Solving optimization problems with high conditioning by means of online whitening

Bagheri, S.; Konen, W.; Bäck, T.; López-Ibánez, M.

Solving Optimization Problems with High Conditioning by Means of Online Whitening

Samineh Bagheri
TH Köln – Univeristy of Applied Sciences
Gummersbach, Germany
samineh.bagheri@th-koeln.de

Wolfgang Konen
TH Köln – Univeristy of Applied Sciences
Gummersbach, Germany
wolfgang.konen@th-koeln.de

Thomas Bäck
Leiden University, LIACS
Leiden, The Netherlands
t.h.w.baeck@liacs.leidenuniv.nl

ABSTRACT

Real-world optimization problems often have expensive objective functions in terms of cost and time. It is desirable to find near-optimal solutions with very few function evaluations. Surrogate-assisted optimizers tend to reduce the required number of function evaluations by replacing the real function with an efficient mathematical model built on few evaluated points. Problems with a high condition number are a challenge for many surrogate-assisted optimizers including SACOBRA. To address such problems we propose a new online whitening operating in the black-box optimization paradigm. We show on a set of high-conditioning functions that online whitening tackles SACOBRA's early stagnation issue and reduces the optimization error by a factor between 10 to 10^{12} as compared to the plain SACOBRA, though it imposes many extra function evaluations. Covariance matrix adaptation evolution strategy (CMA-ES) has for very high numbers of function evaluations even lower errors, whereas SACOBRA performs better in the expensive setting ($\leq 10^3$ function evaluations). If we count all parallelizable function evaluations (population evaluation in CMA-ES, online whitening in our approach) as one iteration, then both algorithms have comparable strength even on the long run. This holds for problems with dimension $D \leq 20$.

CCS CONCEPTS

• **Computing methodologies** → *Model development and analysis*;
• **Mathematics of computing** → *Continuous mathematics*;

KEYWORDS

Surrogate models, high condition number, online whitening

ACM Reference Format:
Samineh Bagheri, Wolfgang Konen, and Thomas Bäck. 2019. Solving Optimization Problems with High Conditioning by Means of Online Whitening. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3322008

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3322008

1 INTRODUCTION

Optimization problems can often be defined as minimization of a black-box objective function $f(\vec{x})$. An optimization problem is called black-box if no analytical information about itself or its derivatives are given.

Many nature-inspired derivative-free optimizers are weak in optimizing functions with high conditioning, despite all their significant contributions in addressing large sets of black-box problems [6, 7]. However, CMA-ES is very successful in tackling high-conditioning problems. The advantage of CMA-ES when solving problems with high conditioning stems from the fact that in each iteration the covariance matrix of the new distribution is adapted according to the evolution path which is the direction with highest expected progress. In other words, the covariance matrix adaptation aims to learn the Hessian matrix of the function in an iterative way.

Evolutionary-based algorithms including CMA-ES often require too many function evaluations which are not affordable in practice. One common approach to address expensive optimization problems efficiently is to employ surrogate models as a replacement for the expensive-to-evaluate functions in order to reduce the number of functions evaluations as much as possible [2, 3, 5].

The SACOBRA [2] optimization framework uses RBF interpolation as surrogate. Although this algorithm is very successful in handling the commonly used constrained optimization problems, the so-called G-function benchmark [2], it performs poorly when optimizing functions with a large condition number[1]. This poor performance is mainly because RBFs fail to provide a useful model for functions with high conditioning. This modeling issue is not exclusive to RBFs. A surrogate-assisted CMA-ES [5] using different modeling approaches addresses this modeling challenge by means of an orthogonal transformation of the evaluated points.

In this work we propose an online whitening approach for the SACOBRA framework to transform objective functions with high conditioning to improve the modeling phase. The proposed method is evaluated on the noiseless single-objective BBOB benchmark [4] and compared to the results of CMA-ES and differential evolution (DE) implementations in R. The detailed description of the algorithm, experimental setup and results can be found in [1].

2 ONLINE WHITENING (OW)

The online whitening scheme is described in Algorithm 1: We seek to transform the objective function $f(\vec{x})$ with high conditioning to another function $g(\vec{x})$ which is easier to model by surrogates:

[1] A function, that has a high ratio of steepest slope in one direction to flattest slope in another direction, has a large condition number. We call this a function with *high conditioning*.

Algorithm 1 Online whitening algorithm. Input: Function f to minimize, population $\mathbf{X} = \left\{\vec{x}_{(k)} | k = 1, \ldots, n\right\}$ of evaluated points, $\vec{x}_{best}$: best-so-far point from SACOBRA.

$\mathbf{H} \leftarrow$ Hessian matrix of function $f(\vec{x})$ at $\vec{x}_{best}$
$\mathbf{M} \leftarrow \mathbf{H}^{-0.5}$ {see Eq. (3)}
Update $\vec{x}_{best}$ with the function evaluations from Hessian calculation
Transformation :
$g(\vec{x}) \leftarrow f(\mathbf{M}(\vec{x} - \vec{x}_{best}))$
$\mathbf{G} \leftarrow \left\{\left(\vec{x}_{(k)}, g(\vec{x}_{(k)})\right) | k = 1, \ldots, n\right\}$ {evaluate all the points in $\mathbf{X}$ on the new function $g(\vec{x})$}
$s(\vec{x}) \leftarrow$ build surrogate model from $\mathbf{G}$
return $s(\vec{x})$ {surrogate model for next SACOBRA step}

$$g(\vec{x}) = f(\mathbf{M}(\vec{x} - \vec{x}_c)), \tag{1}$$

where $\mathbf{M}$ is a linear transformation matrix and $\vec{x}_c$ is the transformation center. The ideal transformation center is the optimum point which is clearly not available. As a substitute, we use the best so-far solution $\vec{x}_{best}$ in each iteration as the transformation center. The transformation matrix $\mathbf{M}$ is chosen in such a way that the Hessian matrix of the new function becomes the identity matrix:

$$\frac{\partial^2 g(\vec{x})}{\partial \vec{x}^2} = \mathbf{I} \tag{2}$$

As shown in [1], a solution for Eqs. (1) and (2) is given by:

$$\mathbf{M} = \mathbf{H}^{-0.5} \tag{3}$$

where $\mathbf{H}$ denotes the Hessian matrix of the objective function f. The transformation matrix $\mathbf{M}$ in Eq. (1) is similar to the so-called Mahalanobis whitening or sphering transformation, commonly used in statistical analysis. The Hessian matrix is determined by means of Richardson's extrapolation which requires $4D + 4D^2$ function evaluations. We call the online whitening scheme only every 10 iterations, since the initial experiments have shown that a more frequent update is not necessary.

3 RESULTS

As shown in Fig. 1, SACOBRA+OW can solve about two times more problems comparing to the plain SACOBRA, but this is only achieved after a significant increase in the number of function evaluations required by OW scheme. CMA-ES outperforms SACOBRA+OW and DE for higher number of function evaluations. However, as shown in Fig. 2, SACOBRA+OW can compete with CMA-ES in an optimistic parallelizable case, where enough computational resources allow parallel computation of the Hessian matrix during OW steps as well as parallel evaluation of new population of solutions in each generation of CMA-ES and DE.

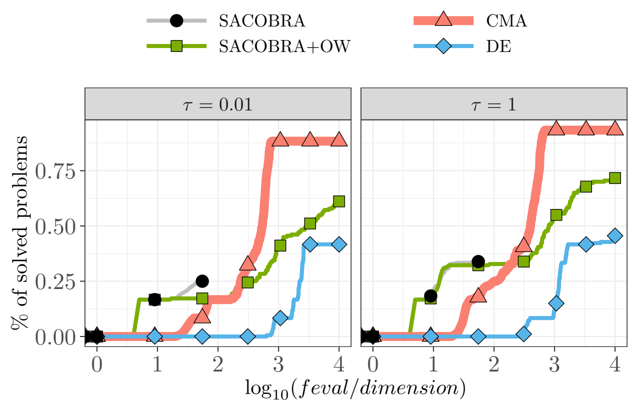

Figure 1: Comparing the overall performances of the algorithms SACOBRA, SACOBRA+OW, DE and CMA-ES, on 12 BBOB problems with $D = 10$. The x-axis has the number of function evaluations, divided by D. The results are shown for two different optimization error tolerances $\tau = \{0.01, 1\}$.

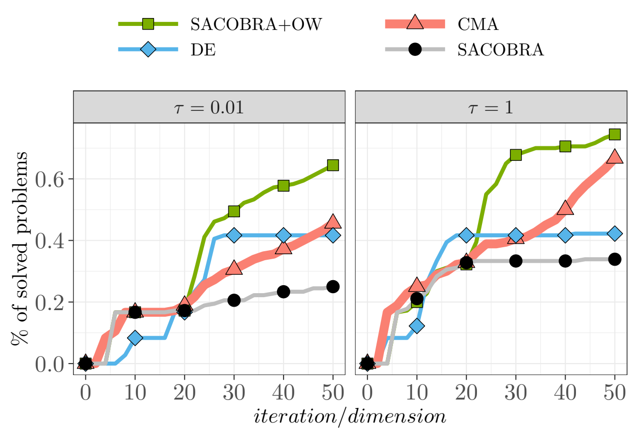

Figure 2: Same as Fig. 1, but now for the 'optimistic parallelizable' case: The x-axis shows the number of iterations (or generations), divided by D.

REFERENCES

[1] Samineh Bagheri, Wolfgang Konen, and Thomas Bäck. 2019. *SACOBRA with Online Whitening for Solving Optimization Problems with High Conditioning.* Technical Report. TH Köln – Univeristy of Applied Sciences. http://www.gm.fh-koeln.de/ciopwebpub/Bagh19a.d
[2] Samineh Bagheri, Wolfgang Konen, Michael Emmerich, and Thomas Bäck. 2017. Self-adjusting parameter control for surrogate-assisted constrained optimization under limited budgets. *Applied Soft Computing* 61 (2017), 377 – 393.
[3] Lukáš Bajer, Zbyněk Pitra, and Martin Holeňa. 2015. Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed. In *Proc. Genetic and Evolutionary Computation Conference GECCO'15*. ACM, New York, 1143–1150. https://doi.org/10.1145/2739482.2768468
[4] Steffen Finck, Nikolaus Hansen, Raymond Ros, and Anne Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Presentation of the Noiseless Functions.* Technical Report 2009/20. Research Center PPE.
[5] Ilya Loshchilov, Marc Schoenauer, and Michèle Sebag. 2012. Self-Adaptive Surrogate-Assisted Covariance Matrix Adaptation Evolution Strategy. *CoRR* abs/1204.2356 (2012). arXiv:1204.2356
[6] Babatunde A Sawyerr, Aderemi O Adewumi, and M Montaz Ali. 2015. Benchmarking rcgau on the noiseless bbob testbed. *The Scientific World Journal* 2015 (2015).
[7] Andrew M Sutton, Monte Lunacek, and L Darrell Whitley. 2007. Differential evolution and non-separability: using selective pressure to focus search. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation.* ACM, 1428–1435.