Consensus-Based Optimization Beyond Finite-Time Analysis

Speaker : **V. Priser**[1]
Co-authors: P. Bianchi[1] R.-A. Dragomir [1]

[1]LTCI, Télécom Paris

Objective

Problem

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f a non-convex with unknown gradient function, and with one unique minimizer x_*.

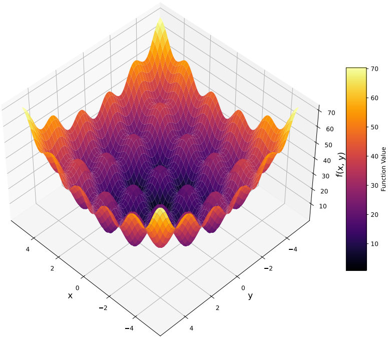


Figure: Example of function f

Solutions

Algorithm: Generate particles to reach the minimizer of f.

- Simulated Annealing Pelletier 1998.
- Genetic and Evolutionary Algorithms Holland 1975.
- Bayesian Optimization Hastings 1970.
- Particle Swarm Optimization (PSO) Kennedy and Eberhart 1995.

Applications: Machine learning, Signal processing, ...
Problem: Lack of theoretical results for these algorithms.

PSO Kennedy and Eberhart 1995

Idea: At time step $k+1$, we want all particles $(X_{k+1}^i)_{i\leq n}$ to move closer to the best particle among all previous positions $(X_j^i)_{j\leq k,\, i\leq n}$.

- Personal best: $p_k^i := \underset{x\in\{X_0^i,\dots,X_k^i\}}{\arg\min} \; f(x)$
- Global best: $g_k := \underset{p\in\{p_k^1,\dots,p_k^n\}}{\arg\min} \; f(p)$

Vanilla PSO Algorithm

$$
\begin{aligned}
v_{k+1}^i &= v_k^i + \eta_1 \big(r_1 \left(p_k^i - X_k^i\right) + r_2 \left(g_k - X_k^i\right)\big),\\
X_{k+1}^i &= X_k^i + \eta_2 \, v_{k+1}^i,
\end{aligned}
$$

where $r_1, r_2 \sim \mathcal{U}(0,1)$ are independent random coefficients.

Problem:

- Heuristic algorithm with few theoretical guarantees.
- We know how to analyze particle systems of the form

$$X_{k+1}^i = F(X_k^1, \dots, X_k^n),$$

but the PSO update rule does not directly fit into this framework.

CBO: a simplification of PSO

$X_k^1, \ldots, X_k^n$: particles generated at time k.
At time $k+1$, we want each particle to move closer to the best particle at time k.

$$\begin{cases} X_{k+1}^1 = X_k^1 + \eta\big(\underset{x \in \{X_k^1, \ldots, X_k^n\}}{\arg\min} f(x) - X_k^1\big) \\ \quad\vdots \\ X_{k+1}^n = X_k^n + \eta\big(\underset{x \in \{X_k^1, \ldots, X_k^n\}}{\arg\min} f(x) - X_k^n\big) \end{cases}$$

Problem: The operator arg min is irregular.

Laplace principle

$$C_\alpha^n := \frac{\sum_{i=1}^n X_k^i \exp(-\alpha f(X_k^i))}{\sum_{i=1}^n \exp(-\alpha f(X_k^i))} \to_{\alpha \to \infty} \underset{x \in \{X_k^1, \ldots, X_k^n\}}{\arg\min} f(x)$$

Consensus-Based Optimization algorithm (CB0) (Pinnau et al. 2017):

$$\forall i \leq n, \quad X_{k+1}^i = X_k^i + \eta\left(C_\alpha^n - X_k^i\right) + \underbrace{\sqrt{\eta}\sigma_k \xi_k^i}_{\text{noise for exploration}}$$

Behavior of CBO

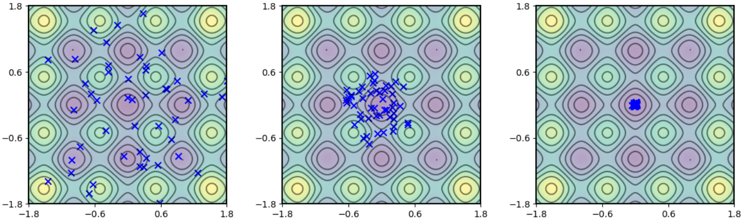


Figure: Demonstration of CBO with $N = 50$ particles on a 2D function.

Literature review

- 1995–today: Development of the PSO algorithm and its variants Kennedy and Eberhart 1995.
- 2013: PSO where the memory is discarded (almost CBO) Askari-Sichani and Jalili 2013.
- 2017: Introduction of the CBO algorithm with smooth $\arg\min$ Pinnau et al. 2017.
- 2018: First convergence proof of the CBO algorithm in the infinite-particle regime ($n :=$ Number of Particles $\to \infty$) under restrictive assumptions Carrillo et al. 2018.
- 2021–2023: Extension of Carrillo *et al.*'s result to the finite-particle regime ($n < \infty$) Hu et al. 2021; Gerber, Hoffmann, and Vaes 2023.
- 2024: New convergence proof in the infinite-particle regime that relaxes Carrillo *et al.*'s assumptions Fornasier, Klock, and Riedl 2024.
- 2024–today: Further improvements of Fornasier *et al.*'s results.
- 2017–today: Development of CBO variants for various tasks, such as sampling, saddle-point search, and constrained optimization.

The CBO algorithm of Fornasier, Klock, and Riedl 2024

Algorithm Fornasier, Klock, and Riedl 2024

$$\forall i \leq n, \quad X_{k+1}^i = X_k^i + \eta \left(C_\alpha^n - X_k^i\right) + \sqrt{\eta} \left\| C_\alpha^n - X_k^i \right\| \xi_k^i \, .$$

In the population limit ($n \to \infty$), particles $X_k^1, \ldots, X_k^n$ are i.i.d.:

$$C_\alpha^n = \frac{\sum_{i=1}^n X_k^i \exp(-\alpha f(X_k^i))}{\sum_{i=1}^n \exp(-\alpha f(X_k^i))} \underbrace{\underset{n \to \infty}{\to} \frac{\mathbb{E}(X_k^1 \mathrm{e}^{-\alpha f(X_k^1)})}{\mathbb{E}(\mathrm{e}^{-\alpha f(X_k^1)})}}_{\text{Law of large number}} =: C_\alpha(\mathrm{Law}(X_k^1)) \, .$$

Mean field Equation ($\eta \to 0, n \to \infty$)

$$dX_t = (C_\alpha(\mathrm{Law}(X_t)) - X_t)dt + \| C_\alpha(\mathrm{Law}(X_t)) - X_t \| \, dB_t$$

B_t: Brownian motion

Result of Fornasier, Klock, and Riedl 2024

$x_* := \arg\min f$.

Result in the mean field limit (Fornasier, Klock, and Riedl 2024)

For some $t \leq T(\alpha)$:

$$W_2(\text{Law}(X_t), \delta_{x_*}) \leq \mathrm{e}^{-Ct} W_2(\text{Law}(X_0), \delta_{x_*}) \, .$$

Choosing α large s. t. $\mathrm{e}^{-CT(\alpha)} \leq \frac{\varepsilon}{W_2(\text{Law}(X_0), \delta_{x_*})}$, we reach an ε precision.

Result of the finite particle regime (Fornasier, Klock, and Riedl 2024)

n: number of particles
For $k \leq \min(K(\alpha), \log n)$ and :

$$W_2(\text{Law}(X_k^i), \delta_{x_*}) \leq \mathrm{e}^{-Ck} W_2(\text{Law}(X_0^i), \delta_{x_*}) \, .$$

Problem: In the finite particles regime, the number of iterations is bounded by a quantity depending on α and n.

Literature's gap

Convergence of CBO:

	Low value of k	**Large value of k**
$n = \infty$	Fornasier, Klock, and Riedl 2024	Carrillo et al. 2018
$n < \infty$	Fornasier, Klock, and Riedl 2024	Our result

- In the case $n = \infty$ with large k, the result of Carrillo et al. 2018 holds under restrictive conditions on the initial distribution.
- In the case $n < \infty$ and for small k, convergence is guaranteed only when $k \ll n$ and $k \leq K(\alpha)$.

Limitation of Fornasier, Klock, and Riedl 2024

Mean field: $dX_t = (C_\alpha(\mathrm{Law}(X_t)) - X_t)dt + \|C_\alpha(\mathrm{Law}(X_t)) - X_t\|\, dB_t$
Assume that $\mathbb{P}(\|X_t - x_*\| \leq \varepsilon) > c$:

$$C_\alpha(\mathrm{Law}(X_t)) = \frac{\mathbb{E}(X_t \mathrm{e}^{-f(X_t)})}{\mathbb{E}(\mathrm{e}^{-\alpha f(X_t)})} \simeq \frac{x_* \mathrm{e}^{-\alpha f(x_*)}\mathbb{P}(\|X_t - x_*\| \leq \varepsilon) + \mathrm{e}^{-\alpha(f(x_*)+\tilde{\varepsilon})}}{\mathrm{e}^{-\alpha f(x_*)}\mathbb{P}(\|X_t - x_*\| \leq \varepsilon)} \to x_*$$

Under, $\mathbb{P}(\|X_t - x_*\| \leq \varepsilon) > c$:

$$dX_t \simeq (x_* - X_t)dt + \|x_* - X_t\|\, dB_t\,, \tag{1}$$

which means $\mathrm{Law}(X_t) \to \delta_{x_*}$.

Problem

- $\mathbb{P}(\|X_t - x_*\| \leq \varepsilon) > c$: true when X_t has a density with large enough variance.
- $\|C_\alpha(\mathrm{Law}(X_t)) - X_t\| \to 0 \implies$ the variance of X_t decreases.
- At some time T, $\mathbb{P}(\|X_t - x_*\| \leq \varepsilon) < c$ and the approximation Eq. (1) does not hold.

Our version of CBO

We want a long-time result on CBO.

Idea

We solve the issue raised in Fornasier, Klock, and Riedl 2024 by using fixed-variance noise.

- Noise with excessively large variance $\implies$ particles are likely to be too far from x_*.
- Variance need to be large enough : $\gamma > \gamma_0$.

Our algorithm

For every $i \leq n$:

$$X_{k+1}^i = X_k^i + \eta \left(C_\alpha^n - X_k^i) \right) + \underbrace{\sqrt{2\eta \frac{\gamma}{\alpha}} \xi_k^i}_{\text{fix variance}}$$

Main result

$\alpha \gg 1$: parameter of algorithm (needed for the Laplace principle).

Our result

$$\mathbb{E} \left\| X_k^i - x_* \right\| \leq C \left(\frac{1}{\sqrt{n}} + \sqrt{\eta} + (1+c)^{-k} \right)$$

- A similar result holds in the Mean Field regime.
- $\gamma > \gamma_0$: the noise has a variance large enough.

Solution of the Mean Field equation

Mean Field Equation

$$dX_t = (C_\alpha(\mathrm{Law}(X_t)) - X_t)dt + \sqrt{2\frac{\gamma}{\alpha}}dB_t$$

It is an Ornstein–Uhlenbeck equation with the solution

$$X_t = x_t + \mathcal{N}(0, \frac{\gamma}{\alpha} I_d) + \underbrace{X_0 \mathrm{e}^{-t}}_{\text{small when } t \gg 1}$$

with

$$\dot{x}_t = C_\alpha(\mathrm{Law}(X_t)) - x_t \, .$$

For large t, $\mathrm{Law}(X_t) \simeq \mathcal{N}(x_t, \frac{\gamma}{\alpha} I_d)$ and by the Laplace principle:

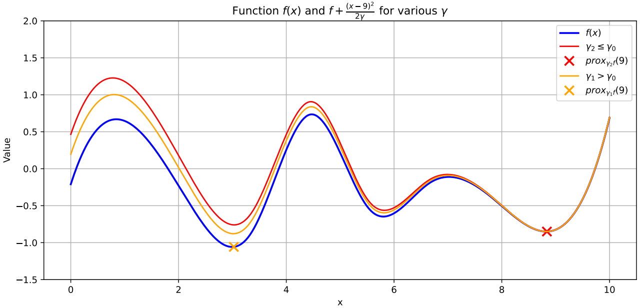
$$C_\alpha(\mathrm{Law}(X_t)) \simeq \frac{\int x \mathrm{e}^{-\alpha\left(f(x) + \frac{\|x - x_t\|^2}{2\gamma}\right)} dx}{\int \mathrm{e}^{-\alpha\left(f(x) + \frac{\|x - x_t\|^2}{2\gamma}\right)} dx} \underset{\alpha \to \infty}{\to} \underbrace{\operatorname*{arg\,min}_{x \in \mathbb{R}^d} \left(f(x) + \frac{\|x - x_t\|^2}{2\gamma} \right)}_{\mathrm{prox}_{\gamma f}(x_t)}$$

With $\mathbb{E}\|\varepsilon_t\| \ll 1$:

$$\dot{x}_t = (\mathrm{prox}_{\gamma f}(x_t) - x_t) + \varepsilon_t$$

The proximal operator

Problem: $\mathrm{prox}_{\gamma f}(x) := \arg\min\{f + \|\cdot - x\|^2 / 2\gamma\}$ requires f convex.



- For γ large enough, $\mathrm{prox}_{\gamma f}(x)$ is sufficiently close to x_*.
- The function f is convex in a neighborhood of x_*.
- For γ large enough, $\mathrm{prox}_{\gamma f}(x)$ lies within the convex region of f.
- Conclusion: There exists a convex function $\bar{f}$ such that $\bar{f} = f$ in a neighborhood of x_* and $\mathrm{prox}_{\gamma f} = \mathrm{prox}_{\gamma \bar{f}}$ for large enough γ.

Alternative point of view: The noise must be sufficiently large to escape local minima.

Convergence of Mean field Equation

The ODE:

$$\dot{x}_t = \text{prox}_{\gamma \bar{f}}(x_t) - x_t + \varepsilon_t \,,$$

is a continuous version of the Proximal Point Method:

$$x_{k+1} = \text{prox}_{\gamma \bar{f}}(x_k) \Leftrightarrow x_{k+1} - x_k = \text{prox}_{\gamma \bar{f}}(x_k) - x_k \,.$$

$\bar{f}$ is convex then

$$\|x_t - x_*\| \leq \|x_0 - x_*\| \, \mathrm{e}^{-Ct} + \sup_s \mathbb{E} \, \|\varepsilon_s\| \,.$$

Recall: $X_t = \mathcal{N}(x_t, \frac{\gamma}{\alpha} I_d)$

Convergence of the Mean Field equation

$$\mathbb{E} \left\| X_t - \mathcal{N}(x_*, \frac{\gamma}{\alpha} I_d) \right\| \leq \mathbb{E} \left\| X_0 - \mathcal{N}(x_*, \frac{\gamma}{\alpha} I_d) \right\| \mathrm{e}^{-Ct} + \underbrace{\sup_s \mathbb{E} \, \|\varepsilon_s\|}_{\frac{C}{\sqrt{\alpha}}}$$

or

$$\mathbb{E} \, \|X_t - x_*\| \leq \frac{C}{\sqrt{\alpha}} + \mathrm{e}^{-Ct} \mathbb{E} \, \|X_0 - x_*\| \,,$$

The finite particle error

The finite particle algorithm

$$\forall i \leq n \quad dX_t^i = (C_\alpha^n - X_t^i)dt + \sqrt{2\frac{\gamma}{\alpha}}dB_t^i\,.$$

The Mean Field equation

$$dX_t = (C_\alpha(\mathrm{Law}(X_t)) - X_t)dt + \sqrt{2\frac{\gamma}{\alpha}}dB_t\,.$$

If $\mathrm{Law}(X_0) = \mathrm{Law}(X_0^i)$:

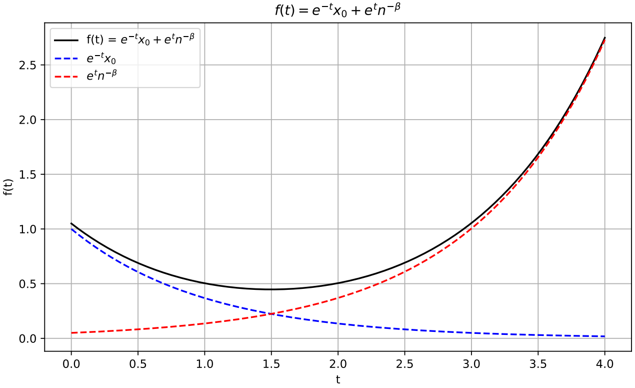
$$W_2(\mathrm{Law}(X_t), \mathrm{Law}(X_t^i)) \leq \frac{\mathrm{e}^{Ct}}{\sqrt{n}}\,.$$

This is a classical issue arising from the use of Grönwall's lemma.

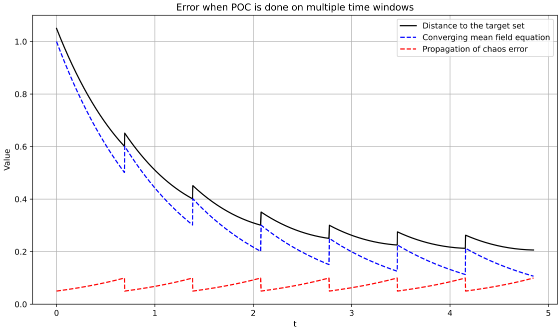
Putting the results together

$$W_2(\mathrm{Law}(X_t^i), \mathcal{N}_\alpha) \leq \underbrace{\mathrm{e}^{-Ct} W_2(\mathrm{Law}(X_0^i), \mathcal{N}_\alpha)}_{\text{Mean Field Equation}} + \underbrace{\frac{\mathrm{e}^{Ct}}{\sqrt{n}}}_{\text{finite particle error}} .$$

Problem: The exponential term in red diverge as $t \to \infty$.



Iterating over multiple time intervals



For T large enough, for any $t \geq 0$,

$$W_2(\mathrm{Law}(X^i_{t+T}), \mathcal{N}_\alpha) \leq \frac{1}{2} W_2(\mathrm{Law}(X^i_t), \mathcal{N}_\alpha) + \frac{e^{CT}}{\sqrt{n}}.$$

Repeat this over N time intervals of width T: $[0, T], \ldots, [(N-1)T, NT]$.

Conclusion

- By using noise with fixed variance, we address an open question regarding the long-time convergence of the CBO algorithm.
- A convergence proof that does not rely on the mean-field equation would provide deeper insight. At this stage, we cannot claim that CBO offers a clear advantage over a simple random search.