Under-Informed Momentum in PSO

Christopher K. Monson, Kevin D. Seppi
Google, Inc. and Brigham Young University

PSO, Distributions, and Determinism

PSO is a stochastic hill climber that iteratively applies simple motion equations to each particle, given here in "inertia weight" form:

$$\mathbf{v}_{t+1} = \omega \mathbf{v}_t + \phi_p \mathbf{U} \circ (\mathbf{p}_t - \mathbf{x}_t) + \phi_g \mathbf{U} \circ (\mathbf{g}_t - \mathbf{x}_t) \qquad \mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{v}_{t+1}$$

Where typically $\omega = 0.72984$ and $\phi_p = \phi_g \approx 1.5$. Each $\mathbf{U}$ is a random vector whose elements are independent draws from a Standard Uniform distribution, and $\circ$ represents element-wise multiplication. The terms $\mathbf{p}$ and $\mathbf{g}$ are, respectively, the individual particle's best known position and the best known position among its neighbors.

In terms of explicit random variables:

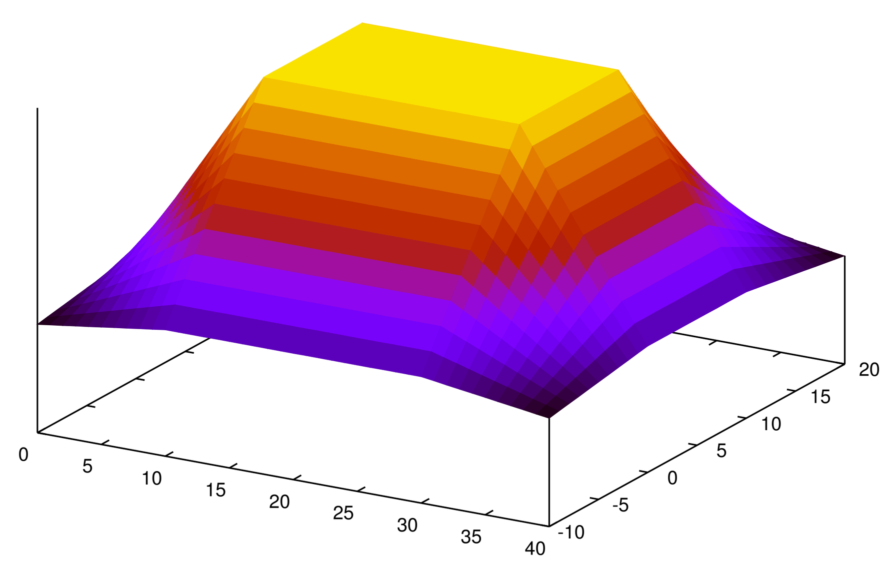
$$\mathbf{P}_{t+1} \sim \mathbf{U}\left[\mathbf{0}, \phi_p(\mathbf{p}_t - \mathbf{x}_t)\right] \qquad \mathbf{G}_{t+1} \sim \mathbf{U}\left[\mathbf{0}, \phi_g(\mathbf{g}_t - \mathbf{x}_t)\right] \qquad \mathbf{v}_{t+1} = \omega \mathbf{v}_t + \mathbf{P}_{t+1} + \mathbf{G}_{t+1}$$

$\mathbf{P}$ and $\mathbf{G}$ are each sampled from a multivariate uniform distribution within the specified hyperrectangle, then used to produce a new velocity.

Re-written as a single distribution $\mathbf{C}$ from the convolution of $\mathbf{P}$ and $\mathbf{G}$:

$$\mathbf{\Theta}_{t+1} \sim \mathbf{C}\left[\phi_p(\mathbf{p}_t - \mathbf{x}_t), \phi_g(\mathbf{g}_t - \mathbf{x}_t)\right] \qquad \mathbf{v}_{t+1} = \omega \mathbf{v}_t + \mathbf{\Theta}_{t+1}$$

That this is mathematically equivalent to the original pso equations. $\mathbf{C}$ can approach the Normal Distribution in the limit of many similar convolutions and is thus reminiscent of the "Bare Bones" PSO algorithm. Note how *momentum is not like the other terms*. It is not directly *informed* by the fitness function, and it does not merely *contribute* information to the final position: instead it arbitrarily shifts the position Θ. **It is highly impactful while being woefully underinformed.**



Diversity Without Momentum

Momentum adds diversity to the system, but has detrimental side effects.

This observation may explain the many published approaches for tuning the inertia weight.

Diversity can be added without these side effects by removing momentum and employing more selective techniques.

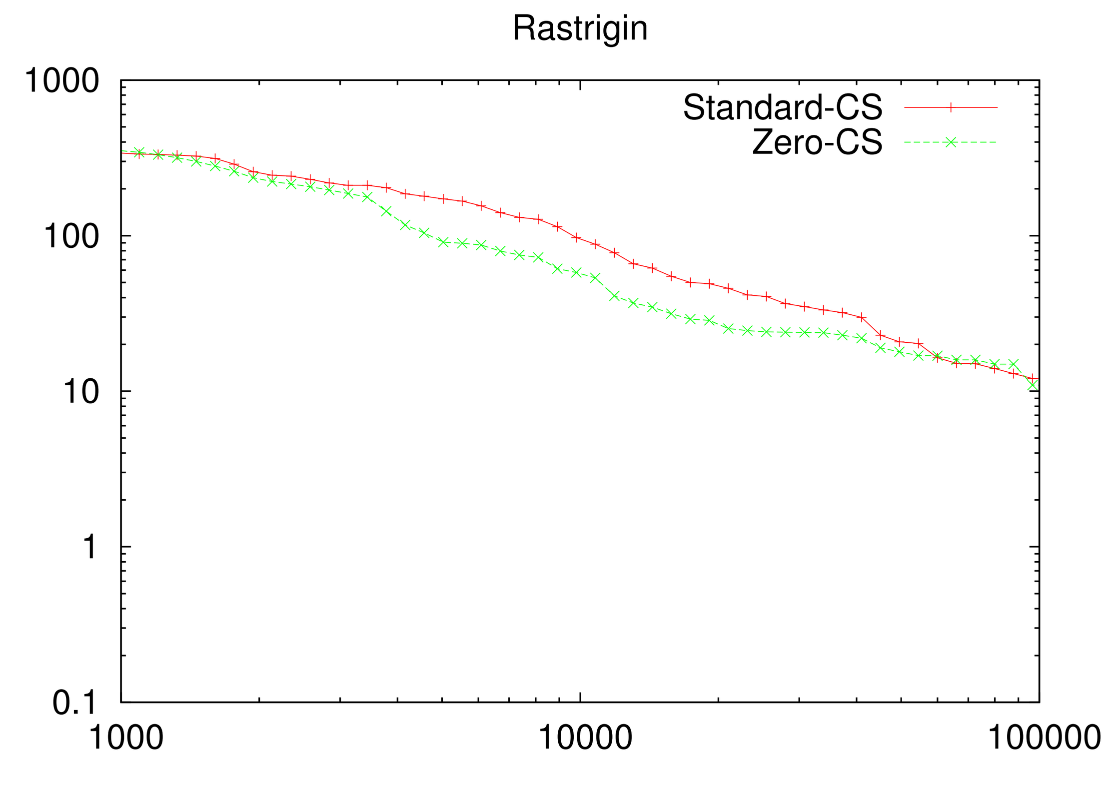
Zero momentum (including removal of constriction and restoration of $\phi_g = \phi_p = 2.05$) produces lackluster results unless *replaced* with other diversity enhancements. This is uncommon in the literature and zero momentum has thus not received due attention.

Results

Stagnation-prevention methods like CRIBS and its predecessors are effective diversity enhancements. They provide intuitive control over exploration / exploitation trade-offs and are effective at stagnation avoidance. Methods like SAC, which employ fitness history feedback to adjust social and cognitive terms, also help with appropriately-applied diversity. These can replace momentum.

Zero momentum on the Sphere function generates decisively improved performance results, elided due to lack of space. This is not surprising, as the function is both convex and smooth; any exploration is likely to be wasted effort. Similar behavior can be observed on Ackley and even Rosenbrock.

On Rastrigin the performance of PSO is not harmed by removal of momentum so long as diversity enhancements are in play: stagnation is prevented by CRIBS and exploration is balanced by SAC.



Christopher K. Monson (chris@highentropy.com) is with Google, Inc. Kevin D. Seppi (k@cs.byu.edu) is with the Department of Computer Science, Brigham Young University.