y} \in \omega_N}} |\boldsymbol{x} - \boldsymbol{y}|^{-s} + N^{s/d} \sum_{\boldsymbol{x} \in \omega_N} q(\boldsymbol{x}).$$

- constant $M_{s,d} := (C_{s,d}(1 + s/d))^{-d/s}$

Theorem (Hardin - Saff - V, '16)

Take A, s as in the BHS-theorem. Every (asymptotically) minimal sequence $\{\tilde{\omega}_N\}_{N\geq 2}$ converges weak:*

$$\frac{1}{N}\sum_{\boldsymbol{x}\in\tilde{\omega}_N}\delta_{\boldsymbol{x}} \xrightarrow{*} M_{s,d}\,(L_1-q)_+^{d/s}\,\mathrm{d}\mathscr{H}_d^A =: \mathrm{d}\mu_q \quad \textit{as } N\to\infty.$$

In particular, the asymptotics of the energy exists:

$$\lim_{N\to\infty}\frac{E_s(\tilde{\omega}_N;q)}{N^{1+s/d}} = \mathfrak{S}(q,A),$$

where

$$\mathfrak{S}(q,A) := \int \frac{L_1 d + q(\boldsymbol{x})s}{d+s}\,\mathrm{d}\mu_q(\boldsymbol{x}).$$

The $L_1 = L_1(q,A)$ is the (unique) constant such that $\mathrm{d}\mu_q$ is a probability measure on A.

VANDERBILT UNIVERSITY

(Re)Producing a distribution

Given an upper semi-continuous $\rho : A \to [0, \infty)$ such that $\rho \, \mathrm{d}\mathscr{H}_d^A$ – probability measure, define

$$w(\boldsymbol{x}, \boldsymbol{y}) := (\rho(\boldsymbol{x})\rho(\boldsymbol{y}) + |\boldsymbol{x} - \boldsymbol{y}|)^{-s/2d}, \quad \text{and} \quad q(\boldsymbol{x}) := -\left(\frac{\rho(\boldsymbol{x})}{M_{s,d}}\right)^{s/d}.$$

Any sequence $\{\tilde{\omega}_N\}_{N\geq 2}$ minimizing either

$$E_s(\omega_N; w) = \sum_{\substack{\boldsymbol{x} \neq \boldsymbol{y} \\ \boldsymbol{x}, \boldsymbol{y} \in \omega_N}} \frac{w(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x} - \boldsymbol{y}|^s}$$

or

$$E_s(\omega_N; q) = \sum_{\substack{\boldsymbol{x} \neq \boldsymbol{y} \\ \boldsymbol{x}, \boldsymbol{y} \in \omega_N}} |\boldsymbol{x} - \boldsymbol{y}|^{-s} + N^{s/d} \sum_{\boldsymbol{x} \in \omega_N} q(\boldsymbol{x}).$$

converges to $\rho \, \mathrm{d}\mathscr{H}_d^A$, $N \to \infty$.

Kernel truncation

- Turns out, one may assume $w(\cdot,\cdot) = w_N(\cdot,\cdot)$ satisfies

$$w(\boldsymbol{x},\boldsymbol{y}) = 0, \quad \text{if} \quad |\boldsymbol{x}-\boldsymbol{y}| > r_N,$$

  where

$$r_N \to 0 \quad \text{so that} \quad r_N N^{1/d} \to \infty.$$

  that is, outside an r_N-neighborhood of $\operatorname{diag}(A \times A)$
- this does not change the limiting measure
- we call this modification the truncated Riesz kernel.
- applies to the $E(\omega_N; q)$ energy as well
- in practice, for a fixed range of N it suffices only to take into account several nearest neighbors

Family of energies

$$E_s(\omega_N; w, q) = \sum_{\substack{\boldsymbol{x} \neq \boldsymbol{y} \\ \boldsymbol{x}, \boldsymbol{y} \in \omega_N}} \frac{w(\boldsymbol{x}, \boldsymbol{y})}{|\boldsymbol{x} - \boldsymbol{y}|^s} + N^{s/d} \sum_{\boldsymbol{x} \in \omega_N} q(\boldsymbol{x}).$$

- flexibility in choosing the functional
- the cost of evaluation of the density defines the costs of w and q
- local forces in w vs global in q

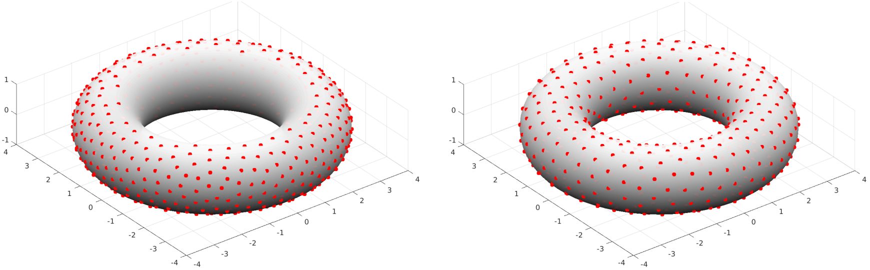
Short-scale properties of minimizers

- Optimal order of minimal pairwise distances, $N^{-1/d}$, for $s > d$.
- Under mild smoothness assumptions minimizers have the optimal covering radius of order $N^{-1/d}$ on any sublevel set $A(u)$ for all $u < L_1$.
- For $d \geq 2$, and $s > d$, the value of constant $C_{s,d}$ is known only numerically. Still, the distribution is stable under small perturbations of the value of $C_{s,d}$.

VANDERBILT UNIVERSITY

Restriction $s > d$ is important

Let A – torus, $\dim A = 2$,



are 500-point approximate minimizers. Left: $s = 0.5$; right: $s = 4$.
This is an artifact of using the ambient, not geodesic, distance.

VANDERBILT UNIVERSITY

Gradient dynamics and its initialization

- Picking a random starting position slows down the optimization.
- The hypersingular kernel is short-ranged; let us try a locally suitable starting set, then apply minimization.
- For Monte Carlo methods piecewise distribution generation: *stratification*.
- Since we prohibit clustering, let's use quasi-Monte Carlo on individual pieces.

Gradient flow

- $\boldsymbol{x}_{j(i,k)}^{(t)}$ – nearest neighbors to $\boldsymbol{x}_i^{(t)}$, $1 \leq k \leq K$; $\Delta\left(\boldsymbol{x}_i^{(t)};\ \omega_N^{(t)}\right)$ – distance to the nearest neighbor
- Perform T iterations, moving in the direction of vector $\boldsymbol{g}_i^{(t)}$:

$$\boldsymbol{g}_i^{(t)} = -\nabla_i E_s\left(\left\{\boldsymbol{x}_{j(i,k)}^{(t)}\right\};w\right) \quad 1 \leq i \leq N.$$

- $C > 0$, constant

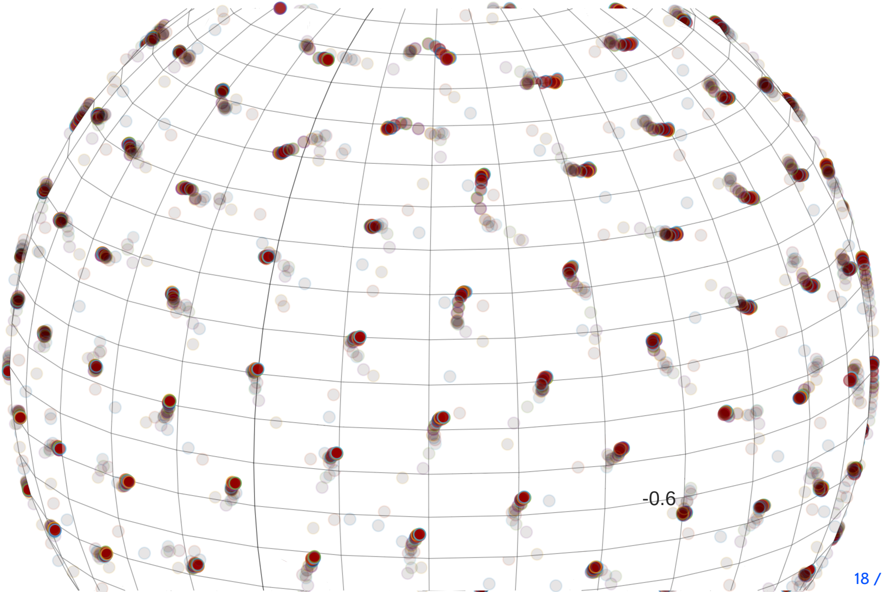
$$\boldsymbol{x}_i^{(t+1)} = \begin{cases} \boldsymbol{x}_i^{(t)} + \dfrac{1}{t+C} \cdot \Delta\left(\boldsymbol{x}_i^{(t)};\ \omega_N^{(t)}\right) \cdot \dfrac{\boldsymbol{g}_i^{(t)}}{\|\boldsymbol{g}_i^{(t)}\|} & \text{if this sum is inside } A; \\ \boldsymbol{x}_i^{(t)}, & \text{otherwise.} \end{cases}$$

- $\approx$ truncated Langevin dynamics $\implies$ considered in the math.phys. community, Chafaï et al, Duerinckx, Serfaty, etc.

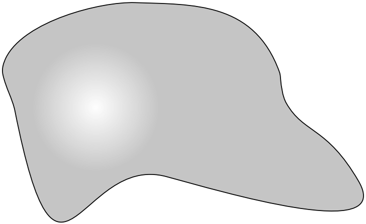
VANDERBILT UNIVERSITY

Dynamics illustration

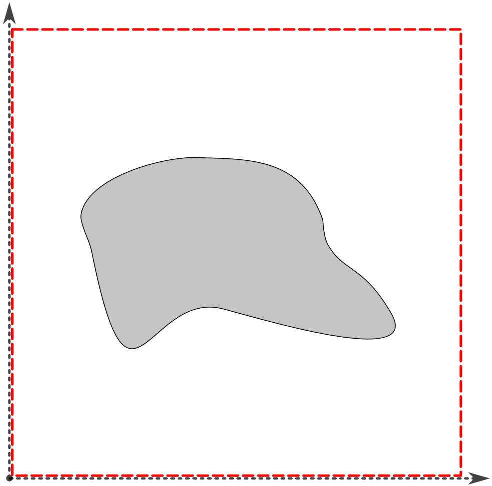
200 spherical points



Algorithm (Flyer - Fornberg - Michaels - V, '17)

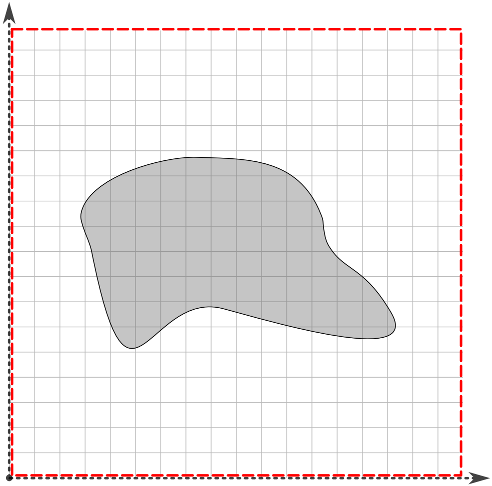


Algorithm (Flyer - Fornberg - Michaels - V, '17)



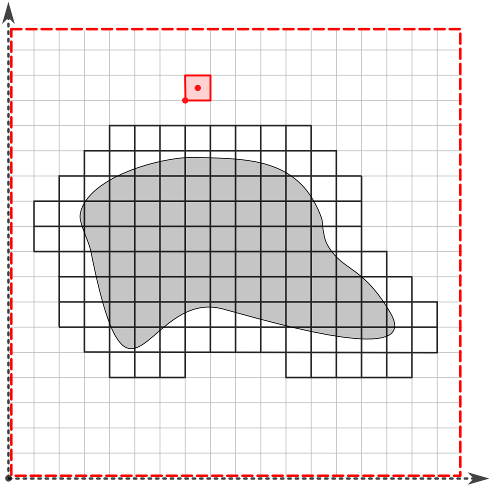
Algorithm (Flyer - Fornberg - Michaels - V, '17)

- partition the domain using a uniform (or adaptive) grid



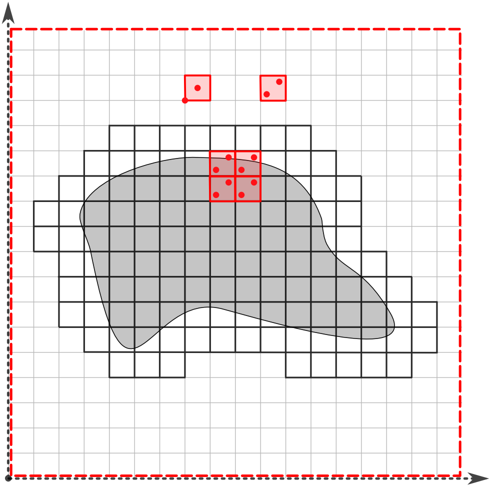
Algorithm (Flyer - Fornberg - Michaels - V, ’17)

- detect support; use only the cells close to it



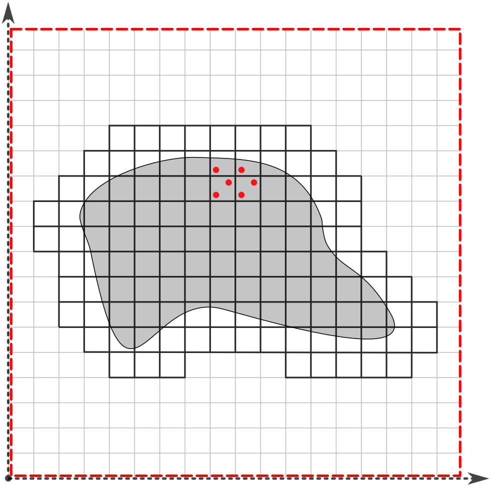
Algorithm (Flyer - Fornberg - Michaels - V, '17)

- place rescaled/translated/otherwise adapted pieces in each cell



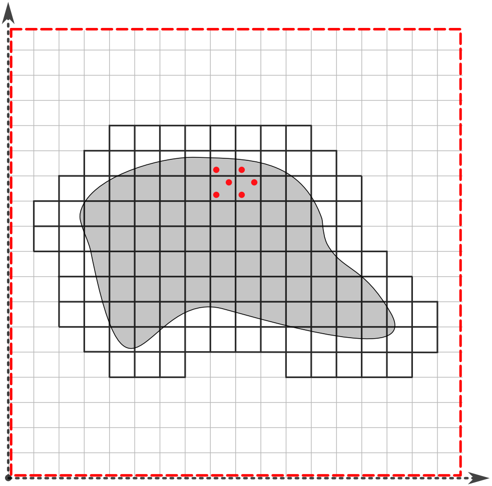
Algorithm (Flyer - Fornberg - Michaels - V, '17)

- make sure no points are outside the support



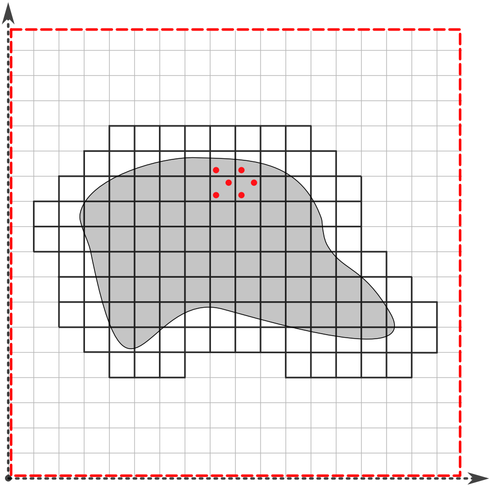
Algorithm (Flyer - Fornberg - Michaels - V, '17)

- apply the gradient flow, weighted with the desired density



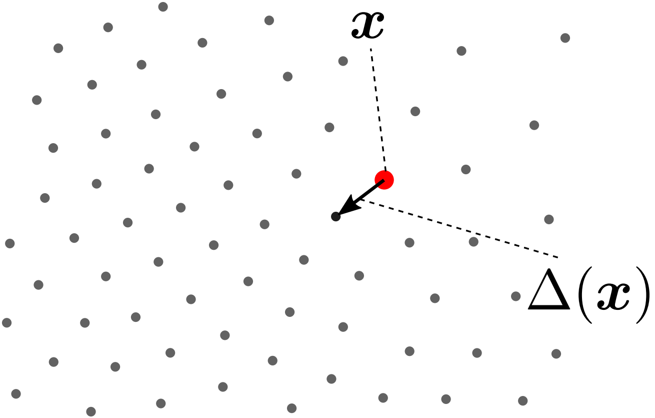
- fill the sparser parts of the distribution by *saturation*

(’greedy’ procedure, conceptually similar to bubble packing by Shimada)



Modified question

Distance from a node to the nearest neighbor has to be approximately equal to a given function of its position: $\Delta(\boldsymbol{x}) \approx \rho(\boldsymbol{x})$



We call ρ the *radial density*. Note: it is Lipschitz-1.

Modified algorithm

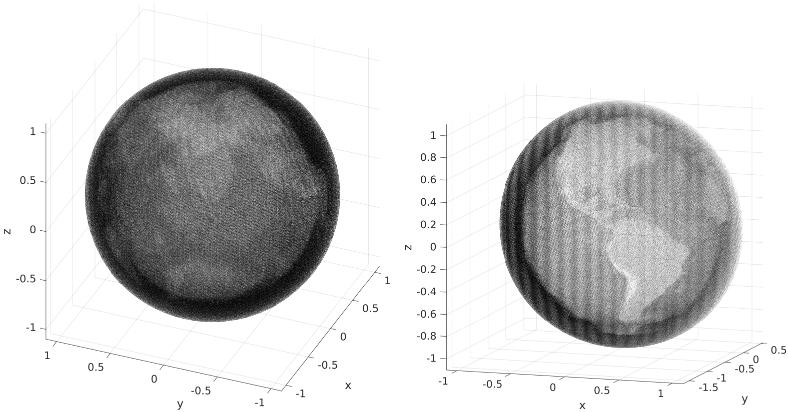
...

- place rescaled/translated/otherwise adapted pieces in each cell
    1. tabulate separation for a pre-determined node sequence (Riesz minimizers or ILs)
    2. use appropriate number of nodes, according to the desired separation
    3. transition between the multiplicative weight and radial density is governed by $w \asymp \Delta^s,\ N \to \infty$

...

Earth surface

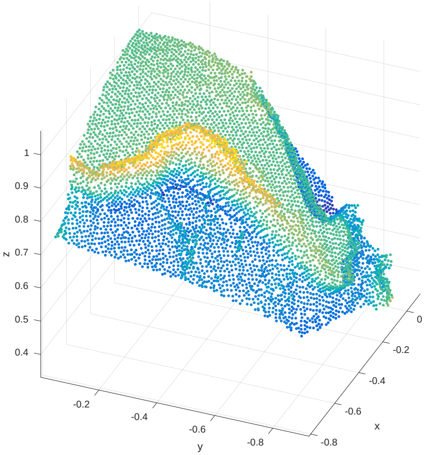
- Goal: regularity on complex surfaces for the uniform density.
  An atmospheric layer with faithful surface recovery:

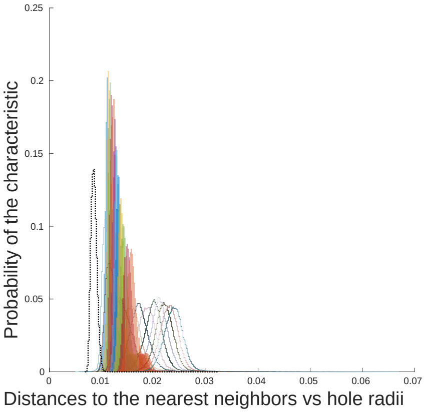


$\approx 1.35M$ nodes, generated in ≈ 3 minutes using ETOPO1 surface data and ray-tracing inclusion algorithm. The $\mathscr{L}_n$ lattice parameters $\alpha_1 = \sqrt{2}$, $\alpha_2 = (\sqrt{5}-1)/\sqrt{2}$.

VANDERBILT UNIVERSITY

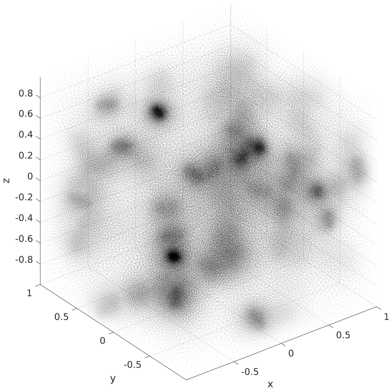
Earth surface: Andes

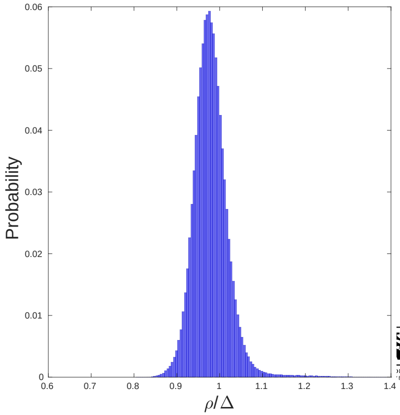




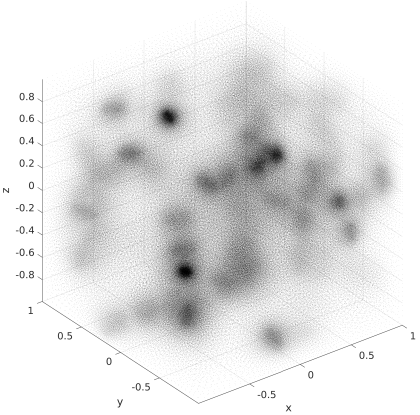
Point cloud

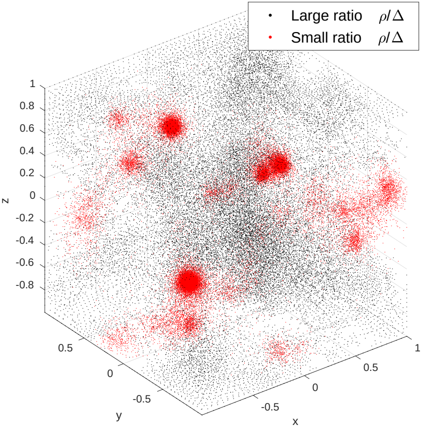
- Random 100 points $\mathscr{P}_{100}$ inside $[-1,1]^3$. Consider the radial density:
$$\rho(\boldsymbol{x}) = \left(\Delta(\boldsymbol{x};\mathscr{P}_{100}) + \Delta^2(\boldsymbol{x};\mathscr{P}_{100})\right)/20,$$
where Δ^2 for the distance to the 2-nd nearest neighbor.
- Goal: density recovery. Output: Left: 577,321 nodes; 200 interations of flow stepping in $\approx$ 12 minutes . Right: ratio of target/actual densities.





Point cloud: error location



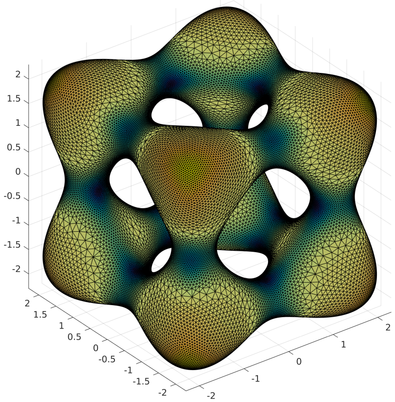


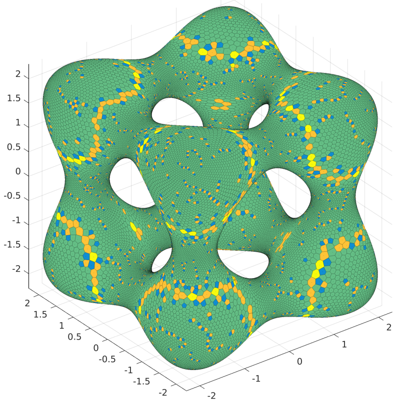
Surface Voronoi

- Chmutov-Banchoff-type surface:

$$x^2(x^2-5)+y^2(y^2-5)+z^2(z^2-5)+11=0$$

- 40K points distributed according to the absolute value of the Gaussian curvature. Left: color-coded Gaussian curvature, blue/orange is lower/higher. Right: surface Voronoi diagram.





Candidates for Quasi-Monte Carlo initialization

- irrational lattices (easy to generate, scalable)
  Pick $\alpha_1, \ldots, \alpha_{d-1}$ linearly independent over $\mathbb{Q}$, fix $0 < \delta < 1$

$$\mathscr{L}_n := \left\{ \left(\left\{ \delta + \frac{i}{n} \right\}, \{i\alpha_1\}, \{i\alpha_2\}, \ldots, \{i\alpha_{d-1}\} \right) \right\}_{i=1}^{n}$$

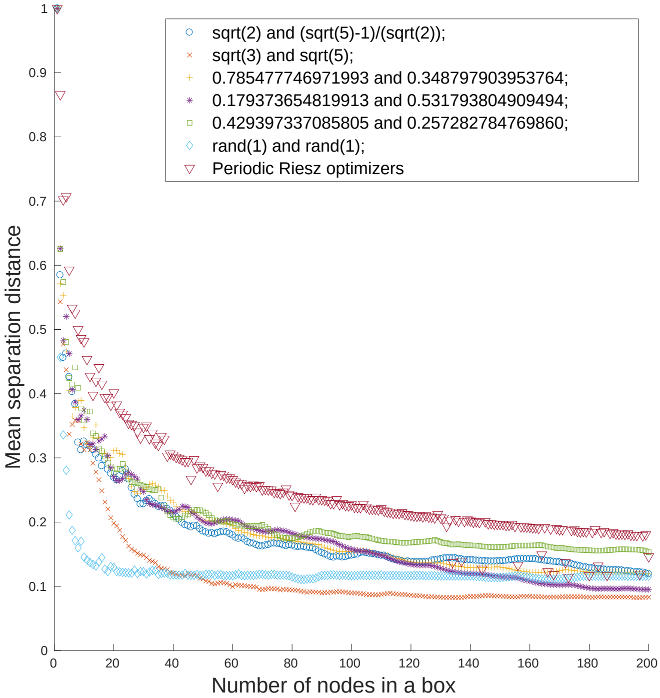
  $\{x\} \equiv x \bmod 1.$

  1. $\mathscr{L}_n$ weakly converge to the uniform distribution
  2. pointwise separation depends on the "irrationality properties" of $\alpha_1, \ldots, \alpha_{d-1}$, apparently on the continued fraction approximation
  3. *Korobov point sets* for the Q-MC community; *Kronecker sequences* for number theorists; *IL* is the low-discrepancy term.

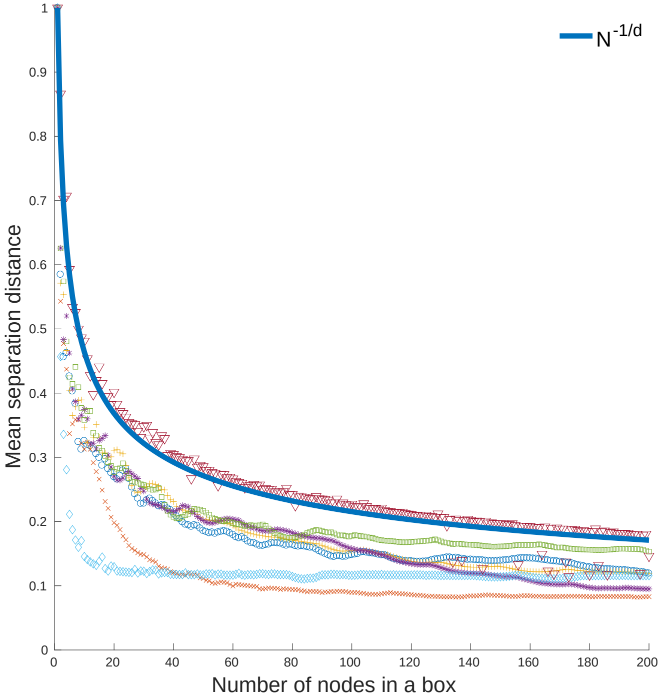
- periodic Riesz minimizers (optimal filling)

VANDERBILT UNIVERSITY

Separation of ILs: a curiosity



VANDERBILT UNIVERSITY



VANDERBILT UNIVERSITY

Implementation

- Matlab prototype; uses the default knn.
  - https://github.com/OVlasiuk/3dRBFnodes
  - https://github.com/OVlasiuk/BRieszk
- Efficient for small to medium scales.

Conclusion

- Riesz energy-based functionals for construction of sets with a predefined density, volumetric and related
- suitable for meshless methods
- parallelizable
- reliably attains optimal separation
- practically suitable for tessellating (2d surfaces)
- (almost) dimension-agnostic
- allows modest (in terms of Wasserstein distance) distribution updates

- on large scales the singularity can cause precision loss, mitigated by smoothing
- relies on finding nearest neighbors

VANDERBILT UNIVERSITY

Thank you!

- D. P. Hardin, E. B. Saff and O. V. *Generating point configurations via hypersingular Riesz energy with an external field*, SIAM J. Math. Anal., 49(1), 646-673, 2017
- S. V. Borodachov, D. P. Hardin, and E. B. Saff. *Low Complexity Methods For Discretizing Manifolds Via Riesz Energy Minimization*. Found. Comput. Math., 14, 2014
- K. Shimada, D. C. Gossard, *Bubble mesh*, ACM SMA '95 (pp. 409–419). NY: ACM Press.
- C. Beltrán, J. Marzo, and J. Ortega-Cerdà. *Energy and discrepancy of rotationally invariant determinantal point processes in high dimensional spheres*, J. Complexity, 37, 76-109, 2016
- N. Flyer, B. Fornberg, T. Michaels, and O. V., *Fast high-dimensional node generation with variable density, submitted*.