Consensus-based Global Optimization

Claudia Totzeck

joint work with
Model: S. Martin, R. Pinnau, O. Tse
Analysis: J. A. Carrillo, Y.-P. Choi, O. Tse

AG Technomathematik
FB Mathematik
TU Kaiserslautern

10.10.2017

Global optimization problem

Let $f : \mathbb{R}^D \to \mathbb{R}^+, D \geq 1$, be a function. The problem we consider is given by

$$\operatorname*{argmin}_{x \in \mathbb{R}^D} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!

Outline:

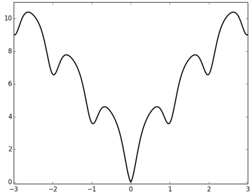
- Metaheuristics for global optimization
- Opinion dynamics & consensus formation
- CBO scheme for particles
- Expected mean-field equation
- Numerical validation of mean-field equation
- Analytical Results
- Numerical Results
- Digression: Pseudo-Inverse distribution function

Setting

Let $f : \mathbb{R}^D \to \mathbb{R}^+, D \geq 1$, be a function. The problem we consider is given by

$$\underset{x \in \mathbb{R}^D}{\operatorname{argmin}} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!



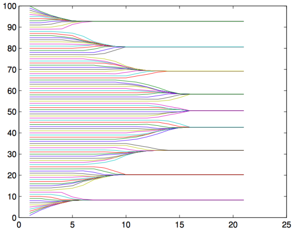
Metaheuristics:

- Random search, Rastrigin (1963)
- Simplex heuristic, Nelder–Mead (1965)
- Evolutionary programming, Fogel et al. (1966)
- Metropolis–Hastings, Hastings (1970)
- Genetic algorithm, Holland (1975)
- Simulated annealing, Kirkpatrick et al. (1983)
- Ant colony, Dorigo (1992)
- Particle swarm, Kennedy et al. (1995)

Consensus formation[1]

Main features:

- Involves multiple agents
- Each agent may adapt other opinions
- Agents may be reluctant to change
- Individual can influence neighbors



Example: Abelson's (averaging) model (1964) $\rightsquigarrow$ Cucker–Smale

$$X_{t+1}^i = \sum_j a_{ij}(X_t^i - X_t^j) = (AX_t)^i$$

where a_{ij} rate of influence of X_t^i over X_t^j

Example: Hegselmann–Krause (bounded confidence) model (2002)

$$X_{t+1}^i = \frac{1}{|N_t^i|} \sum_{j \in N_t^i} X_t^j, \quad N_t^i = \{j \,|\, |X_t^i - X_t^j| \leq \epsilon\}$$

[1] S. Chartterjee, E. Senata (1975, 1977), R. Hegselmann, U. Krause (2002),

Consensus-Based Global Optimization (CBO) Scheme

Combine methods of **swarm interaction** and **opinion dynamics** to obtain the following dynamic:

$$dX_t^i = -\lambda(X_t^i - v_f)\, H[f(X_t^i) - f(v_f)]dt + \sigma|X_t^i - v_f|dB_t^i, \quad i = 1, \dots, N,$$

where the **weighted average** v_f is given by

$$v_f = \frac{\sum_{i=1}^N X_t^i \omega_f^\alpha(X_t^i)}{\sum_{i=1}^N \omega_f(\alpha)(X_t^i)}, \quad \text{with} \quad \omega_f^\alpha(x) = \exp(-\alpha f(x)),$$

supplemented with random initial data $\rho_0 = \text{law}(X_0^i)$. We use the notation

B_t^i	Brownian Motion,
λ	drift parameter,
σ	diffusion parameter and
H	Heaviside function,
α	weight parameter.

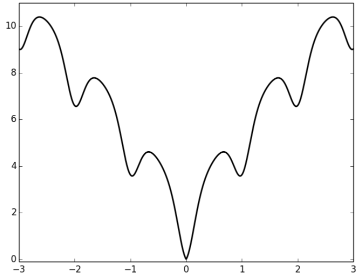
Note: The weighted average v_f is our candidate for the global minimizer!

CBO Particle Scheme

Let $f : \mathbb{R}^D \to \mathbb{R}, D \geq 1$, be a function. The problem we consider is given by

$$\underset{x \in \mathbb{R}^D}{\operatorname{argmin}} f(x).$$

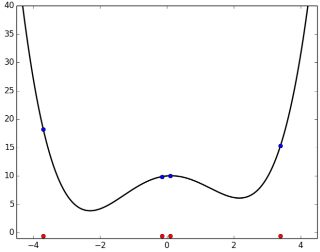
Task: Construct an algorithm that finds the global minimizer of f!



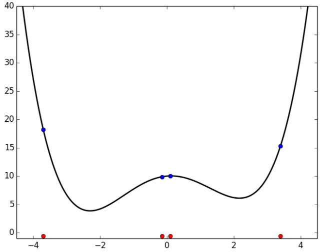
Features:

- global minimum
- particle scheme
- no gradient information
- indistiguishable particles
- mean-field equation for analysis

Importance of Stochasticity



$N = 4,\ \lambda = 1,\ \sigma = 0.0,\ \alpha = 30$
No stochastics!
level set at 9.7



$N = 4,\ \lambda = 1,\ \sigma = 0.7,\ \alpha = 30$
with stochastics

Mean-Field Equation

There is no local or global information involved in the dynamic, i.e., the particles are indistinguishable $\rightarrow$ allows for formal passage to the mean-field limit:

(Postulated) Mean-Field system

As $N \rightarrow \infty$ the SDE system turns into the McKean nonlinear Process

$$d\bar{X}_t = -\lambda(\bar{X}_t - v_f)H[f(X_t) - f(v_f)]dt + \sqrt{2}\sigma|\bar{X}_t - v_f|dB_t,$$

with initial data $\text{law}(\bar{X}_0) = \rho_0$. Here the weighted average is given by

$$v_f = \frac{1}{\int_{\mathbb{R}^d} \omega_f^\alpha d\rho_t} \int_{\mathbb{R}^D} x\omega_f^\alpha d\rho_t, \qquad \rho_t = \text{law}(\bar{X}_t),$$

which we express equivalently by the **nonlocal**, **nonlinear degenerate** PDE

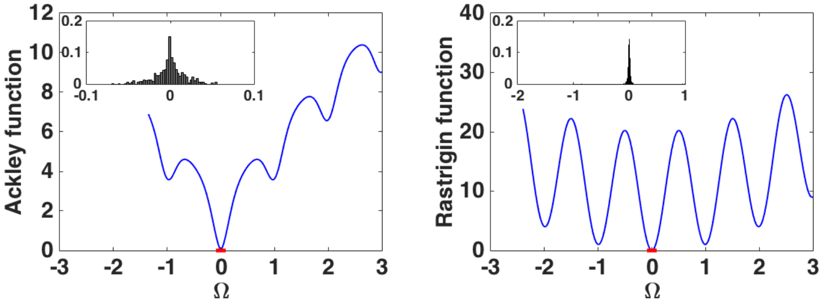
$$\partial_t \rho_t = \Delta(\kappa\rho_t) + \nabla \cdot (\mu\rho_t),$$

with $\qquad \kappa = \sigma^2|x - v_f|^2$ and $\mu = -\lambda(x - v_f)H[f(x) - f(v_f)]$.

Comparison of mean-field and particle simulations

Open Problem: Rigorous proof of the mean-field equation as $N \to \infty$.

Numerical justification:



Computations with 500 samples, $T = 80$.
Histograms show the distribution of the weighted average v_T.
$\text{supp}\rho_T = \{x \in \mathbb{R} : \rho_T > 1e-6\}$.

Analytical Results

- well-posedness of the particle system for locally Lipschitz cost functions
- well-posedness of the mean-field equation
  - for bounded cost functions
  - for cost functions with quadratic growth at infinity

Large time behavior and consensus formation

Let $f \in \mathcal{C}^2(\mathbb{R})$, $\inf f > 0$, $\|\nabla^2 f\|_\infty \le c_f$ and $\Delta f \le c_0 + c_1 |\nabla f|^2$ in $\mathbb{R}^d$ for $c_0, c_1 > 0$.

Then, for any given $0 < \epsilon \ll 1$ arbitrarily small, there exist some $\alpha \gg 1$ and appropriate parameters (λ, σ) such that uniform consensus is obtained at a point $\tilde{x} \in B_\epsilon(x_*)$, i.e.,

$$\rho_t \to \delta_{\tilde{x}} \text{ for } t \to \infty \text{ with } \tilde{x} \in B_\epsilon(x_*).$$

Moreover, we can show that $V(\rho_t) \le V(\rho_0) e^{-qt}$ for some $q = q(\lambda, \sigma) > 0$.

Numerical Results - Success Rates for different N

x_*		N 50	100	200
0	success rate	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$5.21e^{-4}$	$1.18e^{-3}$	$2.47e^{-3}$
1	success rate	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$5.23e^{-4}$	$1.21e^{-3}$	$2.55e^{-3}$
2	success rate	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$5.46e^{-4}$	$1.24e^{-3}$	$2.57e^{-3}$

Ackley function:
20 dimensions,
$\alpha = 30,\ T = 10,$
$\sigma = 5,\ dt = 0.01$

Rastrigin function:
20 dimensions,
$\alpha = 30,\ T = 10,$
$\sigma = 5,\ dt = 0.01$

x_*		N 50	100	200
0	success rate	34.%	61.1%	62.2%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$3.12e^{-1}$	$2.47e^{-1}$	$2.42e^{-1}$
1	success rate	34.5%	57.1%	61.6%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$3.09e^{-1}$	$2.52e^{-1}$	$0.244e^{-1}$
2	success rate	35.5%	54.8%	62.4%
	$\frac{1}{d}\mathbb{E}[\Vert v_f(T) - x_*\Vert^2]$	$3.06e^{-1}$	$2.51e^{-1}$	$2.44e^{-1}$

Numerical Results - Success Rates for different α

x_*		α = 10	20	30	40	50
0	success rate	100%	100%	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	2.55e−4	1.06e−4	6.18e−5	4.21e−5	3.04e−5
1	success rate	100%	100%	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	2.58e−4	1.09e−4	6.31e−5	4.24e−5	3.04e−5
2	success rate	100%	100%	100%	100%	100%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	2.62e−4	1.10e−4	6.46e−5	4.35e−5	3.18e−5

Ackley function: 20 dimensions, $N = 50$, $T = 10$, $\sigma = 5$, $dt = 0.01$

---

x_*		α = 10	20	30	40	50
0	success rate	7.0%	11.6%	61.1%	93.8%	99.7%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	4.32e−1	3.91e−1	2.48e−1	1.35e−1	7.67e−2
1	success rate	7.5%	10.3%	57.1%	96%	99.5%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	4.34e−1	3.94e−1	2.52e−1	1.32e−1	7.59e−2
2	success rate	7.5%	12.2%	54.8%	94.9%	99.3%
	$\frac{1}{d}\mathbb{E}[\|v_f(T) - x_*\|^2]$	4.31e−1	3.92e−1	2.51e−1	1.32−1	7.81e−2

Rastrigin function: 20 dimensions, $N = 50$, $T = 10$, $\sigma = 5$, $dt = 0.01$

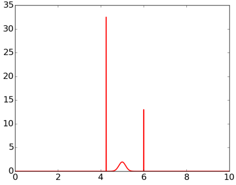
First approaches for the convergence proof in $1D$

Cumulative distribution F_t

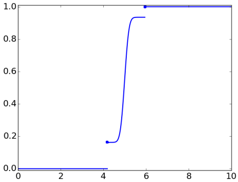
We define the cumulative distribution F_t by

$$F_t(x) = \int_{-\infty}^{x} d\rho_t = \rho_t((-\infty, x])$$

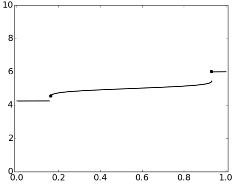
and the pseudo-inverse χ_t of F_t.



measure μ



distribution F



pseudo-inverse χ

Pseudo-Inverse Distribution function in $1D$

General properties of χ:

- tracking the boundary points $\chi(0)$ and $\chi(1)$ allows for proofs of concentration
- when dirac-distributions appear, the pseudo-inverse formulation is better suited for numerics

CBO formulated in terms of the pseudo-inverse

Using the pseudo-inverse χ_t of F_t we derive an equivalent formulation of the mean-field equation:

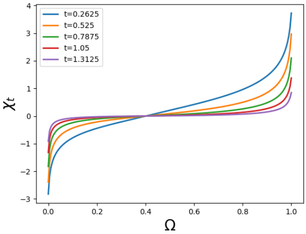
$$\partial_t \chi_t + \mu_t = -\partial_\eta (\kappa_t (\partial_\eta \chi_t)^{-1},$$

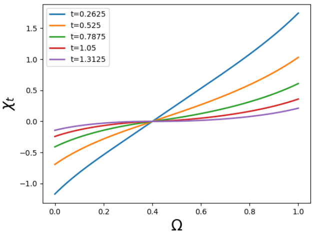
where

$$\mu_t = \lambda(\chi_t - m_f[\chi_t]), \kappa_t = (\sigma^2/2)|\chi_t - m_f[\chi_t]|^2,$$

$$m_f[\chi_t] = \frac{\int_0^1 \chi_t(\eta) \exp(-\alpha f(\chi_t(\eta))) d\eta}{\int_0^1 \exp(-\alpha f(\chi_t(\eta))) d\eta}.$$

Numerical Simulations



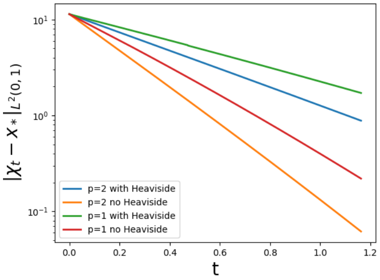


p = 1 p = 2

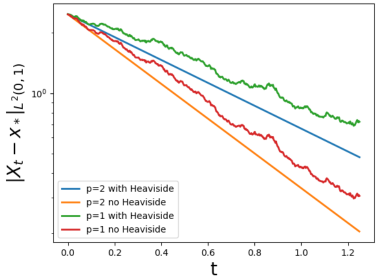
Porous media-type version

$$\partial_t \chi_t + \mu_t = -\partial_\eta(\kappa_t(\partial_\eta \chi_t)^{-p},$$
$$p \in \mathbb{N},\ p \geq 1.$$

Numerical Simulations - convergence

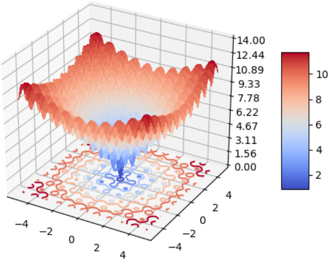


error plot for pseudo-inverse with $p = 1$ in $1d$

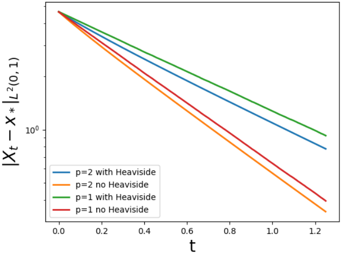


error plot for particle scheme with $p = 1$ in $1d$

Particle Scheme $p = 2$ in $2D$



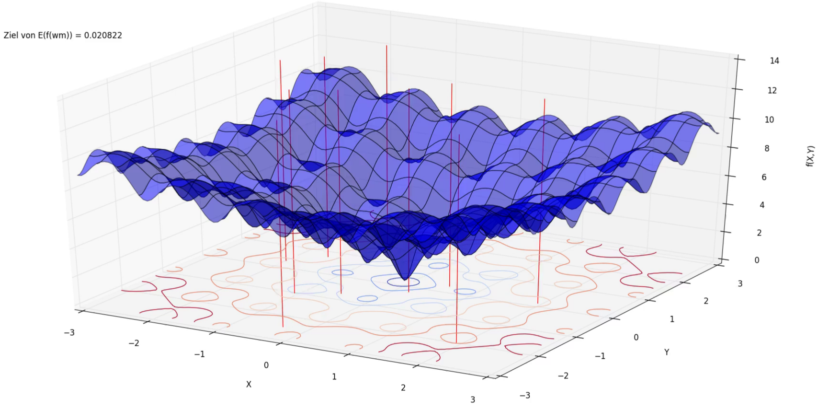
Ackley function in $2d$



particle scheme for $p = 2$ in $2d$, average of 100 MC runs

Unfortunately, the extension of the particle scheme for $p > 2$ is nontrivial.

Particle Scheme $p = 1$ in $2D$



$N = 10,\ \lambda = 2,\ \sigma = 0.7,\ \alpha = 30$

Take Home Message

Consensus-Based Optimization problem

- scheme using indistinguishable particles
  (no local / global best information)
- no gradient information

global optimization + swarming/consensus modeling

$\Rightarrow N \to \infty$ convergence proof using PDE techniques

Thank you for your attention!

References:

- **Model:** Pinnau, T., Tse, Martin (M3AS, 2017)
  A Consensus-Based Model for Global Optimization and Its Mean-Field Limit
- **Analysis:** Carrillo, Choi, T., Tse (submitted, preprint at arXiv)
  An Analytical Framework for a Consensus-Based Global Optimization Method

✉ totzeck@mathematik.uni-kl.de