Inverse Theory - a modern method of data analysis

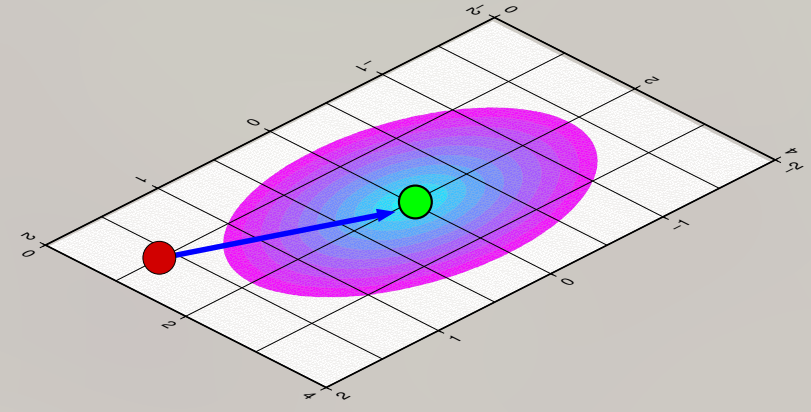
Lecture 5

W. Dębski

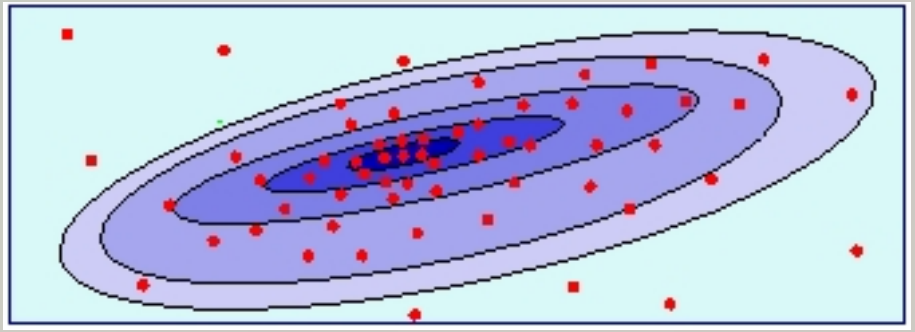
18.05.2021

Optimization approach - generalization

$$\begin{array}{ccc} \mathcal{D} & \Longrightarrow & \mathcal{D} \\ \uparrow & & \downarrow \\ \mathcal{M} & \Longrightarrow & \mathcal{M} \end{array}$$



Search over the $\mathcal{M}$ space for a model which best reproduces $\mathbf{d}^{obs}$



Optimization approach - direct search for best m

Generalization

$$S(\mathbf{m}) = ||\mathbf{d}^{obs} - \mathbf{d}^{th}(\mathbf{m})|| + ||\mathbf{m} - \mathbf{m}^{apr}||$$

solution: search for $\mathbf{m}^{ml}$ minimizing $S(\mathbf{m})$

$$S(\mathbf{m}^{ml}) = \min$$

Optimization approach - main steps

- selection *a priori* model $\mathbf{m}^{apr}$
- selection norms $||\cdot||_{\mathcal{D}}$ and $||\cdot||_{\mathcal{M}}$
- selection optimization algorithm
- run optimization
- post-optimization analysis (residua, resolution, etc.)

Optimization approach - features

- fully nonlinear method
- variety of existing optimization method
- Choice of $S(\mathbf{m})$ - different norms + additional constraints
- problem with error estimation
- is solution unique ?

Norms

l_1	$\|\|\mathbf{d}\|\| = \sum_i \left\| \frac{d^i}{C^i} \right\|$	Absolute value norm
l_2	$\|\|\mathbf{d}\|\| = \sum_{ij} d^i C^{ij} d^j$	Gaussian norm
l_c	$\|\|\mathbf{d}\|\| = \sum_i \log\left(1 + \left(\frac{d^i}{C^i}\right)^2\right)$	Cauchy norm
l_m	$\|\|\mathbf{d}\|\| = \log\left(1 + \sum_i \left(\frac{d^i}{C^i}\right)^2\right)$	Modified Cauchy norm
l_s	$\|\|\mathbf{d}\|\| = \sum_i \log\left[\cosh\left(\frac{x_i}{C_i}\right)\right]$	Hyperbolic secant norm
l_p	$\|\|\mathbf{d}\|\| = \sqrt[p]{\sum_i \left(\frac{\|d^i\|}{C_i}\right)^p}$	generalized Gaussian norm

Optimization approach - features

- fully nonlinear method
- variety of existing optimization method
- Choice of norms - $S(\mathbf{m})$
- problem with error estimation
- is solution unique ?

Solution errors

Optimization approach provides the “best fitting” **single** solution Estimation of accuracy of the find solution is, however, problematic and usually needs special additional afford and develop case-dependent approach. There is NO general method of uncertainty analysis for this approach!

Possible approaches

- linearization around optimum found and calculation covariance matrix
- Monte Carlo simulation
- skipping the problem

Monte Carlo simulation

$$||\mathbf{d}^{obs} - \mathbf{d}^{th}(\mathbf{m}^{est})||_D + ||\mathbf{m}^{est} - \mathbf{m}^{apr}||_M = \min$$

$$\begin{aligned} ||\mathbf{d}^{obs} - \mathbf{d}^{th}(\mathbf{m}^{est}) + \epsilon^{obs} + \epsilon^{th}(\mathbf{m})||_D \quad + \\ ||\mathbf{m}^{est} - \mathbf{m}^{apr} + \epsilon^{apr}||_M \qquad\qquad = \quad \min \end{aligned}$$

$$\Downarrow$$

$$\mathbf{m}^{est}_{ijk} = \mathbf{m}^{est}\left(\epsilon^{obs}_i, \epsilon^{th}_j(\mathbf{m}^{est}), \epsilon^{apr}_k\right)$$

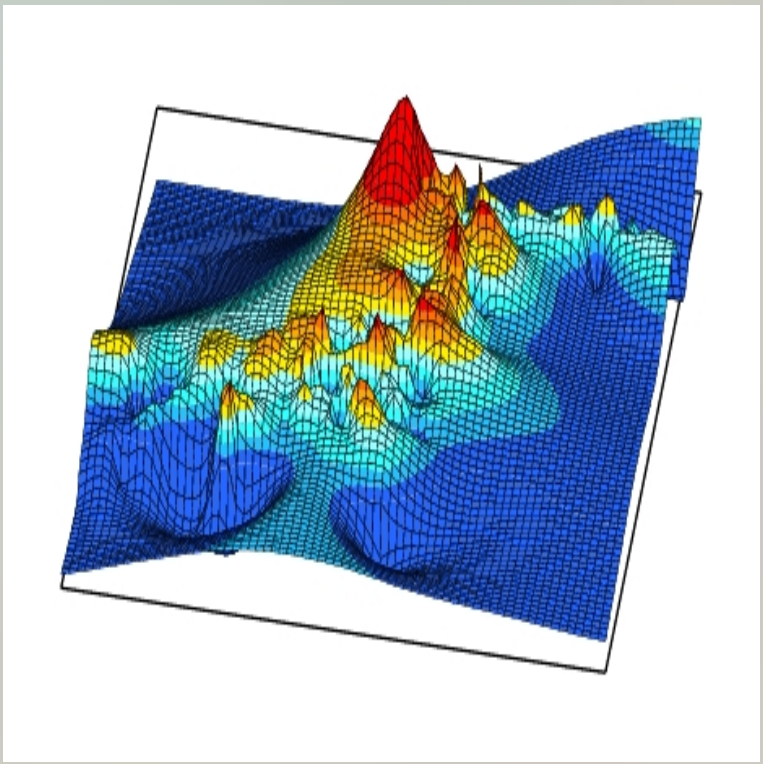
Optimization approach - features

- fully nonlinear method
- variety of existing optimization method
- Choice of norms - $S(\mathbf{m})$
- problem with error estimation
- is solution unique ?

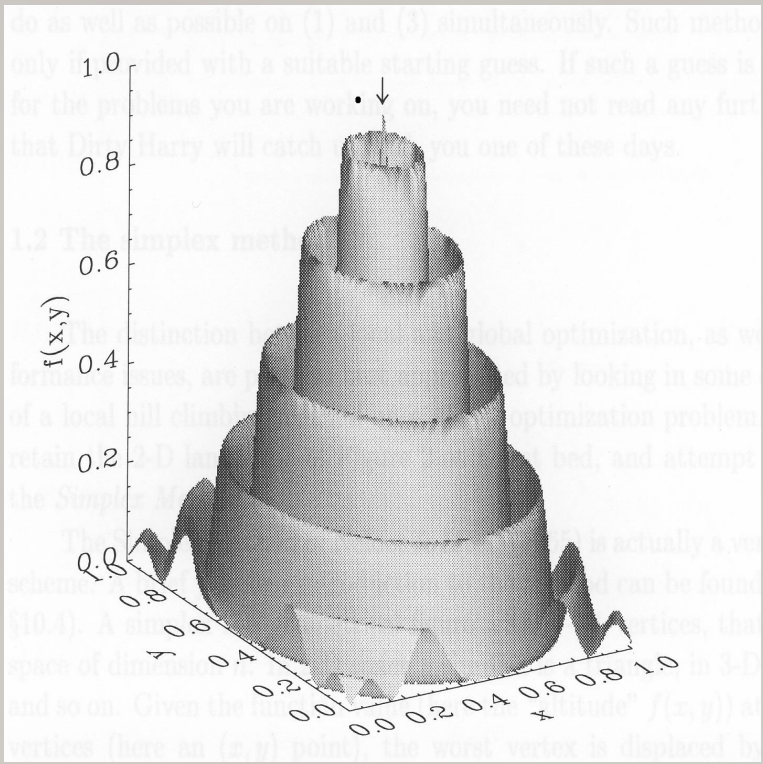
Optimization method - non-uniqness

$$\sigma(\mathbf{m}) \;=\; \exp(-S(\mathbf{m}))$$

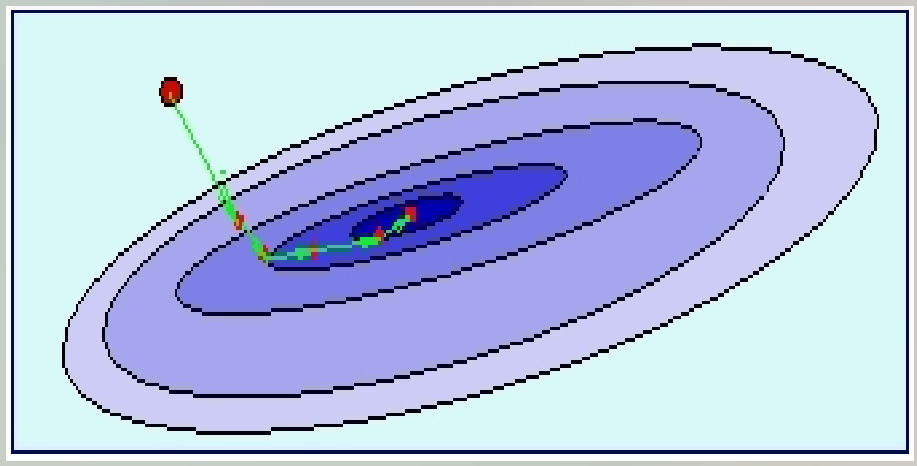
multi-modality

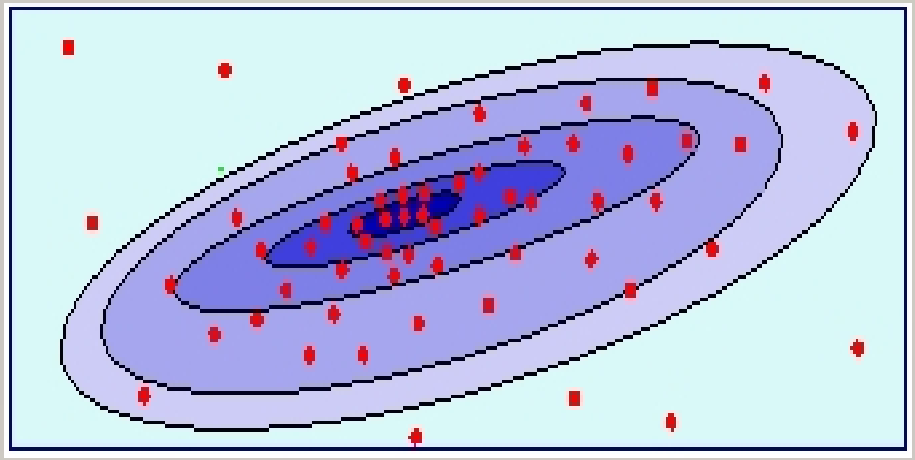


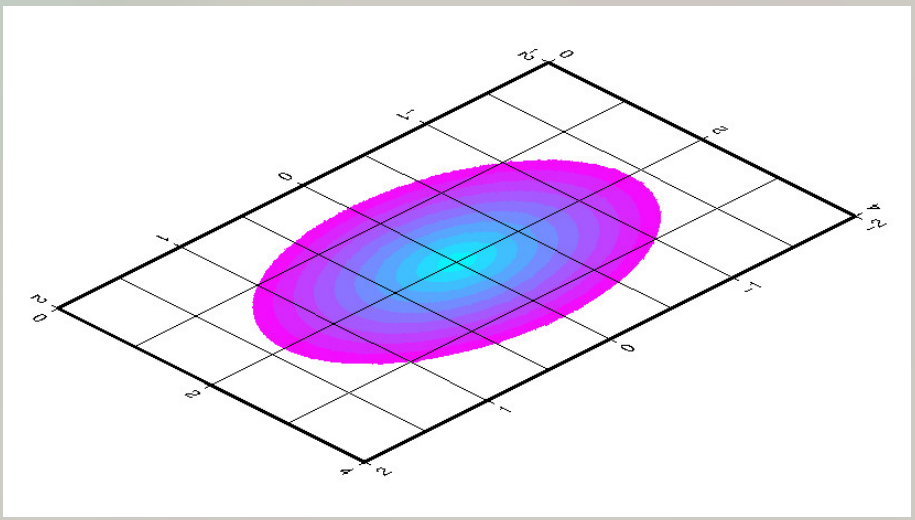
null space



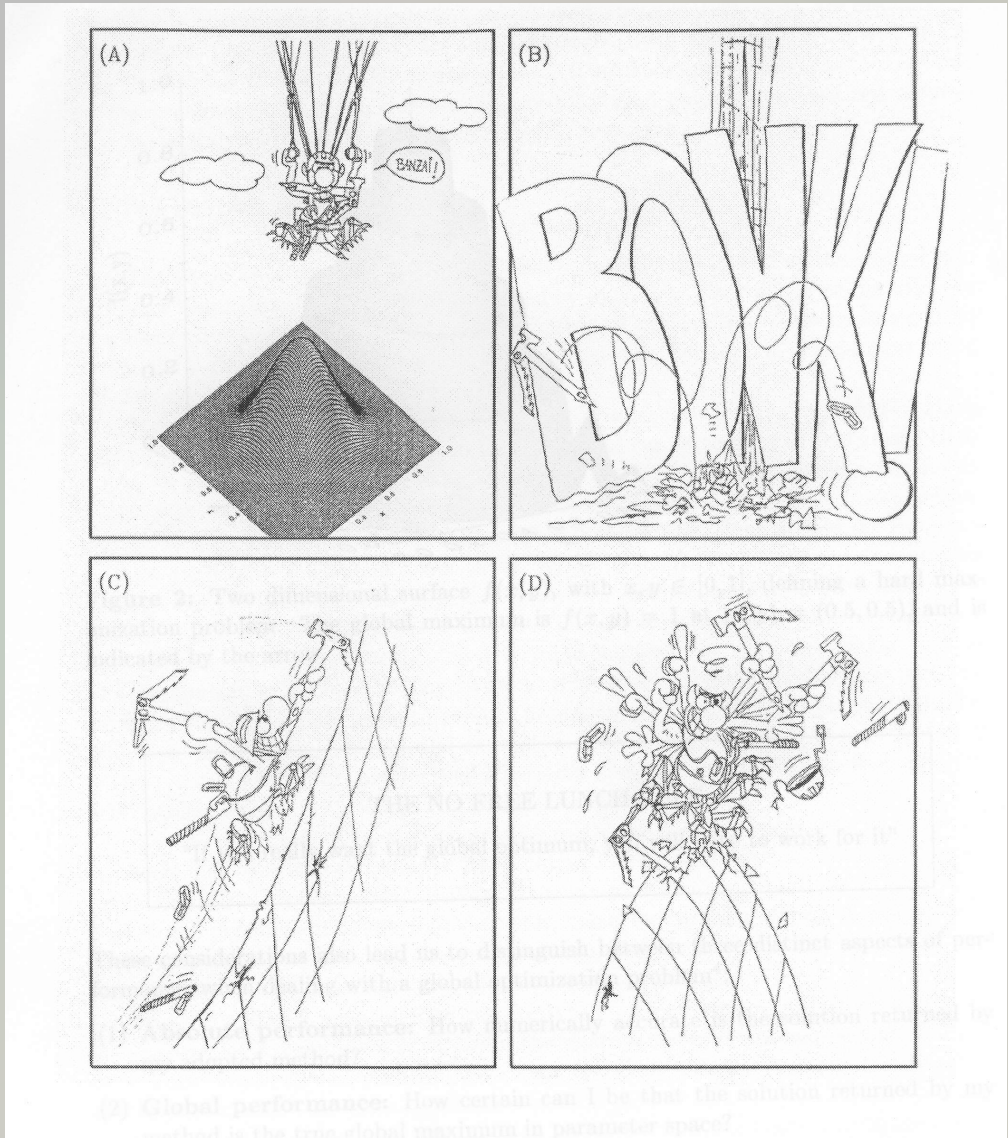
Optimization approach - optimizer selection







Optimization - *gradient methods*

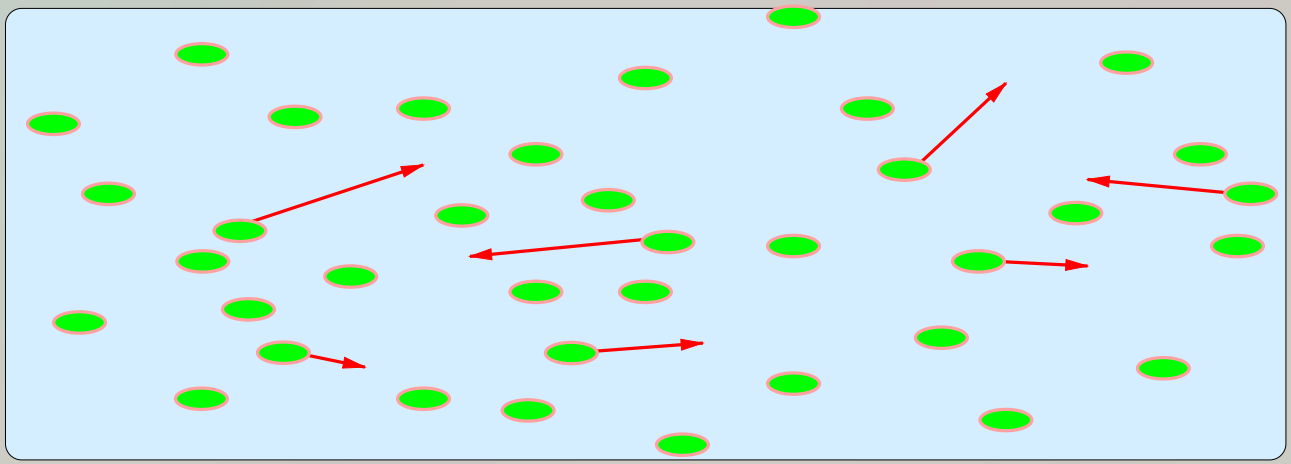


Global optimization algorithm - *Simulated Annealing*

annealed metal, rocks, gases, liquids etc.:

$$\mathbf{1m^3}: \quad N \approx 10^{21}$$

How to describe such systems ??? $(\vec{\mathbf{r}}_\mathbf{i}, \vec{\mathbf{v}}_\mathbf{i})$



Complex systems: statistical approach

- We are mostly interested in *macroscopic* physical parameters: temperature,density, viscosity, density, etc.
- Many *microscopic* states corresponds to the same macroscopic state of the system
- Thermodynamic equilibrium: macroscopic state is **constant** in time but microscopic one is **changing continueously**

Thermodynamic equilibrium

Boltzmann (canonical) distribution

$$p(C) = \frac{1}{Z} exp\left(-\frac{E(C)}{kT}\right)$$

E - energy of given microscopic state

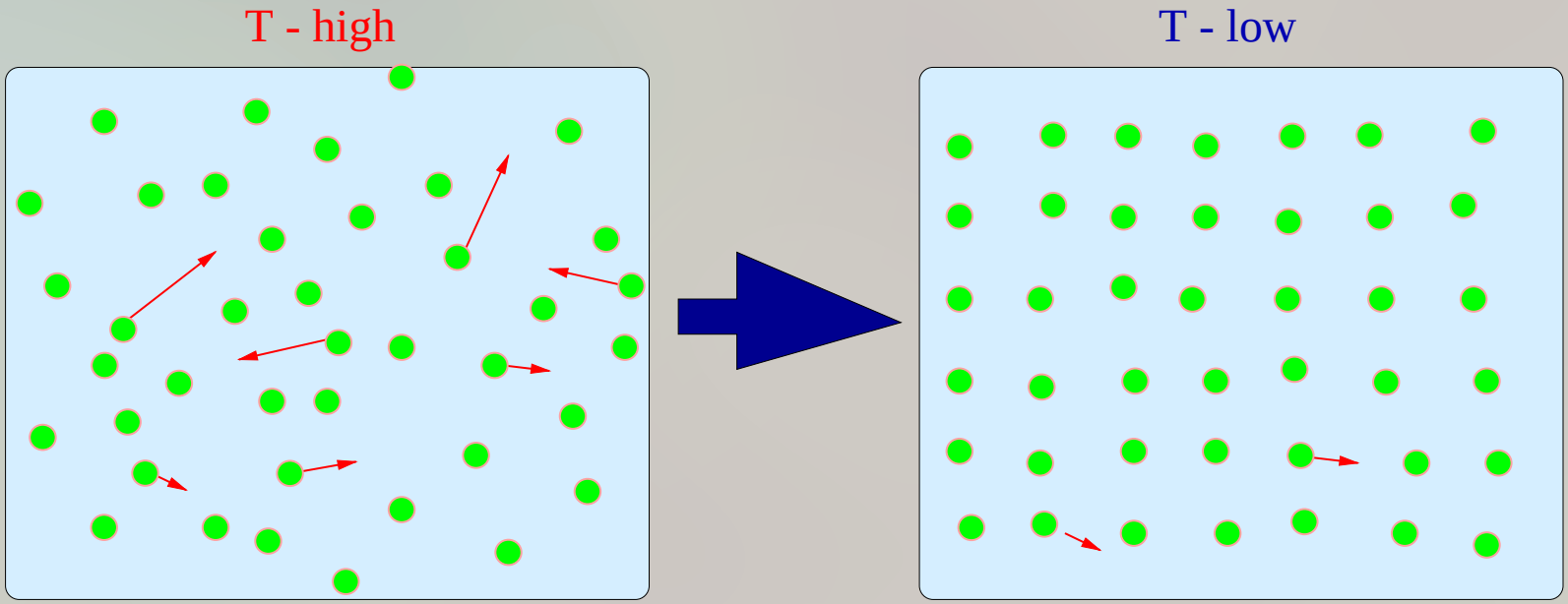
T - temperature

$$Z(T, M, ...) = \sum_C exp\left(-\frac{E(C)}{kT}\right)$$

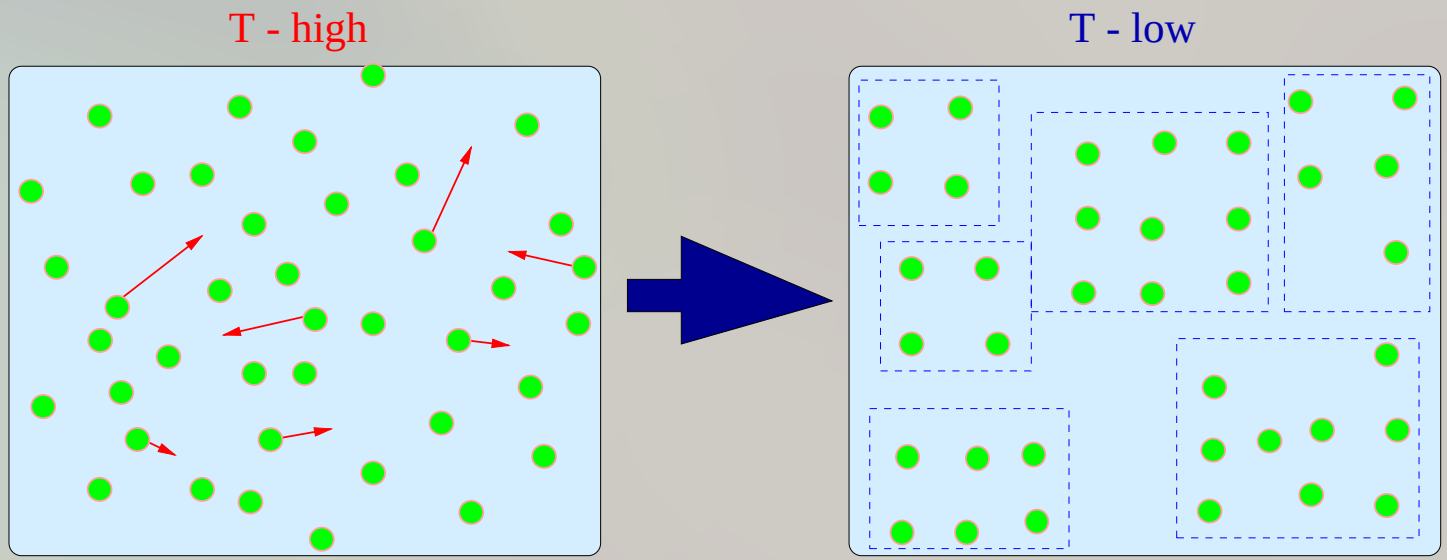
this distribution describes probability of a "realization" of the macroscopic state by a given microscopic state C

Physical application: slow cooling process

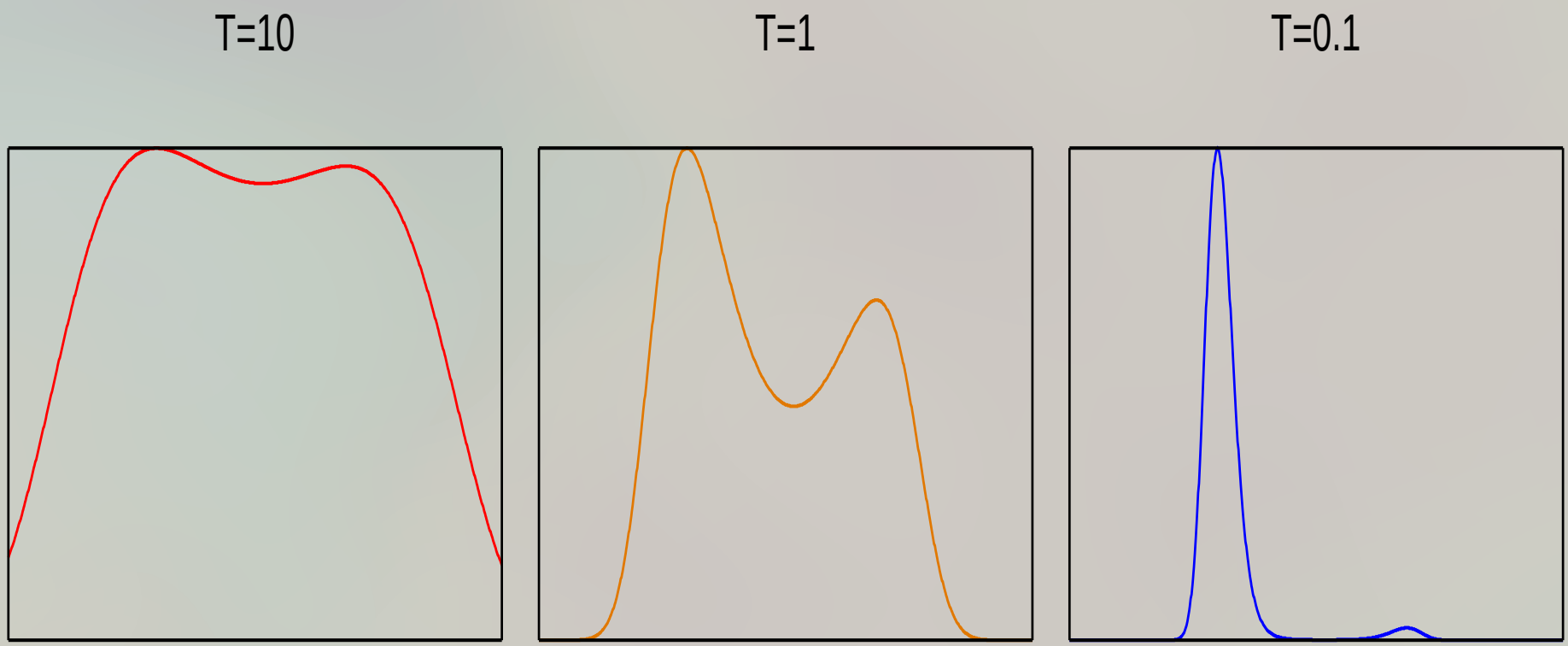
Cooling process - system goes to a state of a minimum internal energy - thermal movements become “frozen”



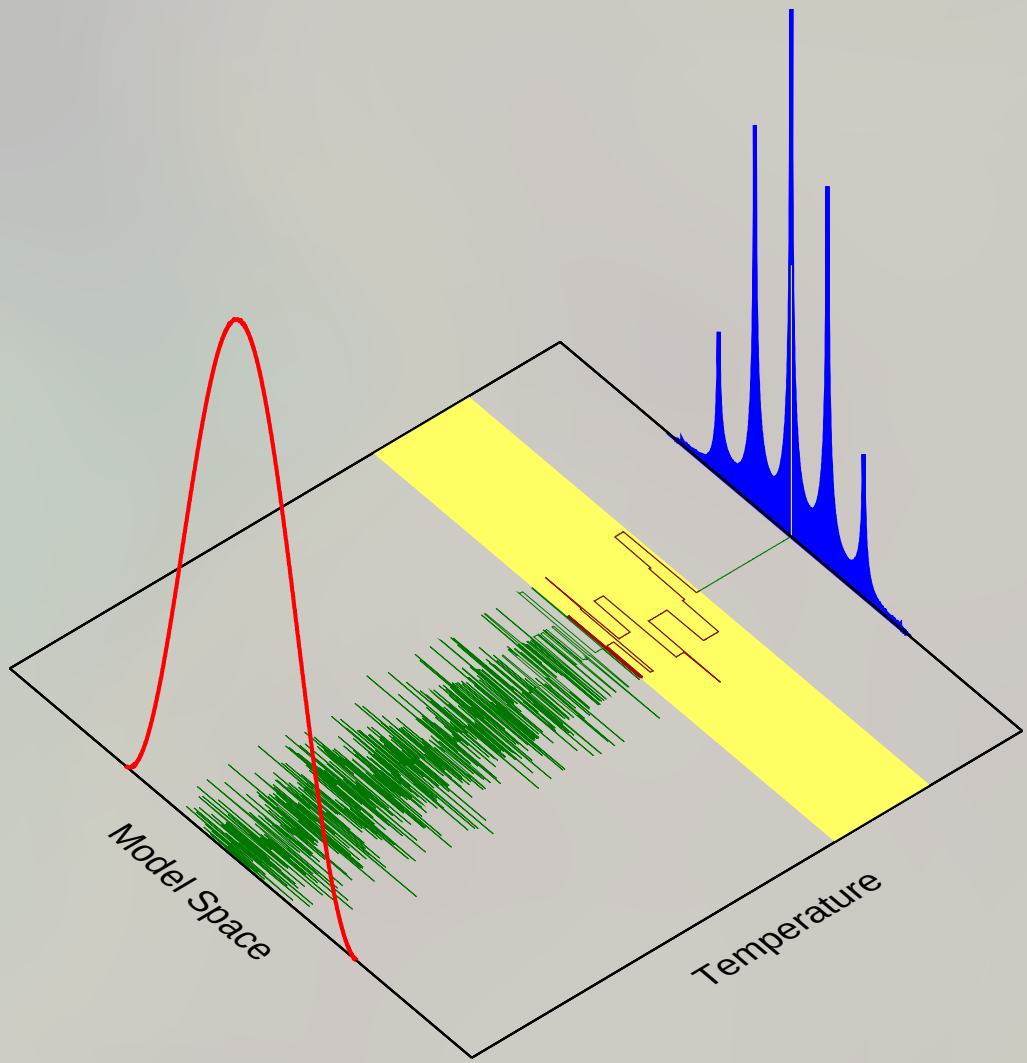
Physical application: fast cooling process



Cooling process - evolution of Boltzmana distribution



Cooling process - evolution of microscopic states



Simulated annealing – Kirkpatrick (1983)

Physical process $C \Rightarrow \bar{C}: \quad E(\bar{C}) = \min$

Boltzman distribution $\sigma(C,T) \Rightarrow \delta(\bar{C})$

$\downarrow$

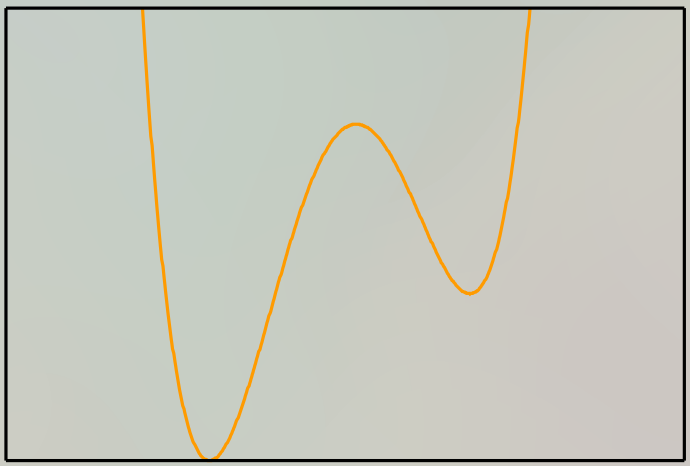
$\mathbf{m} \equiv C \quad \sigma(C,T) \Rightarrow e^{-\frac{S(\mathbf{m})}{T}}$

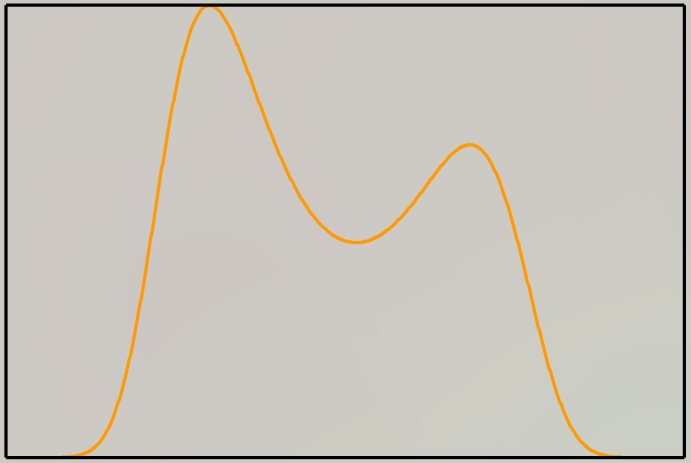
Optimization $S(\mathbf{m}) \Rightarrow S = \min \quad$ while $\quad T \to 0$

Simulated annealing – Kirkpatrick (1983)

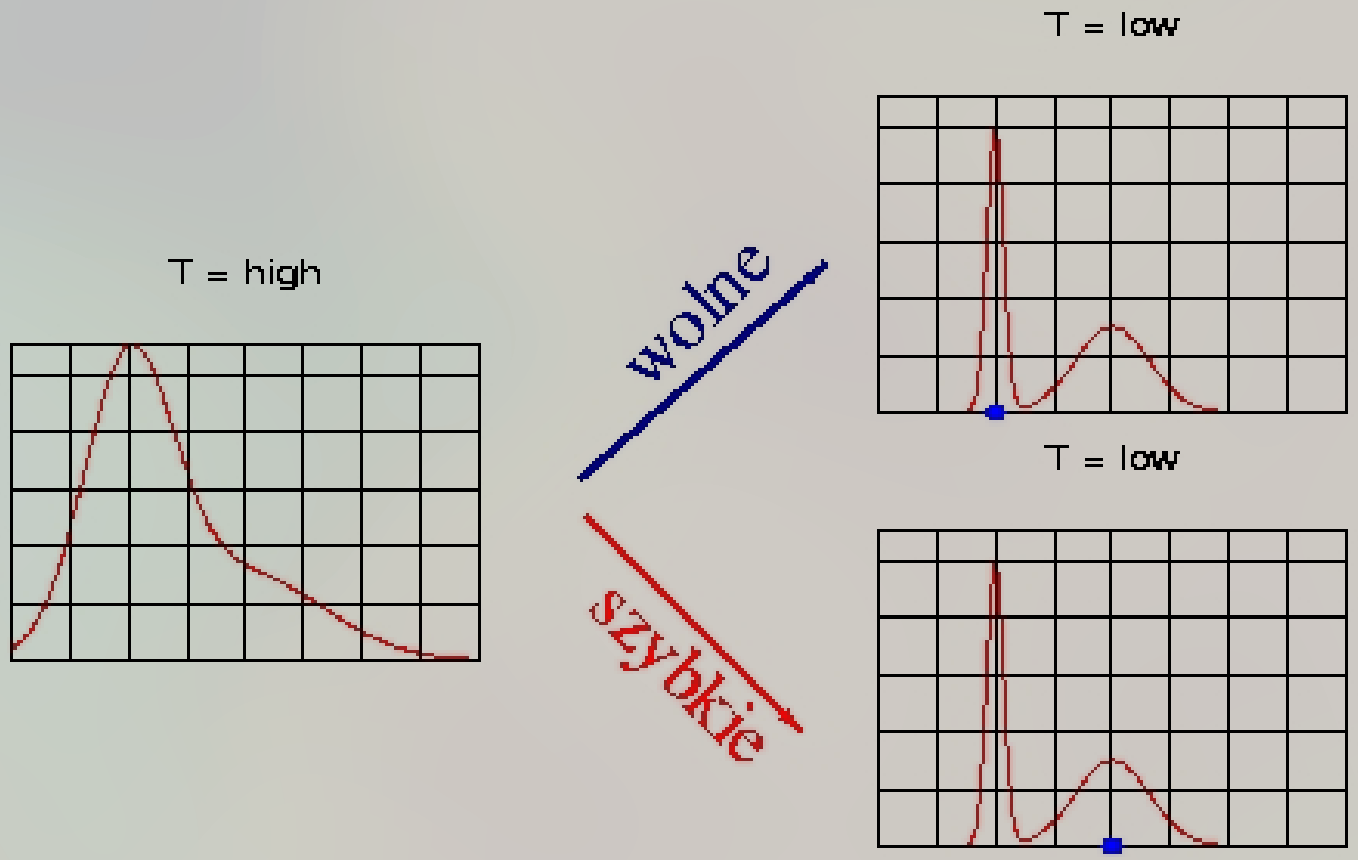
Szukamy absolutnego minimum dodatniej funkcji $S(\mathbf{m}) > 0$

$$S(\mathbf{m}) \Longrightarrow \sigma(\mathbf{m}, T) \;=\; e^{-\frac{S(\mathbf{m})}{T}}$$





Basic state and meta-stable states



Simulating annealing - optimization algorithm

Let assume that we can sample $\sigma(\mathbf{m}, T)$

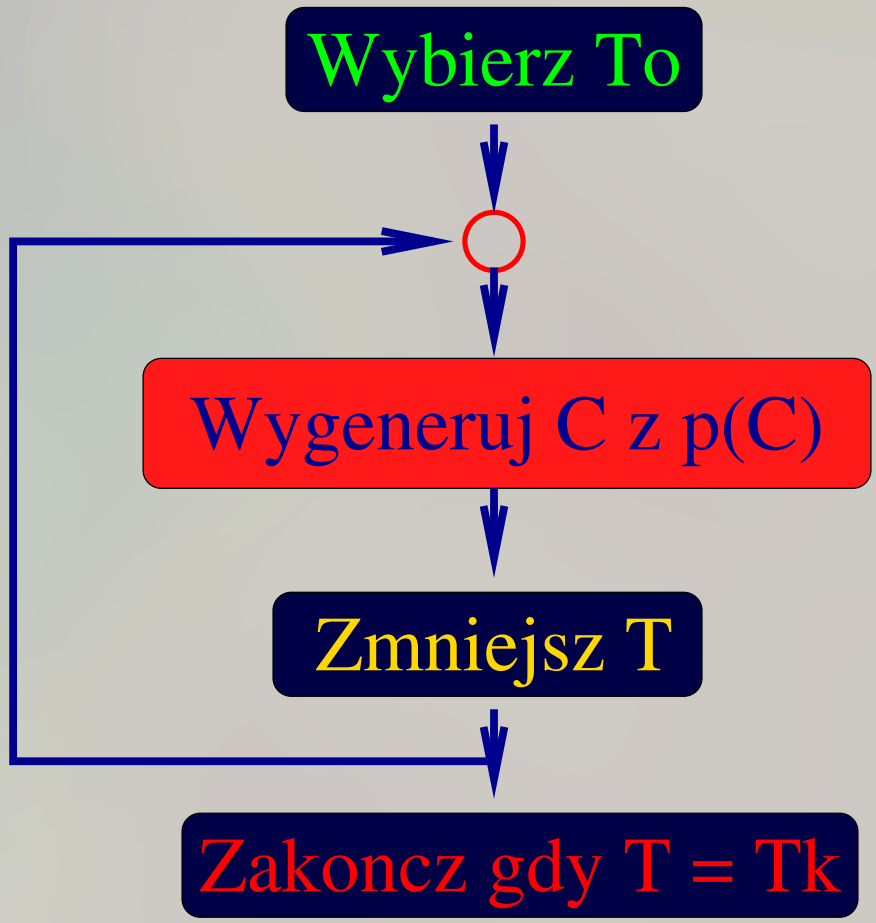
$$p(\mathbf{m}^{\alpha}) \;=\; \sigma(\mathbf{m}^{\alpha}, T)$$

then

$$T \to 0 \quad \Longrightarrow \quad \mathbf{m}^{\alpha} \to \mathbf{m}^{opt}$$

$$S(\mathbf{m}^{opt}) = \min$$

SA algorithm



Simulated annealing - elements

Two most important elements

1. generating sample from $p(\mathbf{m})$
2. scheduling of temperature T change

Efficient sampling - the Metropolis algorithm (generalized by Hastings)
Optimum cooling schedule (Kirkpatrick)

$$T_k = \frac{T_o}{\ln(k)}$$

Metropolisa-Hastings algorithm

- generate sample $\mathbf{m}^0$
- repeat
  - generate sample $\mathbf{m}^\beta = \mathbf{m}^\alpha + \delta\mathbf{m} \qquad (\delta\mathbf{m} - \text{random})$
  - generate random number from uniform distribution $u \sim U(0,1)$
  - if

$$u < P(\mathbf{m}^\alpha, \mathbf{m}^\beta) = \min\left[1, \frac{\sigma(\mathbf{m}^\beta)}{\sigma(\mathbf{m}^\alpha)}\right]; \exp\left(-\frac{S(\mathbf{m}^\beta) - S(\mathbf{m}^\alpha)}{T}\right)$$

$$\mathbf{m}^{\alpha+1} = \mathbf{m}^\beta$$

else

$$\mathbf{m}^{\alpha+1} = \mathbf{m}^\alpha$$

  - continue

Algorytm ASA (Ingber)

Ingber has proposed an enhancend (faster) ASA algorithm $\mathbf{m}^{\beta}$

$$\delta \mathbf{m}_i = r_i(T) \left(m_i^{max} - m_i^{min} \right)$$

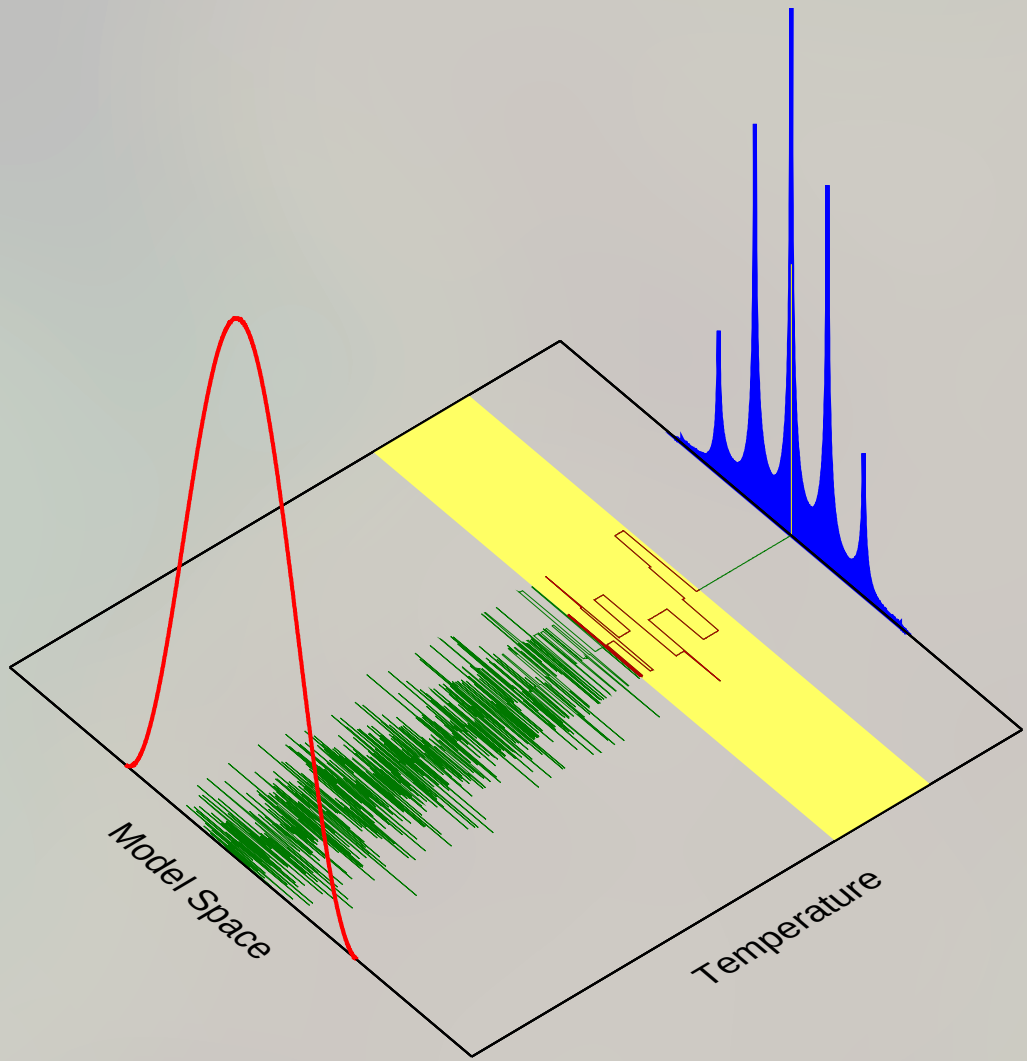
where random factor $r_(T)$ is drawn from the Cauchy-like distribution

$$p(\mathbf{m}, T) = \prod_i^N \frac{1}{2(|m_i| + T)\log(1 + 1/T)}$$

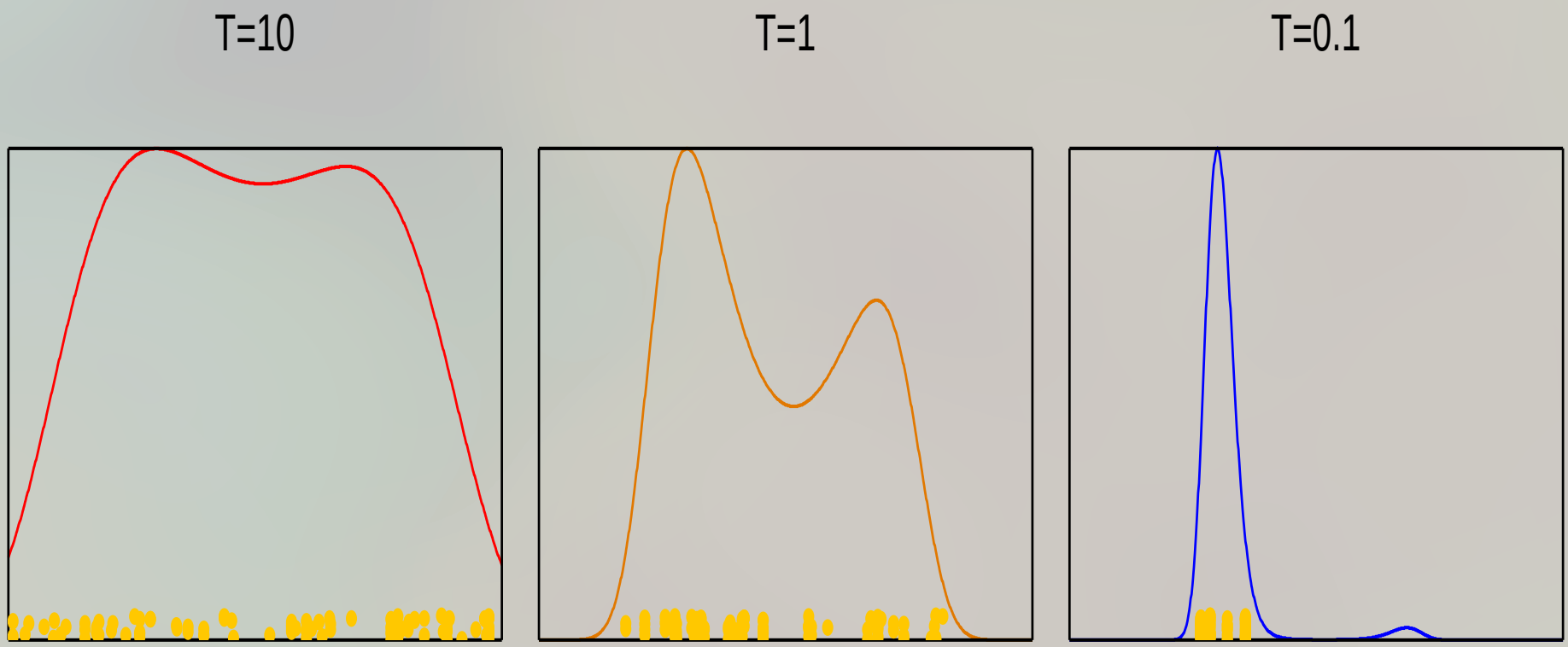
then cooling schedule

$$T_k = T_o \exp\left(-ck^{1/N}\right)$$

ASA algorithm - evolution of $\sigma(\mathrm{m})$



ASA algorithm - generated samples



End