Towards Landscape Analysis in Adaptive Learning of Surrogate Models

Zbyněk Pitra[1,2] and Martin Holeňa[1]

[1] Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
{pitra, martin}@cs.cas.cz
[2] Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague
Břehová 7, 115 19 Prague 1, Czech Republic

Keywords: Adaptive learning · Optimization strategy · Black-box optimization · Landscape analysis · Surrogate model

1 Introduction

A context in which we expect adaptive learning to be promising is the choice of a suitable optimization strategy in black-box optimization. The reason why strategy adaptation is needed in such a situation is that knowledge of the black-box objective function is obtained only gradually during the optimization. That knowledge covers two aspects:

1. the landscape of the black-box objective, revealed through its evaluation in previous iterations;
2. success or failure of the optimization strategies applied to that black-box objective in previous iterations.

To extract landscape knowledge, landscape analysis has been developed during the last decade [7,10,11]. To include also the second aspect, we complement features obtained using the landscape analysis with features describing the optimization employed in previous iterations.

Our interest is in expensive black-box optimization, where the number of evaluations of the expensive objective is usually decreased using a suitable surrogate model. Therefore, the research reported in this extended abstract addresses adaptive learning of surrogate models, more precisely their learning in surrogate-assisted versions of the state-of-the-art black-box optimization method, Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [3].

Considering the results in [2,14] suggesting that the properties of landscape features in connection with surrogate model selection problem should be analysed in more detail, we contribute with this work a first essential step towards a better understanding, by analysing the robustness of feature computation. Such analysis of a large set of landscape features has already been presented only in connection with selection of the most convenient optimization algorithm for problems in fixed dimension [15].

This extended abstract focuses on surrogate model selection task in multiple dimensions and discusses robustness of several classes of features against samples of points from the same distribution.

2 Landscape Analysis for Surrogate Model Selection

Landscape analysis aims at measuring characteristics of the objective function using functions that assign to each dataset a set of real numbers [10]. Let's consider a dataset of N pairs of observations $\left\{(\mathbf{x}_i, y_i) \in \mathbb{R}^D \times \mathbb{R} \cup \{\circ\} \,|\, i = 1, \ldots, N\right\}$, where $\circ$ denotes missing y_i value (e. g., $\mathbf{x}_i$ was not evaluated yet). Then the dataset can be utilized to describe landscape properties using a feature φ : $\bigcup_{N\in\mathbb{N}} \mathbb{R}^{N,D} \times \left(\mathbb{R} \cup \{\circ\}\right)^{N,1} \mapsto \mathbb{R} \cup \{\pm\infty, \bullet\}$, where $\bullet$ denotes impossibility of feature computation.

Feature classes convenient for continuous black-box optimization field are mostly described in [7]. From the available feature classes we mention only those convenient for problems with a high computational complexity (unlike e. g., cell-mapping approach [8]) and at the same time not requiring additional evaluations of the expensive function. Feature classes are able to measure the dissimilarity among points of a subset of the sample (*Dispersion*) [9], express various information content of the landscape (*Information Content*) [11], measure the relative position of each value with respect to quantiles (*Levelset*) [10], extract the information from linear or quadratic regression models (*Meta-Model*) [10] or from the nearest or the better observation neighbours (*Nearest Better Clustering*) [6], and describe the distribution of the objective values (*y-Distribution*) [10]. Moreover, in [13] we have proposed the set of features based on the CMA-ES state variables (*CMA features*).

The surrogate model selection problem tackle the situation in an iteration i of a surrogate-assisted algorithm A, where a set of surrogate models $\mathcal{M}$ are trained using a training set $\mathcal{T}$ selected out of an *archive* $\mathcal{A}$ ($\mathcal{T} \subset \mathcal{A}$) of all points evaluated so far using the objective function f: $\mathcal{A} = \{(\mathbf{x}_i, f(\mathbf{x}_i)) | \, i = 1, \ldots, N\}$. Hereafter, a new set of points $\mathcal{P} = \{\mathbf{x}_k | k = 1, \ldots, \alpha\}$ is evaluated using a surrogate model $M \in \mathcal{M}$, where $\alpha \in \mathbb{N}$ depends on the strategy defining the usage of surrogate model in algorithm A. The research question is: How to select the most convenient M from $\mathcal{M}$ according to $\mathcal{A}$, $\mathcal{T}$, and $\mathcal{P}$?

To tackle the research question connected with the surrogate model selection problem, we have proposed (see [14]) the following metalearning approach visualised in Figure 1:

In the first phase, each model $M \in \mathcal{M}$ is trained on each $\mathcal{T}^{(l)}$ from the set of datasets $\mathcal{D} = \{\mathcal{A}^{(l)}, \mathcal{T}^{(l)}, \mathcal{P}^{(l)}\}_{l=1}^{L}$, $L \in \mathbb{N}$ and its error ε is measured on $\mathcal{P}^{(l)}$. Simultaneously, a set of features Φ is computed on each dataset from $\mathcal{D}$. Hereby, a mapping $S_M : \Phi \to \mathcal{M}$ from the space of landscape features to $\mathcal{M}$ is trained. In the second phase, the trained mapping S_M is utilized in each iteration i of the algorithm A to select the model $M \in \mathcal{M}$ according to the features Φ calculated on $\mathcal{A}^{(i)}$, $\mathcal{T}^{(i)}$, and $\mathcal{P}^{(i)}$. The selected M is utilized to fit $\mathcal{T}^{(i)}$ and afterwards to predict objective function values of points from $\mathcal{P}^{(i)}$.

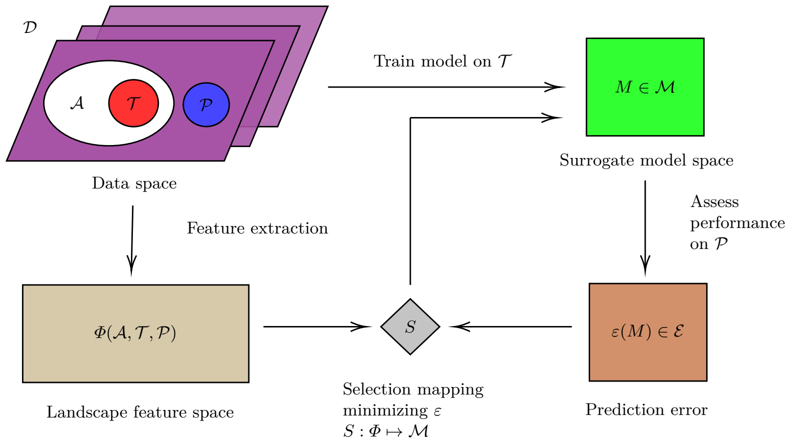

Figure 1: Scheme of the metalearning approach to the surrogate model selection system [14].

3 Feature Robustness

To investigate robustness of feature computation against different samples of points (in the sense of low variance), several independent archive realisations using the same distributions should be available. To gain such realisations, we have created a new set of artificial distributions by smoothing the distributions from real runs of the surrogate algorithm on the set of benchmarks.

First, we have generated a set of datasets $\mathcal{D}$ using independent runs of the 8 model settings from [13] for the DTS-CMA-ES algorithm [1,12] on the 24 noiseless single-objective benchmark functions from the COCO framework [4,5]. All runs were performed in dimensions 2, 3, 5, 10, and 20 on instances 11–15. To gain 100 comparable archives using those runs, we have generated points for new archives using the weighted sum of original archive distributions from $\mathcal{D}$, where the weight vector $\mathbf{w}^{(i)} = \frac{1}{9}(0, \ldots, \underset{i-3}{0}, \underset{i-2}{1}, \underset{i-1}{2}, \underset{i}{3}, \underset{i+1}{2}, \underset{i+2}{1}, \underset{i+3}{0}, \ldots, 0)^\top$ provides distribution smoothing across the available iterations[1]. Second, for all $\mathcal{A}^{(i)}$, $\mathcal{T}^{(i)}$, and $\mathcal{P}^{(i)}$ from $\mathcal{D}$ we have computed all features from the following feature classes: *Dispersion*, *Information Content*, *Levelset*, *Meta-Model*, *Nearest Better Clustering*, *y-Distribution*, *CMA features*.

Once the features are computed, the numbers of $\pm\infty$ and $\bullet$ values of different samples from one iteration are summarized and the rest of feature values is normalized to $[0, 1]$ range using feature minima and maxima over the whole $\mathcal{D}$. We then compare feature means and variances for individual iterations.

[1] Weighted sum of the original archive distributions satisfies $\sum_{n=0}^{i_{\max}} w_n^{(i)} \mathcal{N}\big(\mathbf{m}^{(n)}, \mathbf{C}^{(n)}\big) \sim \mathcal{N}\big(\sum_{n=0}^{i_{\max}} w_n^{(i)} \mathbf{m}^{(n)}, \sum_{n=0}^{i_{\max}} (w_n^{(i)})^2 \mathbf{C}^{(n)}\big)$, where $i_{\max}$ is the maximal iteration reached by particular original archive and $\mathbf{m}^{(n)}$ and $\mathbf{C}^{(n)}$ are mean and covariance matrix in iteration n.

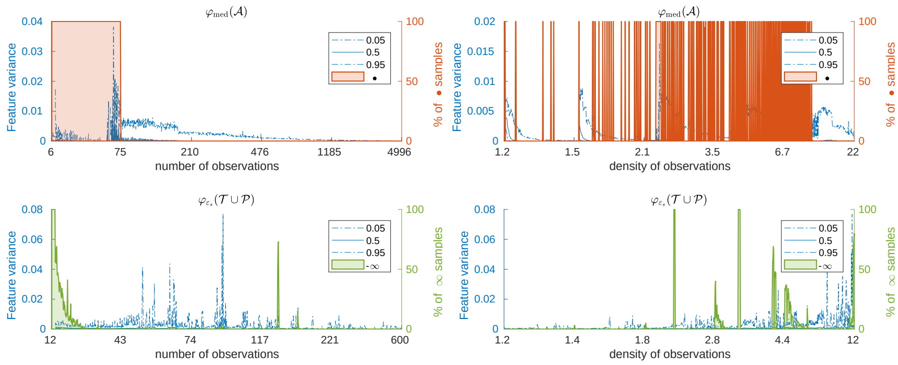

Figure 2: The dependecies of 0.05, 0.5, and 0.95 quantile of feature variance, the median number of $\pm\infty$, or $\bullet$ of feature values on the number of observations N for two features are shown on plots in the first column. The dependencies of the same statistics on the data density $\sqrt[D]{N}$ are presented in the second column. Plots in the first row represent statistics for feature $\varphi_{\text{med}}(\mathcal{A})$ – median distance of the 'best' vs. 'all' objectives in $\mathcal{A}$ (from *Dispersion* feature class) and the second row contains statistics for $\varphi_{\varepsilon_s}(\mathcal{T} \cup \mathcal{P})$ – settling sensitivity of the information content in $\mathcal{T} \cup \mathcal{P}$ (*Information Content*).

Figure 2 shows the dependecies of 0.05, 0.5, and 0.95 quantile of feature variance, the number of $\pm\infty$, or $\bullet$ on the number of observations N in the considered set ($\mathcal{A}$, $\mathcal{T}$, or $\mathcal{P}$) and data density $\sqrt[D]{N}$ for two example features.

The results show that most of the features are robust in the sense of having a low variance, especially for higher numbers of observations. Robustness for lower values of N is not frequently high, or even the feature is not possible to calculate (e. g., some of *Dispersion* features). *CMA* features provided the most robust results probably due to the fact that most of them are sample independent. The lowest variance values, and also high numbers of cases where the feature was impossible to calculate were observed at *Dispersion* features.

4 Conclusion

The extended abstract addressed adaptive learning of a suitable optimization setting in black-box optimization, more precisely, adaptive learning of a surrogate model in a surrogate-assisted version of the CMA-ES. Its main message is the relationship of this kind of adaptive learning to landscape analysis. A formal framework for the learning of a surrogate model based on landscape analysis is given, and considered kinds of landscape features are discussed. In the results obtained so far, attention is paid in particular to feature robustness.

This work in progress is part of a thorough investigation of the possibilities of landscape analysis in the context of surrogate modelling for black-box optimization. That investigation has already brought first results in the past [2,13,14], but much still remains for further research.

Acknowledgements The reported research was supported by the Czech Science Foundation grant No. 18-18080S and by the Grant Agency of the Czech Technical University in Prague with its grant No. SGS20/183/OHK4/3T/14. Further, access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

References

1. Bajer, L., Pitra, Z., Repický, J., Holeňa, M.: Gaussian process surrogate models for the CMA Evolution Strategy. Evolutionary Computation **27**(4), 665–697 (2019)
2. Dvořák, M.: Optimization of surrogate model settings using landscape analysis. Master's thesis, Czech Technical University in Prague, Faculty of Information Technology, Prague, Czech Republic (2020)
3. Hansen, N.: The CMA evolution strategy: A comparing review. In: Towards a New Evolutionary Computation, pp. 75–102. No. 192 in Studies in Fuzziness and Soft Computing, Springer Berlin Heidelberg (Jan 2006)
4. Hansen, N., Auger, A., Finck, S., Ros, R.: Real-parameter black-box optimization benchmarking 2012: Experimental setup. Tech. rep., INRIA (2012)
5. Hansen, N., Finck, S., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Tech. Rep. RR-6829, INRIA (2009), updated February 2010
6. Kerschke, P., Preuss, M., Wessing, S., Trautmann, H.: Detecting funnel structures by means of exploratory landscape analysis. pp. 265–272. GECCO '15, ACM (2015)
7. Kerschke, P., Hoos, H.H., Neumann, F., Trautmann, H.: Automated algorithm selection: Survey and perspectives. Evolutionary computation **27**(1), 3–45 (2019)
8. Kerschke, P., Preuss, M., Hernández, C., Schütze, O., Sun, J.Q., Grimme, C., Rudolph, G., Bischl, B., Trautmann, H.: Cell mapping techniques for exploratory landscape analysis. In: EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V. pp. 115–131. Springer International Publishing (2014)
9. Lunacek, M., Whitley, D.: The dispersion metric and the cma evolution strategy. pp. 477–484. GECCO '06, ACM (2006)
10. Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., Rudolph, G.: Exploratory landscape analysis. pp. 829–836. GECCO '11, ACM (2011)
11. Muñoz, M.A., Kirley, M., Halgamuge, S.K.: Exploratory landscape analysis of continuous space optimization problems using information content. IEEE Transactions on Evolutionary Computation **19**(1), 74–87 (2015)
12. Pitra, Z., Bajer, L., Holeňa, M.: Doubly trained evolution control for the Surrogate CMA-ES. In: Proceedings of the PPSN XIV: 14th International Conference, Edinburgh, UK, September 17-21. pp. 59–68. Springer International Publishing, Cham (2016)

13. Pitra, Z., Repický, J., Holeňa, M.: Landscape analysis of Gaussian process surrogates for the covariance matrix adaptation evolution strategy. pp. 691–699. GECCO '19, ACM (2019)
14. Pitra, Z., Bajer, L., Holeňa, M.: Knowledge-based selection of Gaussian process surrogates. In: Kottke, D., Lemaire, V., Calma, A., Krempl, G., Holzinger, A. (eds.) ECML PKDD 2019: Workshop & Tutorial on Interactive Adaptive Learning. Proceedings. pp. 48–63. ECML PKDD 2019, Würzburg, Germany (Sep 2019)
15. Renau, Q., Dreo, J., Doerr, C., Doerr, B.: Expressiveness and robustness of landscape features. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 2048–2051. GECCO '19, Association for Computing Machinery, New York, NY, USA (2019)