PECULIARITIES OF LARGE DIMENSIONS
and some repercussions

Anatoly Zhigljavsky

Cardiff University, MATHS

Cardiff, November 7, 2018

Plan

I. Large dimensions

II. Applications to global optimization

III. Other repercussions

IV. Conclusions

Chapter I. Large dimensions

Chapter I. Large dimensions

where we learn that our intuition often deceives us

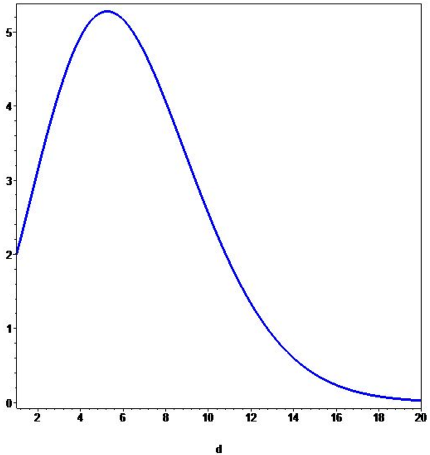
Dimension

$$\mathbb{R}^d$$

Small dimension: $d = 1, 2, 3$
Medium dimension: $d = 10, 20$ (MANY)
Large dimension: $d = 100$ (REALLY MANY)

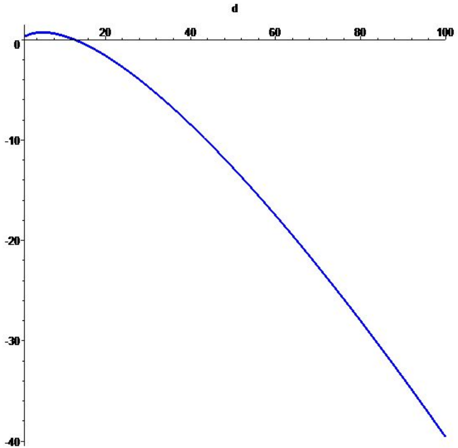
Volume of the d-dimensional unit ball $B(0,1) = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$

$$V_d = \mathrm{vol}(B(0,1)) = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$$



Volume of the d-dimensional unit ball

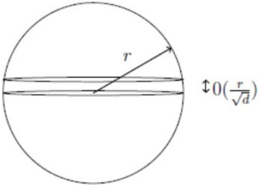
$\log_{10} V_d$ as a function of d:



F.e., $V_{100} \simeq 2.368 \cdot 10^{-40}$

d-dimensional ball

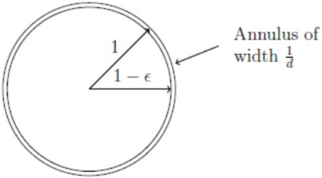
Almost all the volume is near the equator:



Th. *For any $c > 0$, the fraction of the volume of the unit ball above the plane $x_1 = c/\sqrt{d-1}$ is less than $\frac{2}{c}\exp\{-c^2/2\}$.*

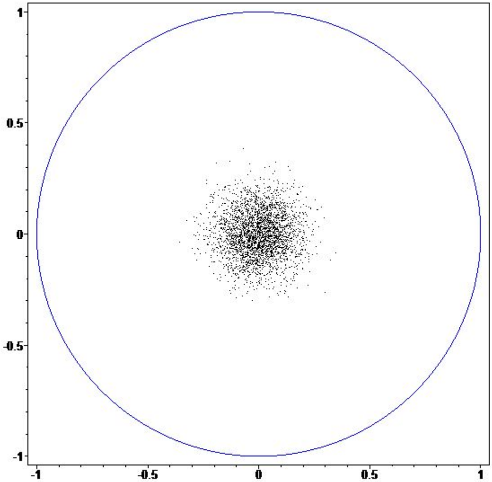
d-dimensional ball

Almost all the volume is also there (in $B(0,1) \setminus B(0,1-\epsilon)$ with $\epsilon = c/d$):

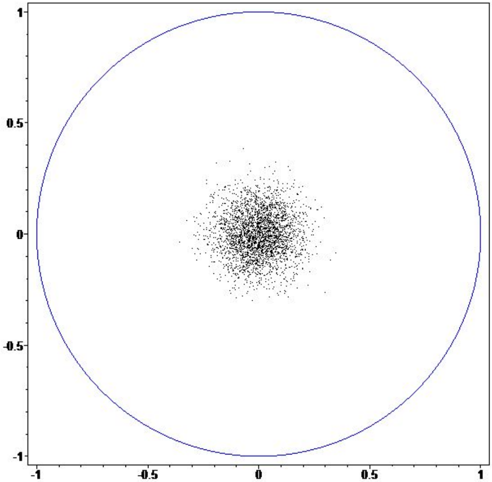


Indeed, $\mathrm{vol}(B(0,1-\epsilon))/\mathrm{vol}(B(0,1)) = (1-\epsilon)^d \simeq 0$ for $\epsilon = c/d$, large d and c fixed but large enough.
Radius of a uniform random point has density $p_d(r) = dr^{d-1}, 0 \leq r \leq 1$.

Random points in a 100-d ball; projection to 2 dimensions



Random points in a 100-d ball; projection to 2 dimensions

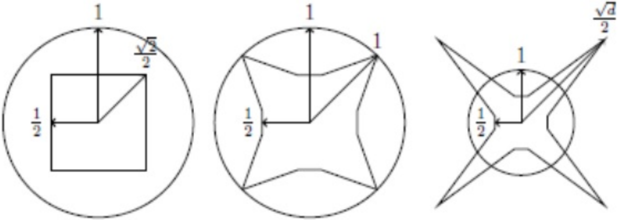


$$B(0,1) = \{x \in \mathbb{R}^d : x_1^2 + x_2^2 + \ldots + x_d^2 \leq 1\}$$

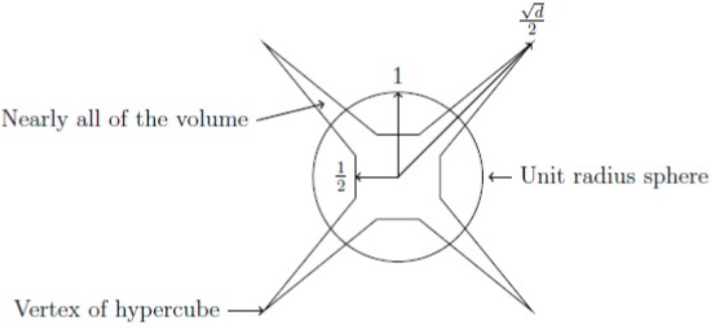
d-dimensional cube and ball

Unit cube: $\{x = (x_1, \ldots, x_d) \in \mathbb{R}^d : |x_i| \leq 1/2\}$
Unit ball: $B(0,1) = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$
Length of the cube's half-diagonal:

$$\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \ldots + \left(\frac{1}{2}\right)^2} = \frac{\sqrt{d}}{2}$$

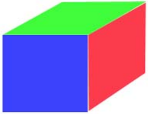


d-dimensional cube

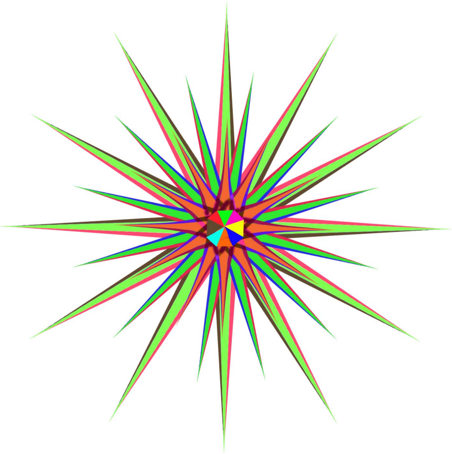


Shape of the d-dimensional cube

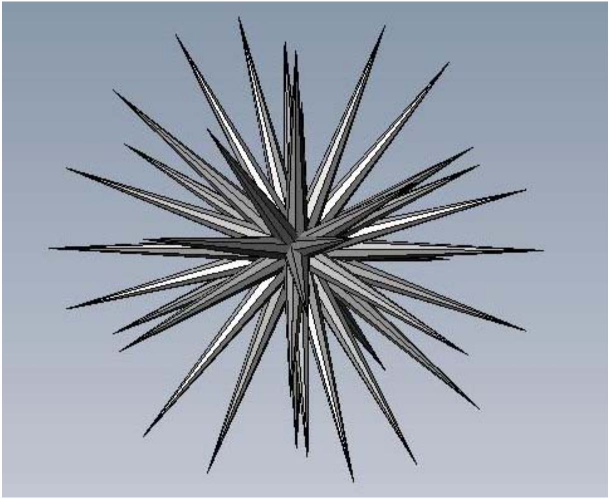
$[-\frac{1}{2}, \frac{1}{2}]^2$

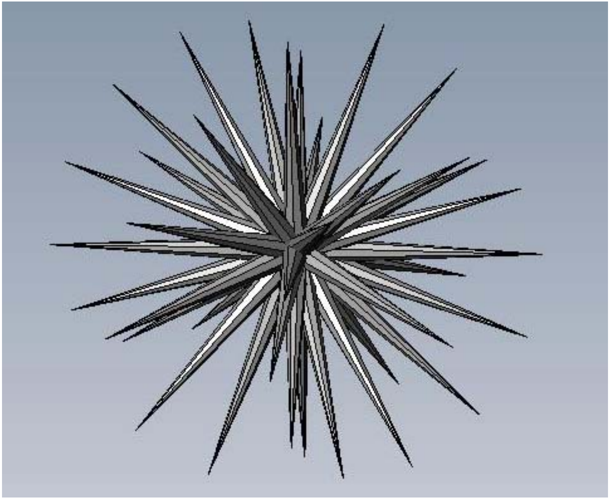


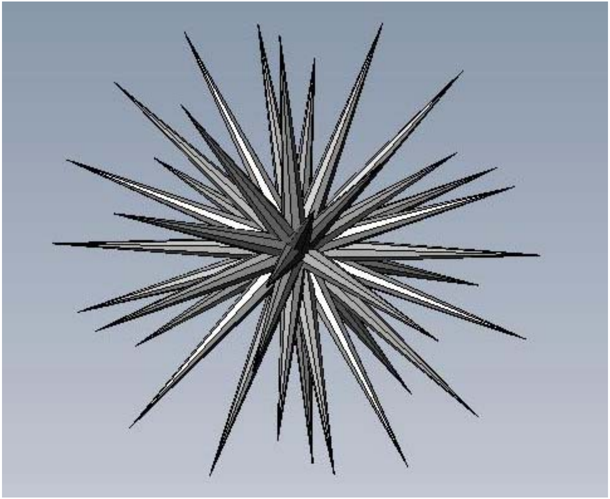
$[-\frac{1}{2}, \frac{1}{2}]^3$

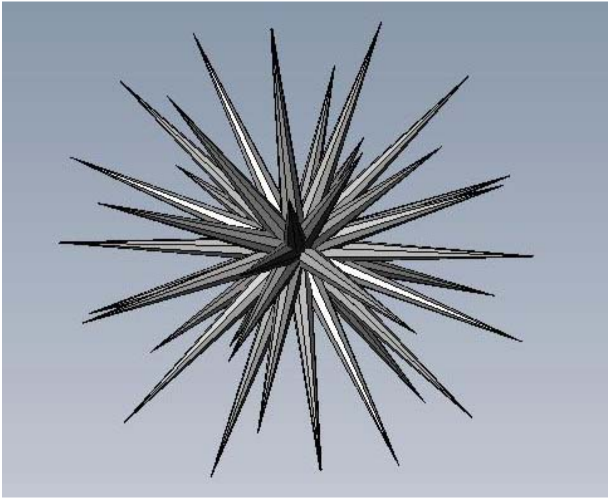


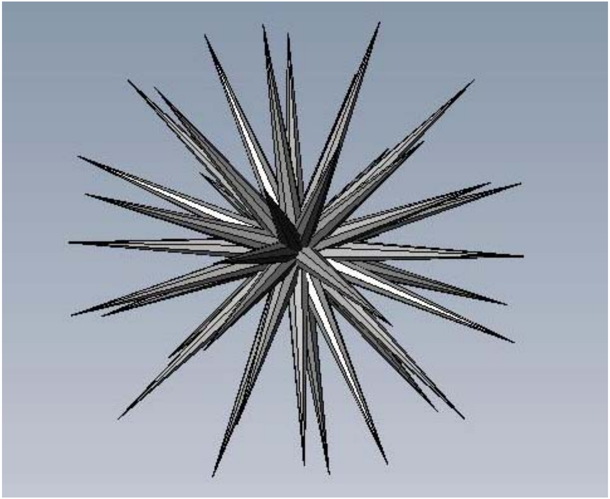
$[-\frac{1}{2}, \frac{1}{2}]^8$





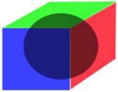


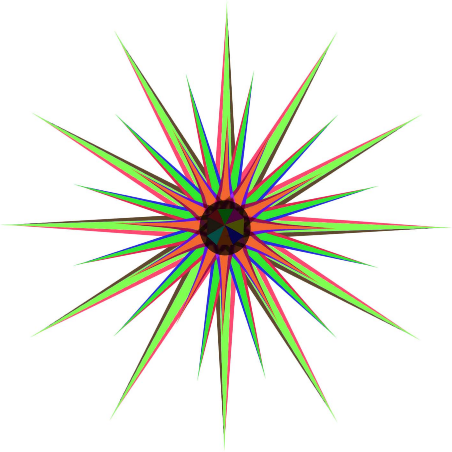




Volume of the largest ball inscribed into the unit cube

Volume of the cube $=1$, $v_d = \frac{\pi^{d/2}}{2^d \Gamma(1+d/2)}$ (volume of the ball of radius $1/2$)

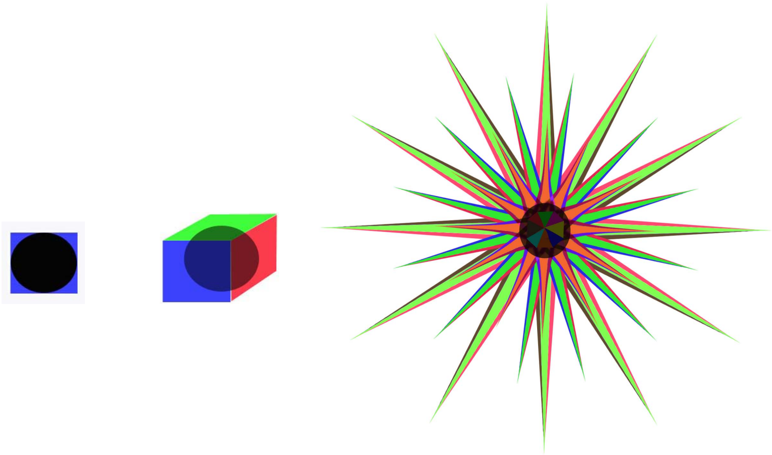




$v_2 = \frac{\pi}{4} \simeq 0.78, \quad v_3 = \frac{\pi}{6} \simeq 0.52$

Volume of the largest ball inscribed into the unit cube

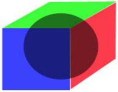
Volume of the cube $=1$, $v_d = \frac{\pi^{d/2}}{2^d \Gamma(1+d/2)}$ (volume of the ball of radius $1/2$)

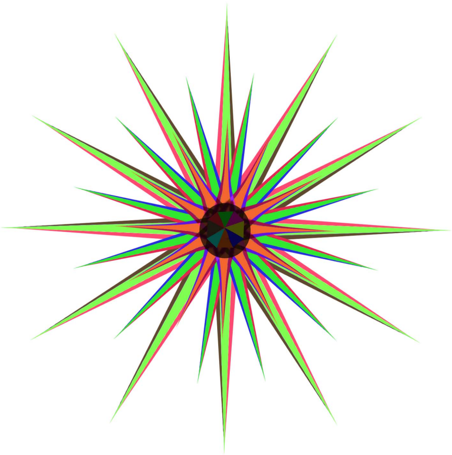


$v_2 = \frac{\pi}{4} \simeq 0.78, \quad v_3 = \frac{\pi}{6} \simeq 0.52, \quad v_{10} \simeq 0.0025$

Volume of the largest ball inscribed into the unit cube

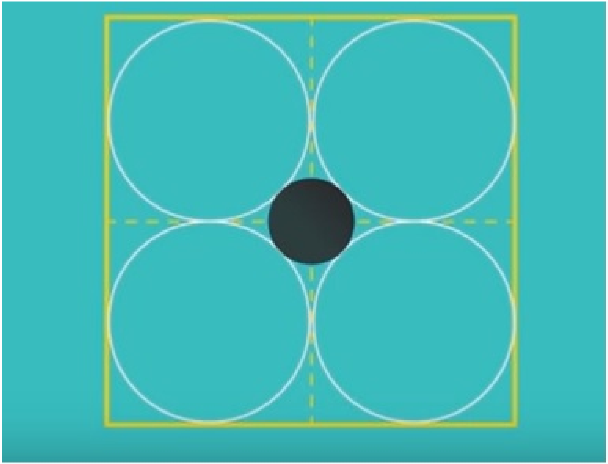
Volume of the cube $=1$, $v_d = \frac{\pi^{d/2}}{2^d \Gamma(1+d/2)}$ (volume of the ball of radius $1/2$)



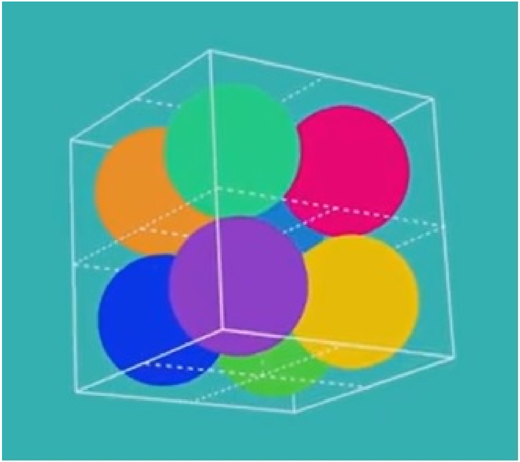


$v_2 = \frac{\pi}{4} \simeq 0.78, \quad v_3 = \frac{\pi}{6} \simeq 0.52, \quad v_{10} \simeq 0.0025,$
$v_{20} \simeq 0.25 \cdot 10^{-7}, \quad v_{100} \simeq 10^{-70}$

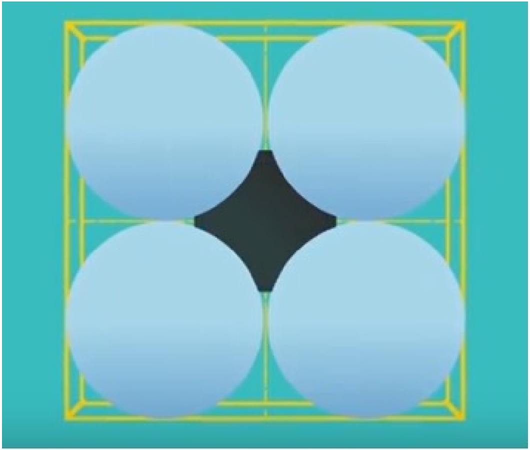
small ball in-between large ones, $d = 2$



small ball in-between large ones, $d = 3$

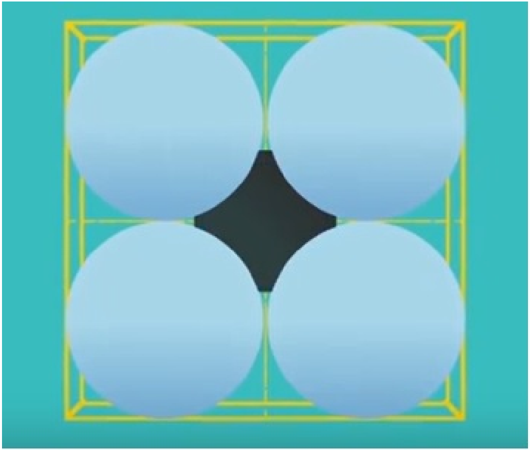


'small' ball in-between 'large' ones, $d \geq 3$



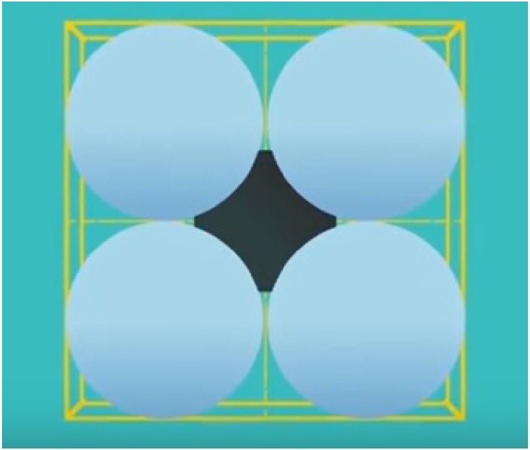
'small' ball in-between 'large' ones, $d \geq 3$

Cube $[-1,1]^d$; centers of 'large' balls of radius $\frac{1}{2}$ are $(\pm\frac{1}{2}, \ldots, \pm\frac{1}{2})$.



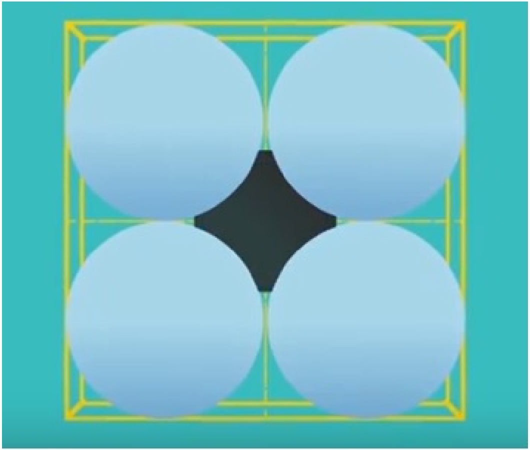
'small' ball in-between 'large' ones, $d \geq 3$

Cube $[-1, 1]^d$; centers of 'large' balls of radius $\frac{1}{2}$ are $(\pm\frac{1}{2}, \ldots, \pm\frac{1}{2})$.
Therefore the radius of the 'small' ball is $r_d = \frac{1}{2}(\sqrt{d} - 1)$.



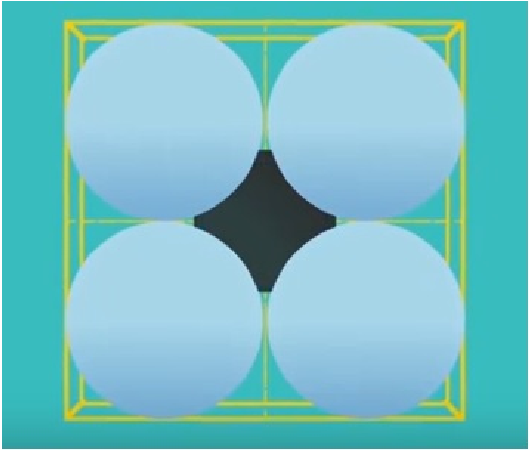
‘small’ ball in-between ‘large’ ones, $d \geq 3$

Cube $[-1,1]^d$; centers of ‘large’ balls of radius $\frac{1}{2}$ are $(\pm\frac{1}{2},\ldots,\pm\frac{1}{2})$.
Therefore the radius of the ‘small’ ball is $r_d = \frac{1}{2}(\sqrt{d}-1)$.
F.e., $r_1 = 0$, $r_2 \simeq 0.207$, $r_3 \simeq 0.366$, $r_4 = \frac{1}{2}$, $r_9 = 1$, $r_{100} = 4.5$

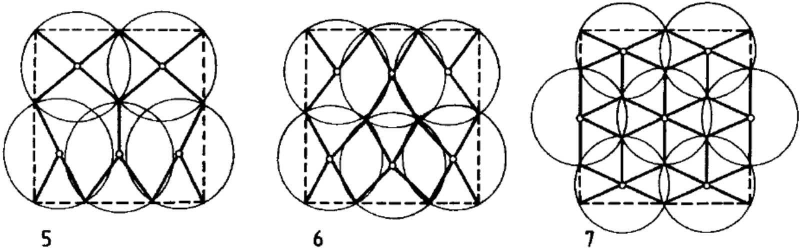


‘small’ ball in-between ‘large’ ones, $d \geq 3$

Cube $[-1, 1]^d$; centers of ‘large’ balls of radius $\frac{1}{2}$ are $(\pm\frac{1}{2}, \ldots, \pm\frac{1}{2})$.
Therefore the radius of the ‘small’ ball is $r_d = \frac{1}{2}(\sqrt{d} - 1)$.
For $d > 1205$, the volume of the ‘small’ ball is larger than 2^d!



Covering of the space (Conway & Sloan)



Θ_d (thickness) $=$ average number of balls that contain a random point.
Some values of this thickness are:
$\Theta_2 \simeq 1.2092$, $\Theta_3 \simeq 1.4635$, $\Theta_{10} \simeq 5.2517$, $\Theta_{20} \simeq 31.14$.

Packing (Conway & Sloan)

Δ_d (density) $=$ proportion of the space occupied by the balls.
Some values of this density are:
$\Delta_2 \simeq 0.906$, $\Delta_3 \simeq 0.74$, $\Delta_{10} \simeq 0.099$, $\Delta_{20} \simeq 0.0032$

Covering and packing, $d = 100$

Θ_d (thickness of covering) = average number of balls that contain a random point.

Δ_d (packing density) = proportion of the space occupied by the balls.

$\Theta_2 \simeq 1.2092$, $\Theta_3 \simeq 1.4635$, $\Theta_{10} \simeq 5.2517$, $\Theta_{20} \simeq 31.14$, $\Theta_{100} \simeq ?$

$\Delta_2 \simeq 0.906$, $\Delta_3 \simeq 0.74$, $\Delta_{10} \simeq 0.099$, $\Delta_{20} \simeq 0.0032$, $\Delta_{100} \simeq ?$

Packing and covering, $d = 100$

Θ_d (thickness of covering) $=$ average number of balls that contain a random point.
Δ_d (packing density) $=$ proportion of the space occupied by the balls.
$\Theta_2 \simeq 1.2092$, $\Theta_3 \simeq 1.4635$, $\Theta_{10} \simeq 5.2517$, $\Theta_{20} \simeq 31.14$
$\Delta_2 \simeq 0.906$, $\Delta_3 \simeq 0.74$, $\Delta_{10} \simeq 0.099$, $\Delta_{20} \simeq 0.0032$, $\Delta_{100} \simeq ?$

$\Theta_{100} \simeq 4.28 \cdot 10^7$ (an average point is covered more than 40 million times!)

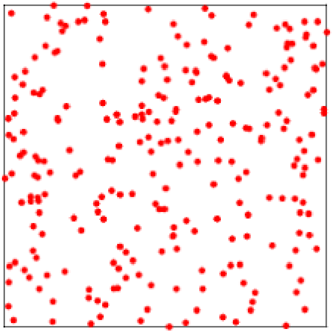
Packing and covering, $d = 100$

Θ_d (thickness of covering) $=$ average number of balls that contain a random point.
Δ_d (packing density) $=$ proportion of the space occupied by the balls.
$\Theta_2 \simeq 1.2092$, $\Theta_3 \simeq 1.4635$, $\Theta_{10} \simeq 5.2517$, $\Theta_{20} \simeq 31.14$
$\Delta_2 \simeq 0.906$, $\Delta_3 \simeq 0.74$, $\Delta_{10} \simeq 0.099$, $\Delta_{20} \simeq 0.0032$

$\Theta_{100} \simeq 4.28 \cdot 10^7$ (an average point is covered more than 40 million times!)

$\Delta_{100} < 10^{-26}$ (less than 0.000000000000000000000001% of the space is occupied by the balls!)

Uniform random points on a square



Uniform points in a cube are at almost the same distance from each other

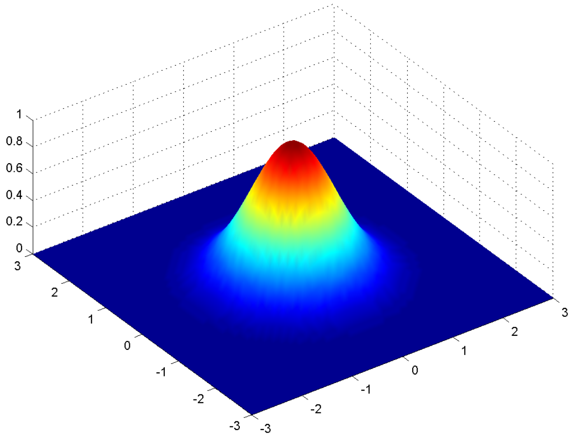
The distribution of the distances

$$\|x - y\| = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

is concentrated around its expected value which is approximately $\sqrt{d/6}$.

Similar results hold for the unit ball and for the distributions different from the uniform.

Gaussian distribution (density function)



Gaussian random vectors

If x is Gaussian $N(0, I_d)$ then the distance from the origin

$$r = \sqrt{\sum_{i=1}^{d} x_i^2}$$

is very close to $\sqrt{d}$.
More precisely, for any $0 < \beta < \sqrt{d}$,

$$Pr\{\sqrt{d} - \beta \leq r \leq \sqrt{d} + \beta\} \geq 1 - 3\beta^2/64$$

Two i.i.d. Gaussian vectors are almost orthogonal to each other. Similar for uniform r.v. in a ball and in a cube.

Random projections

Johnson-Lindenstrauss Lemma. For any $0 < \varepsilon < 1$ and any integer n, let $k \geq c\varepsilon^2 \log n$ for some $c > 0$. For any set of n points in $\mathbb{R}^d$, the random projection $f : \mathbb{R}^d \to \mathbb{R}^k$ has the property that for all pairs of points v_i and v_j, with probability at least $1-\frac{3}{2n}$,

$$(1-\varepsilon)\|v_i - v_j\| \leq \|f(v_i) - f(v_j)\| \leq (1-\varepsilon)\|v_i - v_j\|$$

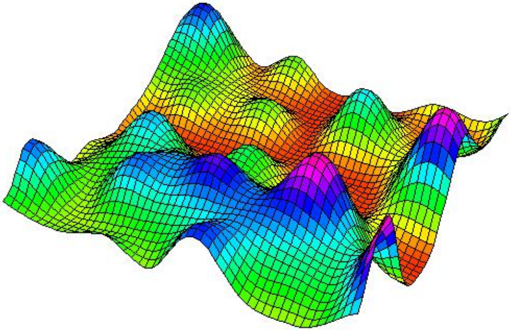
Chapter II. Applications to global optimization

Chapter II. Applications to global optimization

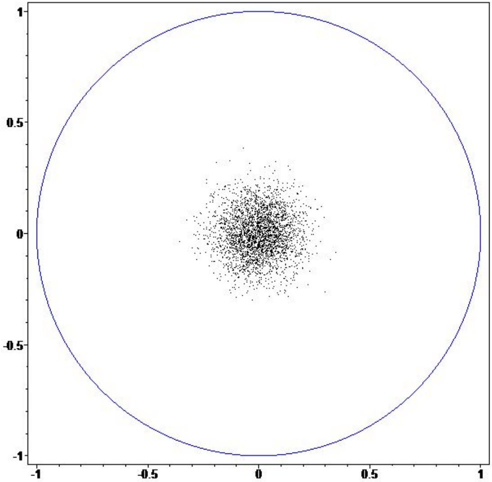
where we do not see many reasons for optimism

Global optimization

$$f(x) \to \min_{x \in A}; \quad x_* = \arg\min_{x \in A} f(x)$$



Random points in a ball; projection to 2 dimensions



How far are the points from the boundary? $d \in [5, 200]$

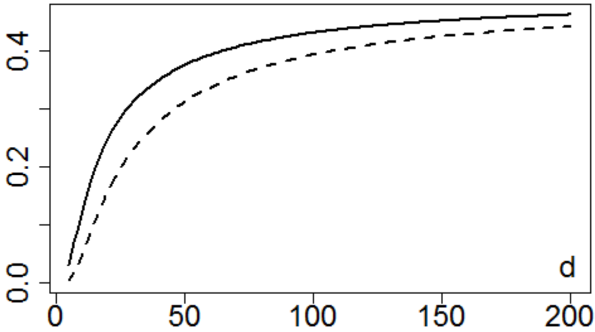


Figure: The difference $y_{1,n} - f_*$ for $n = 10^6$ (solid) and $n = 10^{10}$ (dashed), where $y_{1,n}$ is the record of evaluations of the function $f(x) = e_1^T x$ at points $x_1, \ldots, x_n$ with uniform distribution in the unit ball in the dimension d as d varies in $[5, 200]$.

Are quasi-random points any better?

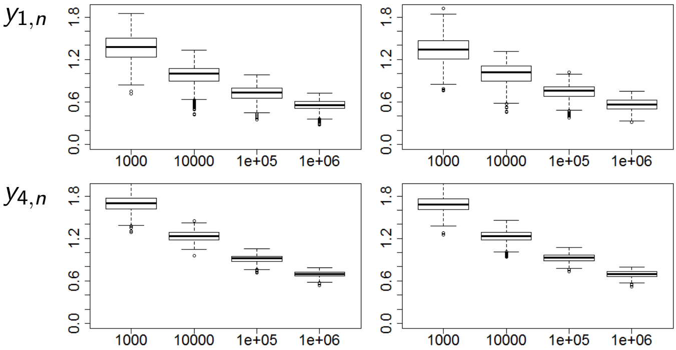


Figure: Boxplots of $y_{1,n}$ and $y_{4,n}$ for 500 runs with points generated from the Sobol low-dispersion sequence (left) and the uniform distribution (right), $d = 20$.

Rate of convergence of the simple random search

The number of points n_γ required to hit a ball or radius ε centered at the minimizer, with probability $\geq 1-\gamma$, for different dimensions d:

d	$\gamma = 0.1$			$\gamma = 0.05$		
	$\varepsilon = 0.5$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.5$	$\varepsilon = 0.2$	$\varepsilon = 0.1$
1	0	5	11	0	6	14
2	2	18	73	2	23	94
3	4	68	549	5	88	714
5	13	1366	43743	17	1788	56911
10	924	$8.8 \cdot 10^6$	$9.0 \cdot 10^9$	1202	$1.1 \cdot 10^7$	$1.2 \cdot 10^{10}$
20	$9.4 \cdot 10^7$	$8.5 \cdot 10^{15}$	$8.9 \cdot 10^{21}$	$1.2 \cdot 10^8$	$1.1 \cdot 10^{16}$	$1.2 \cdot 10^{22}$
50	$1.5 \cdot 10^{28}$	$1.2 \cdot 10^{48}$	$1.3 \cdot 10^{63}$	$1.9 \cdot 10^{28}$	$1.5 \cdot 10^{48}$	$1.7 \cdot 10^{63}$
100	$1.2 \cdot 10^{70}$	$7.7 \cdot 10^{109}$	$9.7 \cdot 10^{139}$	$1.6 \cdot 10^{70}$	$1.0 \cdot 10^{110}$	$1.3 \cdot 10^{140}$

n_γ is roughly ε^{-d}/V_d (multiplied by $-\ln \gamma$); recall $V_{100} \simeq 10^{-40}$.

Convergence: Borel-Cantelli lemma

Global random search algorithm converges if

$$\sum_{j=1}^{\infty} \inf P_j(B(x_*, \varepsilon)) = \infty \qquad (1)$$

for any $\varepsilon > 0$, where $B(x_*, \varepsilon) = \{x \in A : ||x - x_*|| \leq \varepsilon\}$; the infimum in (1) is taken over all possible previous points and the results of the objective function evaluations at them.
Standard choice of probability distributions to guarantee convergence:

$$P_{j+1} = \alpha_{j+1} P_U + (1 - \alpha_{j+1}) Q_j \,, \quad \sum_j \alpha_j = \infty \,.$$

Example: $P_{j+1} = \alpha_{j+1} P_U + (1-\alpha_{j+1}) Q_j\,, \quad \alpha_j = 1/j\,.$

Using the approximation $\sum_{j=1}^{n} \alpha_j \simeq \ln n$, we obtain
$n_\gamma \simeq \exp\{-\ln\gamma / P_U(B)\}$.
If $A = [0,1]^d$ this gives $n_\gamma \simeq \exp\{-\ln\gamma / P_U(B)\}$.
Assuming further $B = B(x_*, \varepsilon)$ we obtain $n_\gamma \simeq \exp\{\mathrm{const} \cdot \varepsilon^{-d}\}$, where $\mathrm{const} = (-\ln\gamma)/V_d$ (if x_* lies closer to the boundary of A than ε then $n(\gamma)$ is even larger).
For example, for $\gamma = 0.1$, $d = 10$ and $\varepsilon = 0.1$, n_γ is a number larger than $10^{1000000000}$
Even for $d = 3$, $\gamma = 0.1$ and $\varepsilon = 0.1$, the value of n_γ is huge: $n_\gamma \simeq 10^{238}$.

Simulated Annealing (SA) and Quantum Annealing (QA)

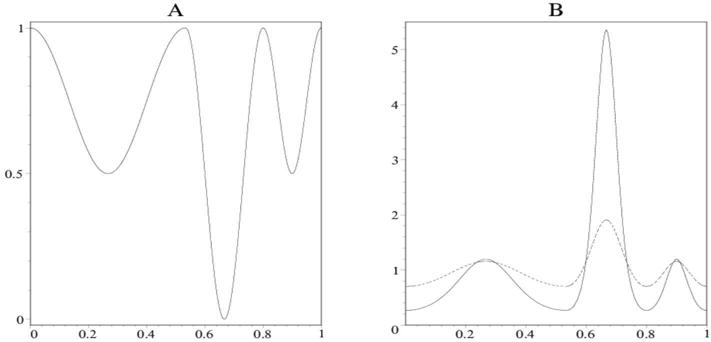
Simulated Annealing (SA) and Quantum Annealing (QA)

can they help us in getting faster convergence?

SA and Gibbs densities

SA accepts the move $x_k \to x_{k+1}$ w.p. 1 if $f(x_{k+1}) \leq f(x_k)$ and $\exp(-(f(x_{k+1}) - f(x_k))/(Kt_k))$ if $f(x_{k+1}) > f(x_k)$.

$$\pi_\beta(x) = \exp\{-\beta f(x)\} \Bigg/ \int_A \exp\{-\beta f(z)\} dz \quad \beta = 1/(Kt)\,.$$



(A) Graph of the objective function f; (B) Gibbs densities with $\beta = 1$ (dotted line) and $\beta = 3$ (solid line)

SA, convergence

Geman S., Geman D. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." IEEE Transactions on pattern analysis and machine intelligence 6 (1984): 721-741. Cited by more than 22,000.

Introduction, p.3:

Roughly speaking, it says that if the temperature $T(k)$ employed in executing the kth site replacement (i.e., the kth image in the iteration scheme) satisfies the bound

$$T(k) \geqslant \frac{c}{\log(1+k)}$$

for every k, where c is a constant independent of k, then with probability converging to one (as $k \to \infty$), the configurations generated by the algorithm will be those of minimal energy.

Geman & Geman (the theorem)

Let

$$\Omega_0 = \{\omega \in \Omega : U(\omega) = \min_{\eta} U(\eta)\}, \tag{12.4}$$

and let π_0 be the uniform distribution on Ω_0. Finally, define

$$U^* = \max_{\omega} U(\omega),$$

$$U_* = \min_{\omega} U(\omega),$$

$$\Delta = U^* - U_*. \tag{12.5}$$

Theorem B (Annealing): Assume that there exists an integer $\tau \geqslant N$ such that for every $t = 0, 1, 2, \cdots$ we have

$$S \subseteq \{n_{t+1}, n_{t+2}, \cdots, n_{t+\tau}\}.$$

Let $T(t)$ be any decreasing sequence of temperatures for which

a) $T(t) \to 0$ as $t \to \infty$;
b) $T(t) \geqslant N\Delta / \log t$
for all $t \geqslant t_0$ for some integer $t_0 \geqslant 2$.

Then for any starting configuration $\eta \in \Omega$ and for every $\omega \in \Omega$,

$$\lim_{t \to \infty} P(X(t) = \omega \mid X(0) = \eta) = \pi_0(\omega). \tag{12.6}$$

Geman & Geman (comment after the theorem)

$N/logk = t \ \Rightarrow \ \log k = \exp(N/t)$

$N = 20000, \ t = \frac{1}{2} \ \Rightarrow \ k = \exp(40000) \simeq 6 \cdot 10^{17371}$

Travelling salesman with 10 cities:
$N = 10! = 3628800 \ \Rightarrow \ k = \exp(3628800) \simeq 6.5 \cdot 10^{1575967}$
If we take $\log_2$ from this number we get $\simeq 5 \cdot 10^6$.
For 20 cities we get $20! = 2432902008176640000$ and $\simeq 7 \cdot 10^{18}$.

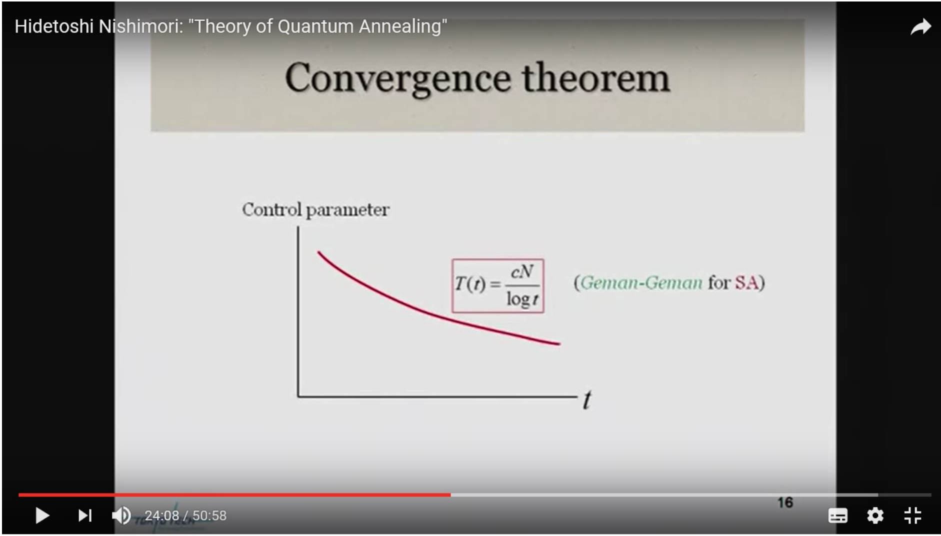
The major practical weakness is b); we cannot truly follow the "schedule" $N\Delta/\log t$. For example, with $N = 20{,}000$ and $\Delta = 1$, it would take $e^{40,000}$ site visits to reach $T = 0.5$.

SA, convergence

The formula

$$T(k) = \frac{c}{\log k} \;\text{ with }\; c = N\Delta$$

for the temperature decrease in SA is one of the most famous formulas in optimization; see e.g. 24-th minute in the celebrated Google talk by Hidetoshi Nishimori *Theory of Quantum Annealing*: https://www.youtube.com/watch?v=OQ91L96YWCk

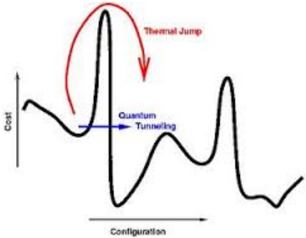

Hidetoshi Nishimori: "Theory of Quantum Annealing"
Convergence theorem
Control parameter
$T(t) = \dfrac{cN}{\log t}$
(Geman-Geman for SA)
t
16
24:08 / 50:58

My comment on SA in 1985/1991

AZ(1985, 1991):

the time required to approach the stationary Gibbs-distribution increases exponentially with 1/T and may reach astronomical values for small T (as confirmed also by numerical results). This can be explained by the fact that for small T a homogeneous simulated annealing method tends to be like the local random search algorithm that rejects unprofitable steps, and so its global search features are poor.

QA versus SA



QA in words

QA uses a quantum field instead of a thermal gradient. In order to explore the landscape of the objective function, SA and its variants use "thermal" fluctuations associated to temperature gradients, while QA uses quantum fluctuations.
When the QA is applied to a minimization problem, a current state is replaced by a “neighbor state“ chosen randomly (or chosen by a more sophisticated method).
Main area where QA may be efficient: combinatorial optimization, like the classical “Traveling Salesman Problem”.

QA

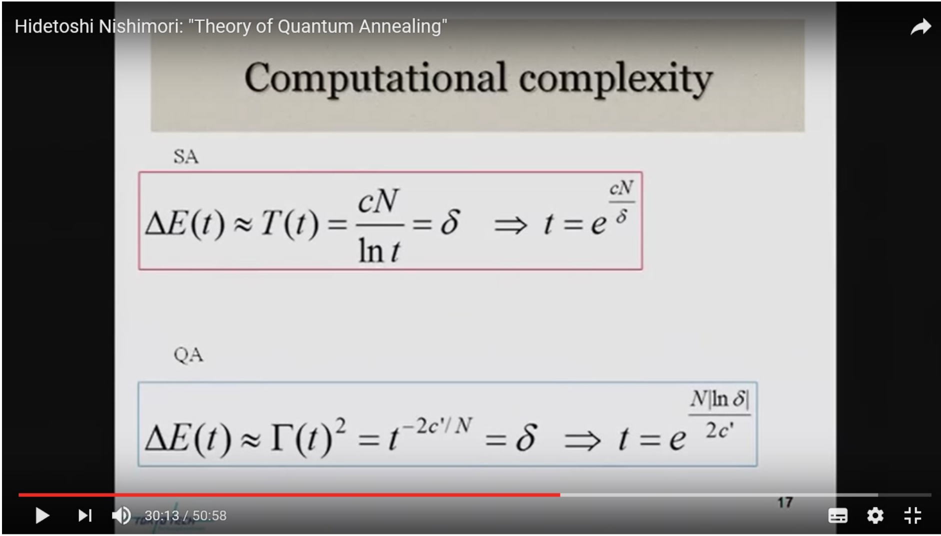
- Main idea: Hamiltonian at time t :

$$H(t) = \left(1 - \frac{t}{T}\right) H_0 + \frac{t}{T} H_q, \;\; 0 \leq t \leq T \,.$$

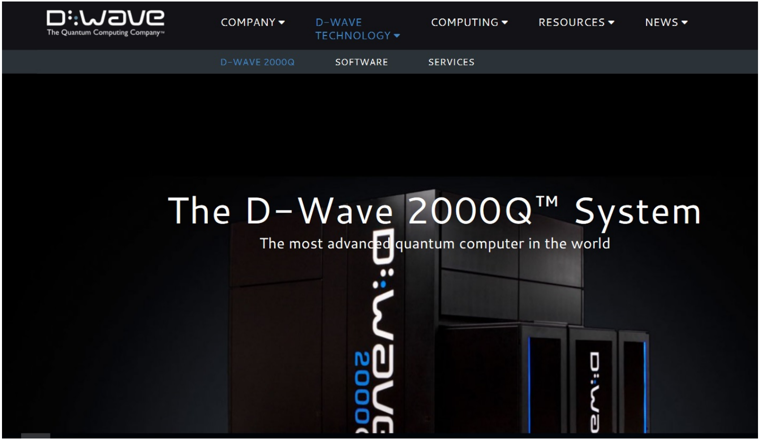
- Suited to: QUBO (Quadratic Unconstrained Binary Optimization):

$$\sum_{i,j=1}^{n} Q_{i,j}\, x_i x_j \to \min_{x \in \{-1,+1\}^n}$$

QA versus SA


Hidetoshi Nishimori: "Theory of Quantum Annealing"
Computational complexity
SA
$\Delta E(t) \approx T(t) = \dfrac{cN}{\ln t} = \delta \quad \Rightarrow t = e^{\frac{cN}{\delta}}$
QA
$\Delta E(t) \approx \Gamma(t)^2 = t^{-2c'/N} = \delta \quad \Rightarrow t = e^{\frac{N|\ln \delta|}{2c'}}$
17
30:13 / 50:58

Quantum computer D-Wave



What have we reached with quantum computers so far?

Factorization into prime factors: $21 = 3 \times 7$
(this was a record in 2012; now it is slightly larger, like $56153 = 233 \times 241$)

QUBO with D-Wave:

$$\sum_{i,j=1}^{n} Q_{i,j}\, x_i x_j \to \min_{x \in \{-1,+1\}^n}$$

Largest n?

Can DNA computers help?

Can the infinity computer help?

Roughly, the grossone-based infinity computer operates with infinitesimals as fast as with ordinary numbers. It's not built yet.

Some references

A.Blum, J. Hopcroft, R. Kannan (2018) Foundations of Data Science

K. Ball (1997) An elementary introduction to modern convex geometry

AZ. Theory of global random search, Kluwer, 1991

AZ, A. Zilinskas. Stochastic global optimization, Springer, 2001

A. Zilinskas, A. Pepelyshev, AZ (2017) Performance of global random search algorithms for large dimensions. JOGO

Thank you

Thank you very much

Thank you very much
for listening

Thank you very much
for listening,
for participating in this meeting

Thank you very much
for listening,
for participating in this meeting,
for your interest in the area of big data