Some Measurements on the Effects of the Curse of Dimensionality

Stephen Chen
School of Information Technology
York University
Toronto, Canada
1-416-736-2100 x:30526
sychen@yorku.ca

James Montgomery
School of Engineering and ICT
University of Tasmania
Hobart, Australia
61-3-6226-7294
james.montgomery@utas.edu.au

Antonio Bolufé-Röhler
School of Mathematics and CS
University of Havana
Havana, Cuba
53-5-242-0064
bolufe@matcom.uh.cu

ABSTRACT
The existence of the curse of dimensionality is well known, and its general effects are well acknowledged. However, perhaps due to this colloquial understanding, specific measurements on the curse of dimensionality and its effects are not as extensive. In continuous domains, the volume of the search space grows exponentially with dimensionality. Conversely, the number of function evaluations budgeted to explore this search space usually grows only linearly. New experiments show that particle swarm optimization and differential evolution have super-linear growth in convergence time as dimensionality grows. When restricted by a linear growth in allotted function evaluations, this super-linear growth in convergence time leads to a decrease in the allowed population size.

Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search – *Heuristic methods*

General Terms
Algorithms

Keywords
Curse of dimensionality, large scale global optimization.

1. INTRODUCTION
Our intuitive understanding of population-based heuristic search techniques is often based on examples formed in two dimensions. Of note, particle swarm optimization (PSO) starts with an image of birds circling in towards a food source in a corn field [1]. As we picture the 15-30 birds involved with this original conceptualization, we can imagine how the proposed search technique will achieve excellent coverage of the search space. However, these intuitive ideas can be inaccurate and even misleading in higher dimensional search spaces.

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/10.1145/2598394.2602271

A simple guideline from differential evolution (DE) is to use a population size (n) that is ten times the dimensionality (d) of the search space [2]. However, going from d = 2 to d = 42 can increase the size of the search space by a factor of about 10^{12}. Whereas we might imagine 20 birds being able to fully explore a 10m x 10m courtyard, we should have a different visualization of 420 birds in an area similar to the size of the Atlantic Ocean.

Despite this increasing sparseness in the coverage of the search space, popular guidelines for the recommended size of the population in PSO and DE include constant sized (e.g. [3] for PSO) and linearly increasing sizes (e.g. [2][4] for DE). The following experiments show that even these population sizes can become too large to allow convergence when PSO and DE are applied to large scale global optimization (i.e. $d \geq 1000$).

2. PARTICLE SWARM OPTIMIZATION
The experiments measure what it takes for PSO and DE to converge "close" to the optimum on sphere. The sphere function is chosen because it is the simplest (unimodal) baseline function used in many benchmark sets. If a search technique is unable to reach the optimum on sphere, the error on sphere suggests a minimum error that will exist for any (local) optimum on a deceptive (multi-modal) problem.

The current implementation of sphere uses the range of +/− 100 and the limit of 10,000 * d function evaluations. Starting with uniform random solutions on this range, we record the number of function evaluations required to first produce a solution within the +/− 0.1 hypercube around the origin (where the un-shifted global optimum is located). Thirty independent trials are run for each data point. Since variations amongst the trials were negligible (with the excess of trials mostly to confirm the key trends), only averages of the converged trials are presented.

The first set of experiments is a 10 by 10 grid – dimensions d from 50 to 500 in steps of 50 and population sizes n from 50 to 500 also in steps of 50. The average number of generations (i.e. function evaluations divided by population size) to first find a solution within the target hypercube is reported in Table I. In general, the number of generations required for convergence is mostly constant with respect to population size with a slight benefit (i.e. fewer generations required) for larger values of n. However, the approximate growth rate in generations required to converge as a function of dimensions d is faster than a linear function with a constant of 1. Since the allowed function evaluations increases linearly (i.e. function evaluations = 10,000 * d), the growth rate in generations eventually eclipses the growth rate in function evaluations. PSO with n = 450 cannot converge for $d \geq 450$ dimensions, and n = 500 fails for $d \geq 250$ dimensions.

Table 1. Number of generations required by PSO to converge on the sphere function.

n	Dimensions (*d*)									
	50	100	150	200	250	300	350	400	450	500
50	959.4	1982.9	3099.6	4251.3	5465.6	6816.1	8121.4	9346.4	10807.0	12361.0
100	935.9	1961.3	3047.1	4212.2	5416.2	6653.1	7905.1	9234.7	10579.8	11899.0
150	924.4	1963.1	3053.3	4174.4	5371.7	6611.6	7879.1	9180.5	10480.0	11870.7
200	920.1	1942.8	3021.8	4177.8	5348.0	6562.4	7830.4	9107.8	10436.3	11738.2
250	920.6	1944.1	3015.7	4148.9	5335.3	6559.6	7814.1	9078.4	10416.2	11776.3
300	918.2	1932.7	3016.9	4134.9	5299.7	6563.9	7763.8	9059.8	10439.3	11739.0
350	914.2	1938.3	3008.6	4129.6	5296.7	6557.4	7793.1	9045.3	10392.5	11692.9
400	900.5	1918.4	3003.2	4118.1	5307.1	6548.5	7805.5	9069.5	10312.4	11685.8
450	909.4	1919.1	3016.5	4121.0	5284.3	6504.5	7710.9	8821.3	-	-
500	913.3	1929.3	2971.6	3974.0	-	-	-	-	-	-

The observation that the resources required for convergence grow super-linearly with dimensions is more easily seen in Figure 1. The blue diamonds indicate the actual number of generations required to first produce a solution within the target hypercube, and the red crosses indicate a linear extrapolation of the first point for $d = 50$ and the origin.

3. DIFFERENTIAL EVOLUTION

A similar super-linear growth rate in the number of generations required for convergence is also observed in DE. In Figure 2, a standard DE/rand/1/bin is tested on sphere using a target hypercube of +/− 1, $n = 50$, and $F \in \{0.5, 0.6, 0.7, 0.8\}$. The shaded area represents generation counts above the maximum imposed by the FE limit. These results also indicate that lowering F can lead to improved search efficacy – the curves from top to bottom have $F = 0.8$, 0.7, 0.6, and lastly 0.5. However, while the trend in the plot suggests that continuing to reduce F would produce further improvements, experiments with $F = 0.4$ resulted in premature convergence for all $d > 50$. The balance between allowing the population to contract (so that it can make effective exploratory moves) while maintaining sufficient diversity is evidently critical to DE's success, but can be difficult to achieve.

4. SUMMARY

Since the budget of function evaluations usually grows only linearly with dimension d, it becomes clear that for any population size n, there will eventually be a problem size d for which this linear budget of function evaluations (e.g. 10,000 * d) will become insufficient to achieve a desired level of convergence. For large scale global optimization, the trends in Figures 1 and 2 suggest that populations sizes will have to become much smaller than the problem dimensionality (i.e. $n \ll d$) to ensure convergence in a reasonable amount of time. This limitation may affect the performance of PSO and DE in high dimensional search spaces.

5. REFERENCES

[1] Kennedy, J. and Eberhart, R. C. 1995. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks* (Perth, Australia, Nov 27-Dec 01, 1995). ICNN'95. IEEE Press, Piscataway, NJ, 1942-1948. DOI= http://dx.doi.org/10.1109/ICNN.1995.488968.

[2] Storn, R. 1996. On the usage of differential evolution for function optimization. In *Proceedings of the 1996 Biennial Conference of the North American Fuzzy Information Processing Society* (Berkeley, USA, June 19-22, 1996). NAFIPS'96. IEEE Press, Piscataway, NJ, 519-523. DOI= http://dx.doi.org/10.1109/NAFIPS.1996.534789.

[3] Bratton, D. and Kennedy, J. 2007. Defining a standard for particle swarm optimization. In *Proceedings of the Swarm Intelligence Symposium* (Honolulu, USA, April 01-05, 2007). SIS'07. IEEE Press, Piscataway, NJ, 120-127. DOI= http://dx.doi.org/10.1109/SIS.2007.368035.

[4] Mallipeddi, R. and Suganthan, P. N. 2008. Empirical study on the effect of population size on differential evolution algorithm. In *Proceedings of the IEEE Congress on Evolutionary Computation* (Hong Kong, June 01-06, 2008). CEC'08. IEEE Press, Piscataway, NJ, 3663-3670. DOI= http://dx.doi.org/10.1109/CEC.2008.4631294.

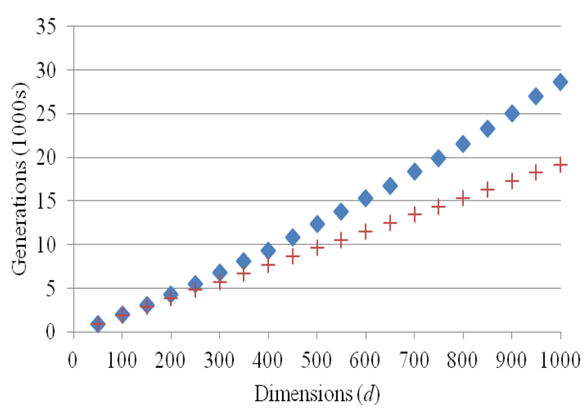

Figure 1. Convergence time grows super linearly in PSO.

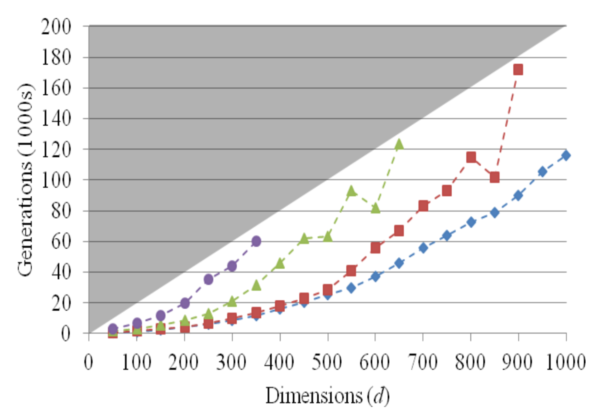

Figure 2. Convergence time grows super linearly in DE.