Under-Informed Momentum in PSO

Christopher K. Monson
Google, Inc.
Washington, D.C.
chris@highentropy.com

Kevin D. Seppi
BYU Computer Science Dept.
3361 TMCB
Provo, UT 84602-6576
kseppi@byu.edu

ABSTRACT

Particle Swarm Optimization is fundamentally a stochastic algorithm, where each particle takes into account noisy information from its own history as well as that of its neighborhood. Though basic information-theoretic principles would suggest that less noise indicates greater certainty, the momentum term is simultaneously the least directly-informed and the most deterministically applied. This dichotomy suggests that the typically confident treatment of momentum is misplaced, and that swarm performance can benefit from better-motivated processes that obviate momentum entirely.

1. INTRODUCTION

Particle Swarm Optimization (PSO) is a stochastic hill-climbing algorithm especially suitable for use in continuous domains. Several "particles" are initialized with random positions and velocities in a limited "feasible region" of the function's domain. At each time step, all particles sample the fitness function, observe the history of their neighbors, and unilaterally reposition themselves using a simple update equation. This is repeated for the duration of the optimization session. Each particle maintains its current velocity and position, as well as the best known position and fitness [6].

An effort has been made to codify a standard for PSO that can be used as a baseline comparison for ongoing research [2]. This standard, basic algorithm is an impressively performant general approach when implemented correctly and thus serves as an effective starting point. Standard PSO utilizes the symmetric "ring" topology with 20 particles and the following update equations:

$$\mathbf{v}_{t+1} = \chi\left(\mathbf{v}_t + \phi_p \mathbf{U} \circ (\mathbf{p}_t - \mathbf{x}_t) + \phi_g \mathbf{U} \circ (\mathbf{g}_t - \mathbf{x}_t)\right) \tag{1}$$

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{v}_{t+1} \ . \tag{2}$$

Here $\phi_p = \phi_g = 2.05$ are the "cognitive" and "social" coefficients [7], each $\mathbf{U}$ is a distinct vector whose elements are independently drawn from a standard uniform distribution for each use, and $\circ$ represents element-wise multiplication. The "constriction" coefficient χ is typically specified as $\chi \approx 0.72984$ [2]. The terms $\mathbf{p}$ and $\mathbf{g}$ are, respectively, the individual particle's best known position and the best known position among its neighbors.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
GECCO'14, July 12-16, 2014, Vancouver, BC, Canada
ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598490.

An important alternative formulation of (1) uses an inertia weight ω instead of the constriction factor χ [10]:

$$\mathbf{v}_{t+1} = \omega \mathbf{v}_t + \phi_p \mathbf{U} \circ (\mathbf{p}_t - \mathbf{x}_t) + \phi_g \mathbf{U} \circ (\mathbf{g}_t - \mathbf{x}_t) \ . \tag{3}$$

Formulations (1) and (3) are equivalent when $\omega = 0.72984$ and $\phi_p = \phi_g \approx 1.5$. Nevertheless, for our purposes the flexibility of (3) is more convenient. When using this formulation with different parameter settings, a per-dimension velocity cap $V_{\max}$ is typically employed to avoid velocity explosion.

In spite of the maturity of PSO, the most effective inertia setting continues to be an area of active and not entirely fruitful research. For a given problem it is often possible to find an effective value or strategy, but tuning this parameter is usually a problem-specific exercise and a matter of trial and error [1, 3, 4].

All of the terms in the update equations are both noisy *and informed* by fitness values, except momentum. The momentum term, while accompanied by noise in several variants of PSO, is at best only ever indirectly informed by fitness samples. The disproportionate impact of this term on swarm exploration would suggest that algorithmic benefits arise from a better-motivated formulation.

2. DISTRIBUTIONS AND DETERMINISM

The notation of (3) is convenient from an implementation point of view, but for our purposes a formulation in terms of random variables will be helpful. Consider the following completely equivalent formulation of (3):

$$\mathbf{P}_{t+1} \sim \mathbf{U}\left[\mathbf{0}, \phi_p(\mathbf{p}_t - \mathbf{x}_t)\right] \tag{4}$$

$$\mathbf{G}_{t+1} \sim \mathbf{U}\left[\mathbf{0}, \phi_g(\mathbf{g}_t - \mathbf{x}_t)\right] \tag{5}$$

$$\mathbf{v}_{t+1} = \omega \mathbf{v}_t + \mathbf{P}_{t+1} + \mathbf{G}_{t+1} \tag{6}$$

Here, $\mathbf{P}$ and $\mathbf{G}$ are each sampled from a multivariate uniform distribution over independent variables, specified by extreme corners of a hyperrectangle. These are then used to produce a new velocity. We can take this one step further, producing a single distribution $\mathbf{C}$ from the convolution of $\mathbf{P}$ and $\mathbf{G}$ (shaped like, e.g., Figure 1):

$$\mathbf{\Theta}_{t+1} \sim \mathbf{C}\left[\phi_p(\mathbf{p}_t - \mathbf{x}_t), \phi_g(\mathbf{g}_t - \mathbf{x}_t)\right] \tag{7}$$

$$\mathbf{v}_{t+1} = \omega \mathbf{v}_t + \mathbf{\Theta}_{t+1} \ . \tag{8}$$

Again, this is exactly equivalent to (3). Interestingly, it is somewhat reminiscent of the "Bare Bones" PSO algorithm, which uses the Normal distribution to position particles[1] [5].

The distribution $\mathbf{C}$ is noisily but directly informed by locations of historically good fitness. A closer look at (8) reveals that momentum, on the other hand, does not *inform* the final position dis-

[1] Fittingly, the Normal Distribution is approached in the limit by multiple similar convolutions.

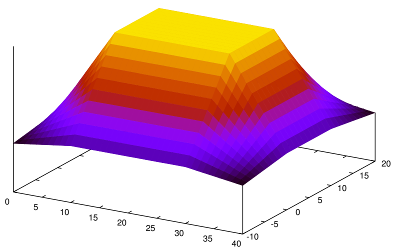

Figure 1: Convolution U $[(0,0),(30,20)] * \mathbf{U}\,[(0,0),(10,-10)]$**.**

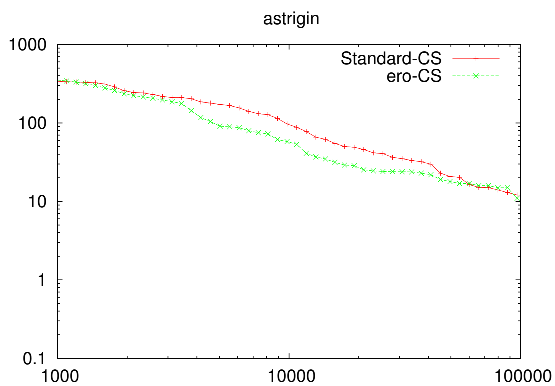

Figure 2: CRIBS+SAC, Standard and Zero Momentum.

tribution, but arbitrarily *moves it elsewhere*. It is simultaneously highly impactful and woefully under-informed.

3. DIVERSITY WITHOUT MOMENTUM

The rather arbitrary nature of momentum presents difficulties when attempts are made to find consistent inertia weight tuning strategies. Swarm exploration as controlled by ω is indirect and unpredictable, suggesting a more direct approach. Methods like CRIBS and its predecessors, for example, provide intuitive control over exploration/exploitation trade-offs and are effective at stagnation avoidance [9]. Methods like SAC, which employ fitness history feedback to adjust social and cognitive terms, also help with appropriately-applied diversity [8]. These can replace momentum.

Zero momentum is not new [5], but such results have not received deserved attention, likely because such results are not typically accompanied by other diversity enhancements in the literature. In the results of Figure 2, we demonstrate the effects of removing momentum while leaving SAC and CRIBS in place to improve diversity and reduce stagnation. As constriction is not needed without momentum, we restore $\phi_g = \phi_p = 2.05$ for Standard PSO.

Omitting momentum on the Sphere function (a result elided due to lack of space) decisively improves performance. This is not surprising, as the function is both convex and smooth. Any exploration is likely to be wasted effort. Similar behavior was observed on Ackley and even Rosenbrock. On Rastrigin the performance of PSO is not harmed by complete removal of momentum; stagnation is prevented by CRIBS and exploration is balanced by SAC.

The omission of momentum tends to reduce swarm diversity for harder functions in the long run, necessitating additional diversity injection, but additional diversity-increasing methods are necessary on harder functions *anyway*, and the idea that the responsibility for the remaining effects of momentum might be successfully turned over to more directly-informed strategies is very attractive. If we cannot find consistently motivated and effective ways to use momentum to control swarm diversity and exploration, it may well be possible to drop it entirely in favor of something else.

4. REMARKS AND CONCLUSION

Momentum is an arbitrary and difficult-to-tune approach to swarm exploration tradeoffs. Fortunately, when PSO is augmented with more straightforward diversity injection mechanisms, momentum is simply not needed. As these mechanisms are needed on difficult and highly multi-modal functions regardless of the presence of momentum, this does not impose any greater burden on developers or practitioners than was already present. The performance of momentumless PSO can be dramatically improved by further simple changes to existing PSO terms, to appear in future work.

References

[1] J. C. Bansal, P. K. Singh, M. Saraswat, A. Verma, S. S. Jadon, and A. Abraham. Inertia weight strategies in particle swarm optimization. In *Proceedings of the World Congress on Nature and Biologically Inspired Computing (NaBIC)*, pages 633–640, 2011.

[2] D. Bratton and J. Kennedy. Defining a standard for particle swarm optimization. In *Proceedings of the IEEE Swarm Intelligence Symposium (SIS 2007)*, pages 120–127, Honolulu, HI, 2007.

[3] M. A. M. de Oca, T. Stützle, M. Birattari, and M. Dorigo. Frankenstein's PSO: A composite particle swarm optimization algorithm. *IEEE Transactions on Evolutionary Computation*, 13(5):1120–1132, October 2009.

[4] R. C. Eberhart and Y. Shi. Tracking and optimizing dynamic systems with particle swarms. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2001)*, Seoul, Korea, 2001.

[5] J. Kennedy. Bare bones particle swarms. In *Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003)*, pages 80–87, Indianapolis, Indiana, 2003.

[6] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *International Conference on Neural Networks IV*, pages 1942–1948, Piscataway, NJ, 1995. IEEE Service Center.

[7] J. Kennedy and R. C. Eberhart. *Swarm Intelligence*. Morgan Kaufmann Publishers, 2001.

[8] C. K. Monson. Simple adaptive cognition for PSO. In *Proceedings of the Congress on Evolutionary Computation (CEC 2011)*, pages 1657–1664, New Orleans, Louisiana, 2011. IEEE.

[9] C. K. Monson and K. D. Seppi. Adaptive diversity in PSO. In *Proceedings of the 8th Annual conference on Genetic and Evolutionary Computation (GECCO 2006)*, pages 59–66, Seattle, Washington, 2006. ACM.

[10] Y. Shi and R. C. Eberhart. A modified particle swarm optimizer. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 1998)*, Piscataway, New Jersey, 1998.