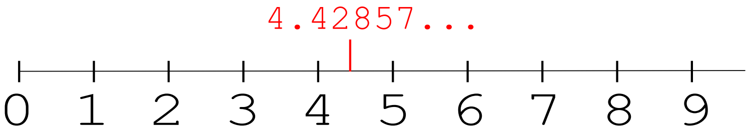
On Near Optimal Lattice Quantization of Multi-Dimensional Data Points

Manuel Finckh, Holger Dammertz, Hendrik P. A. Lensch

7. Mai 2015

Quantization

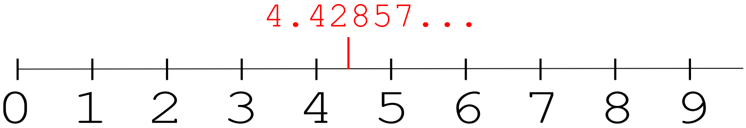
- analog to digital conversion
  - recording
  - photography
  - digital measuring
  - convert to numbers the computer can handle
  - `byte, int, float, ...`
- reduce precision
  - use less bits for each datum
  - reduce size of data (compression)



Lattices

A *lattice* is a discrete additive subgroup of $\mathbb{R}^s$, i.e. a subset $\Lambda \subseteq \mathbb{R}^s$ which satisfies:

- subgroup
  - Λ is closed under addition and subtraction
- discrete
  - there is an $\epsilon > 0$ such that two distinct lattice points $x \neq y \in \Lambda$ are at distance at least $\| x - y \| \geq \epsilon$



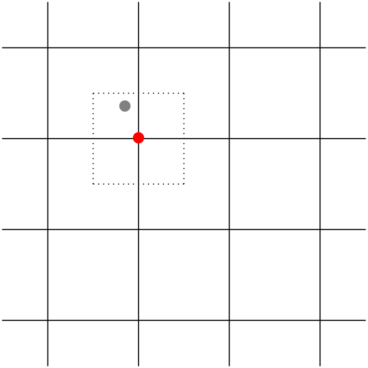
Quantization - Goals

Predefined: target bit depths b (fixed, but arbitrary)

Input: s-dimensional vector $\mathbf{v} \in [0, 1)^s$
Output: index i with b-bits

- controllable number of bits
- low quantization error
- fast indexing
  - lattice-based quantization
  - no data-dependent vector quantization

Lattice Quantization: Quantization Error



Quantization error:

- all Voronoi cells are congruent
- place origin at centroid of Voronoi cell

$$G(\Lambda) = \frac{\frac{1}{s}\int_{\Pi} x \cdot x \; dx}{Vol(\Pi) \cdot Vol(\Pi)^{\frac{2}{s}}}$$

- find an s-dimensional lattice Λ for which $G(\Lambda)$ is a minimum.

Quantization Error of some Lattices

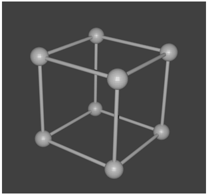
- Gain of optimal lattices increases with dimension

dim	Optimal Lattice	$G(\Pi)$	Cartesian Grid	$G(\Pi)$
1	$\mathbb{Z}$	0.083333	$\mathbb{Z}$	0.083333
2	hexagonal	0.080188	$\mathbb{Z}^2$	0.083333
3	BCC	0.078543	$\mathbb{Z}^3$	0.083333
4	D_4	0.076603	$\mathbb{Z}^4$	0.083333
5	D_5^*	0.075625	$\mathbb{Z}^5$	0.083333
6	E_6^*	0.074244	$\mathbb{Z}^6$	0.083333

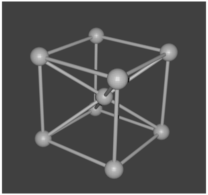
- Conway, Sloane 2010: Sphere Packings, Lattices and Groups

Encoding Lattice Points

$\mathbb{Z}^3$



body centered cubic



$\mathbb{Z}^3$:

- n points per dimensions $\rightarrow N = n^3$

BCC:

- construction: union of two $\mathbb{Z}^3$ lattices
- second one shifted to body center of first one
- $N = 2 \cdot n^3$

Number of points N most often not a power of 2!

Goal: controllable bit depth

Need a different type of lattice: rank-1 lattices

Rank-1 Lattices, Korobov 1959

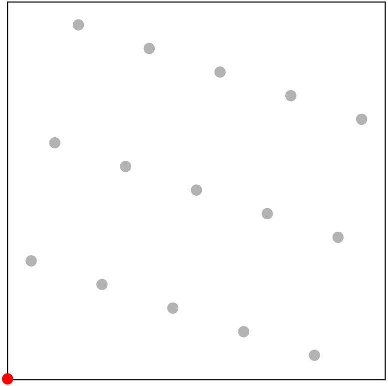
Given:

- number n of points and dimension s
- generator vector $\mathbf{g} = g_1, g_2, ..., g_s \in \mathbb{N}^s$
- then lattice point $\mathbf{x_i}, i = 0, ..., n-1$ are

$$\mathbf{x_i} := \frac{1}{n}(i \cdot \mathbf{g} \mod n)$$

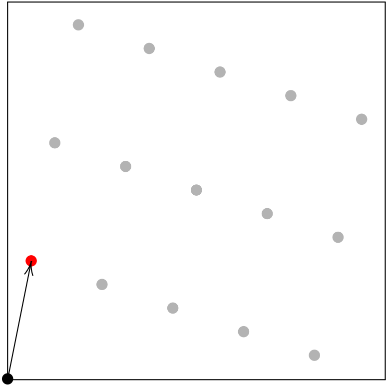
- all $\mathbf{x_i}$ are inside unit cube $[0, 1)^s$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



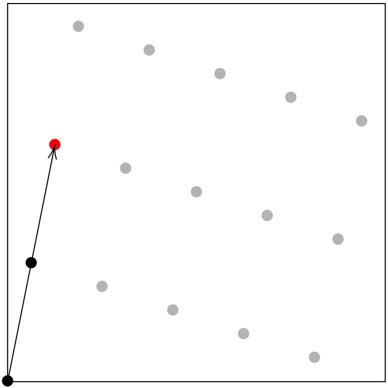
$$\mathbf{x}_0 = \frac{1}{16}(0 \cdot (1,5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



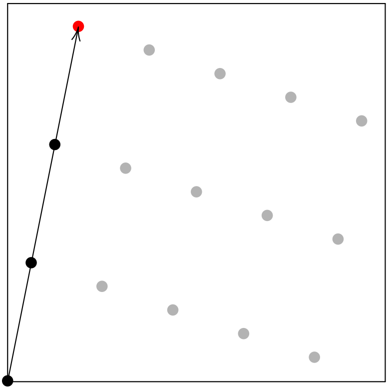
$$\mathbf{x}_1 = \frac{1}{16}(1 \cdot (1,5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



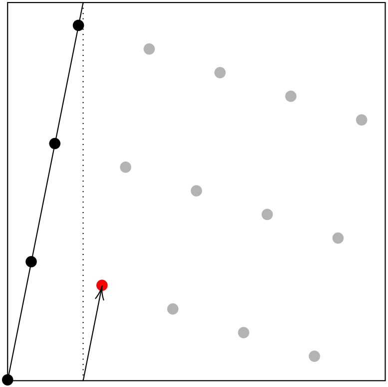
$$\mathbf{x}_2 = \frac{1}{16}(2 \cdot (1,5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



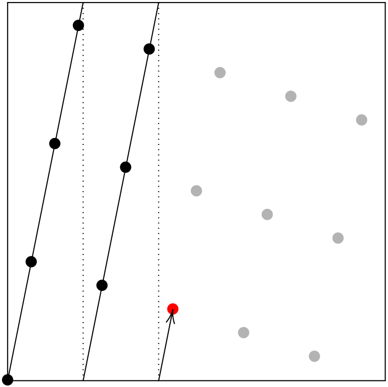
$\mathbf{x}_{3} = \frac{1}{16}(3 \cdot (1,5) \mod 16)$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



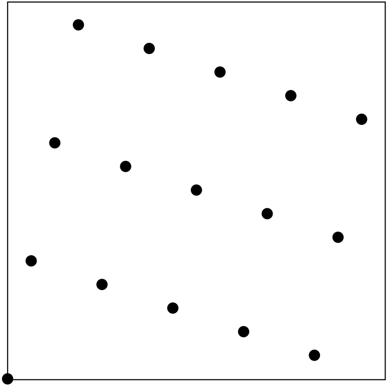
$$\mathbf{x}_4 = \frac{1}{16}(4 \cdot (1,5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



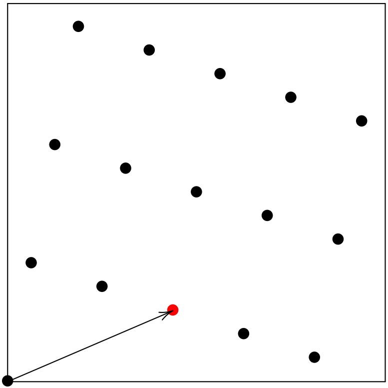
$$\mathbf{x}_{7} = \frac{1}{16}(7 \cdot (1, 5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(1,5)}$



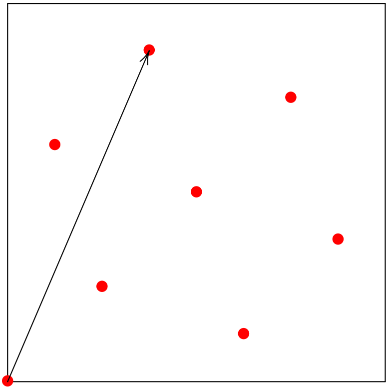
$$\mathbf{x_i} = \frac{1}{16}(i \cdot (1, 5) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(7,3)}$



$$\mathbf{x_i} = \frac{1}{16}(i \cdot (7,3) \mod 16)$$

Rank-1 Lattices: Examples $L_{16,(6,14)}$



$$\mathbf{x_i} = \frac{1}{16}(i \cdot (6, 14) \mod 16)$$

Rank-1 Lattices: Generator Vector

Generator vector $\mathbf{g} = (g_1, g_2, ..., g_s)$ meets the condition

$$\mathbf{gcd}(g_1, g_2, ..., g_s, n) = 1$$

where $\mathbf{gcd}(\cdot)$ is the greatest common divisor of all the arguments.

- Rank-1 lattices exist for any number of point in $I^s = [0, 1)^s$

Rank-1 Lattices: Finding Parameters

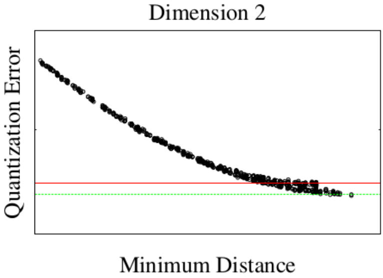
- find $\mathbf{g} = (g_1, ..., g_s)$ that minimizes quantization error
- Quantization error $G(\Lambda)$ costly to compute
  - analytic integration gets costly in higher dimensions
  - Monto Carlo noisy, needs lots of samples

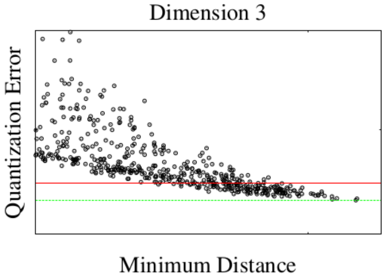
Instead use simple criteria and show that it works

- Maximize the minimum distance between lattice points
  - works for sampling efficiency (Dammertz 2009)
  - up to dimension 3 the best sampling lattices are also the best quantization lattices
    1. Integer lattice $\mathbb{Z}$
    2. Hexagonal lattice
    3. Body centered cubic lattice (BCC)

Rank-1 Lattices: Quantization and Minimum Distance

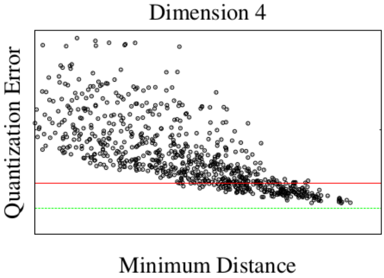
- generate random generator vectors
- plot quantization error vs. minimum distance
  - maximizing minimum distance minimizes quantization error

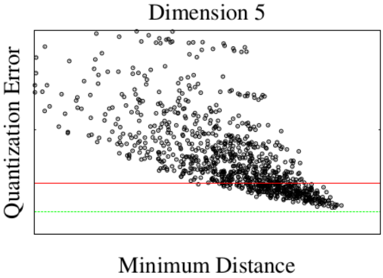




Rank-1 Lattices: Quantization and Minimum Distance

- generate random generator vectors
- plot quantization error vs. minimum distance
  - maximizing minimum distance minimizes quantization error





Rank-1 Lattice: Minimum Distance

Lattice Basis

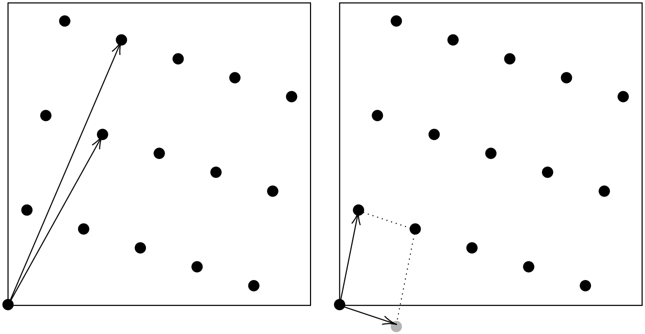
Each point of an s-dimensional lattice can be represented as integer linear combination of s linear independent basis vectors $\mathbf{b_1}, ..., \mathbf{b_s}$:

$$\mathbf{x} = \sum_{i=1}^{s} \xi_i \mathbf{b_i}$$

where $\xi = (\xi_1, ..., \xi_s)$ is an arbitrary integer vector.

- determine reduced basis
  - basis with shortes basis vectors
- shortest vector in reduced basis $\rightarrow$ minimum distance
- basis further used for quantization

Rank-1 Lattice: Basis Reduction



- reduced basis spans fundamental parallelepiped

Rank-1 Lattice: Basis Reduction

Given lattice basis $\mathbf{b_1}, ..., \mathbf{b_s}$:

- apply basis reduction algorithm to find the shortest vector
  - approximations do not work (like LLL algorithm)
  - brute force too costly
  - use greedy reduction algorithm, Nguyen, Stehle 2004
    - exact for dimensions up to 6
    - efficient up to dimension 4, feasible for dimension 5 and 6

Basic idea of the algorithm:

- order basis $[\mathbf{b_1}, ..., \mathbf{b_s}]_{\leq}$
- greedily reduce $\mathbf{b_k}$ with $L[\mathbf{b_1}, ..., \mathbf{b_{k-1}}]$
- start with $k = 2$
- no length reduction: set $k = k + 1$
- shorter vector found: insert into ordered basis

Parameter Search

1. Choose generator vector (enumerate or at random)
2. Construct initial lattice basis
3. Apply basis reduction algorithm
4. $\rightarrow$ shortest vector in rank-1 lattice
5. goto step 1. until "good" lattice found

Quantization with Rank-1 Lattices

Vector $\mathbf{v} \in [0,1)^s$

1. Solve linear equation system

$$\mathbf{v} = \xi[\mathbf{b_1}, ..., \mathbf{b_k}]$$

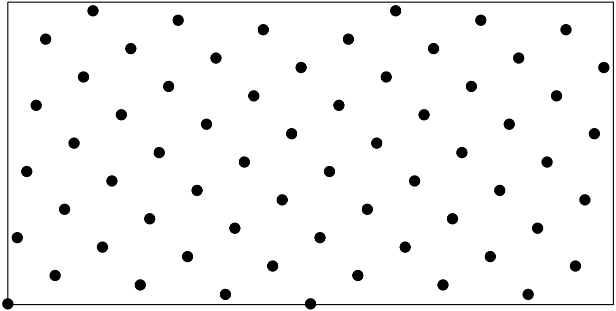
2. clamp coefficients to integer values
   - Cartesian coordinates of lattice point $\mathbf{p}$
   - $\mathbf{p}$ anchor point of parallelepiped

$$(\xi + \Delta)[\mathbf{b_1}, ..., \mathbf{b_k}], \Delta \in [0,1)^s$$

3. enumerate corner points to find $\mathbf{p_{near}}$
   - enumerate also neighboring points for dimension > 4

Rectangular Region instead of Unit Cube

- optimize minimum distance in scaled domain $[0,1)^s \to [0,x_1]) \times \ldots \times [0,x_s)$
- scale basis before basis reduction



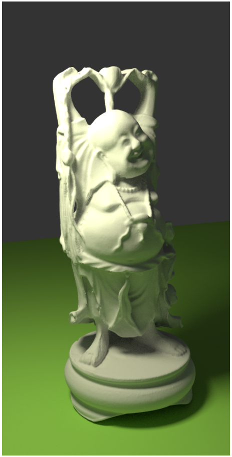
Practical Application of Rank-1 Lattice Quantization

1. Target bit-depth and dimension of input data
2. Find generator vector
   - search
   - look-up
3. Quantization
   - find index i for each datum
4. Reconstruction

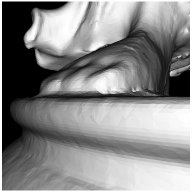
$$d_i = \frac{1}{n}(i \cdot \mathbf{g} \mod n)$$

Applications: Mesh Vertex Quantization

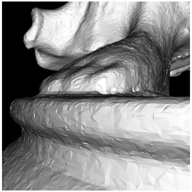
- Stanford Buddha vertices quantized to 32-bits per vertex
- Quantization domain: bounding box (AABB)



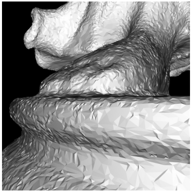
Original
RMSE: –



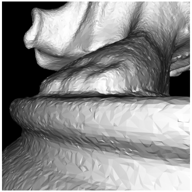
$10 + 12 + 10$-bit grid, AABB
RMSE: $1.029 \cdot 10^{-4}$



32-bit rank-1 lattice $[0, 1)^3$
RMSE: $1.726 \cdot 10^{-4}$



32-bits rank-1 lattice, AABB
RMSE: $0.956 \cdot 10^{-4}$

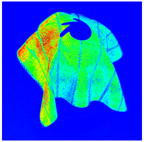


Applications: Spectral Reflectance Quantization

- Example: spectral texture with 6D data per texel

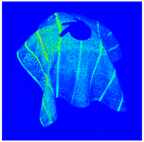
reference

6×4-*bit* regular



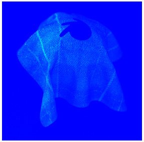
0.0210

24-*bit* rank-1



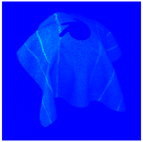
0.0125

6×6-*bit* regular



0.00422

36-*bit* rank-1



0.00415

Conclusion

- Arbitrary bit-depth b, independent of dimension
- Close to optimal quantization error
- Fast reconstruction, independent of dimension
- Quantization is
  - very fast: dimensions 2 to 4
  - practical: dimensions 5 and 6

Limitations

- Exponential complexity of basis reduction in higher dimensions
  - basis reduction algorithms for higher dimensions
  - brute-force basis reduction for small bit-depths (up to 32-bits)