Effects of function translation and dimensionality reduction on landscape analysis

Mario A. Muñoz and Kate Smith-Miles
School of Mathematical Sciences, Monash University, Clayton, VIC Australia
e-mail: {mario.munoz,kate.smith-miles}@monash.edu

***Abstract*—Exploratory Landscape Analysis (ELA) measures have been shown to predict algorithm performance; hence, they are being applied on critical tasks such as automatic algorithm selection and problem generation. This paper provides a cautionary examination on their use in black-box continuous optimization. We explore the effect that translations have on the measures, when the cost function is defined within a bound-constrained region. Furthermore, we examine the robustness of the neighborhood structure after dimensionality reduction. The results demonstrate that a measure may transition abruptly due a translation. Therefore, we should not generalize the measures of an instance nor report average values of a measure as belonging to the generating function. Moreover, dimensionality reduction could alter the neighborhood structure, such that the regions corresponding to significantly different functions overlap.**

***Index Terms*—Black-box continuous optimization, Exploratory landscape analysis, Fitness landscape analysis, Stochastic optimization**

I. Introduction

The ubiquitous nature of black-box continuous optimization problems in science, engineering, and other fields, as well as greater access to computing infrastructure, has led to intense research on stochastic search algorithms over the past decades. As a result, we now have significant algorithmic diversity, which unexpectedly has compromised our ability to master —or even be familiar with— all reported algorithms [1]. Therefore, significant innovations in the field are being obfuscated [2] and ideas are being recycled [3]–[6]. Furthermore, our theoretical understanding of the strengths and weaknesses of most algorithms on real-world problems is still limited, even after significant advances [7]. Therefore, selecting an appropriate algorithm for a given problem is at best cumbersome [8], even with expert knowledge of search algorithms, and skills in algorithm engineering and statistics [9].

The extensive and successful work on the algorithm selection problem in machine learning and combinatorial optimization [10]–[13] has reignited the interest in Exploratory Landscape Analysis (ELA) measures as quantifiers of problem characteristics, such as modality, ruggedness, and variable dependencies, of black-box continuous optimization problems [14], [15]. Evidence suggest that ELA measures are effective predictors of algorithm performance [16]–[18]; however, they do have limitations. For example, there is no single, all encompassing, measure of problem complexity, as it is the interplay between characteristics that defines difficulty [19]–[21]. Furthermore, ELA methods are sampling based. Hence, they require a dataset that grows exponentially with the problem dimension to converge [21]–[23], which cannot be guaranteed in polynomial time [24]. Therefore, we must rely on approximated measures in practice. Arguably, we could save computational time by measuring the characteristics during an algorithm run [25], if the sampling bias created by the algorithm is accounted for [26]. Nevertheless, our practical and theoretical understanding of ELA measures and their limitations remains modest.

A little explored topic is the effect that small incremental changes on the function structure, such as translations, have on a measure when the function is defined within a bound-constrained region. The assumption being that robust measures should capture the fundamental characteristics of a function, even if they may slightly vary across instances. Therefore, we may feel inclined to average a measure across all the instances generated from the same function, or to generalize the measure resulting from a single instance. On the contrary, we suspect that such robustness level is unachievable and these effects could be substantial. Furthermore, due to the high dimensional nature of the measure space, we increasingly rely on dimensionality reduction techniques to visualize the relationships between functions. Since the resulting low dimensional space is a compromise, we will obtain altered neighborhood structures that could hide important function characteristics. These issues should be acknowledged if we must rely on ELA measures for critical decision making during algorithm selection, problem analysis and generation.

This paper aims to answer two questions: How does the translation of a function affect the landscape measures? How is the instance neighborhood affected by a two dimensional projection? To find an answer, we calculated nine measures for a set of reference functions —the noiseless benchmarks set from the Comparing Continuous Optimizers (COCO) package [27]— and test functions —the Sphere, Rastrigin and Bent cigar functions translated in the input and output spaces. To understand the measures' transitions on the resulting nine dimensional space, we calculated the distance between the problems and examined their relationships. We then generated a two dimensional representation of the measure space, which allowed us to visualize the path created by the test

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

problems and their location within the space. The results demonstrate that a measure may transition abruptly due a translation. Therefore, we should not generalize the measures of an instance nor report average values of a measure as belonging to the generating function. We should also avoid isolated interpretations of a measure, as this is likely to lead to incorrect conclusions. Furthermore, dimensionality reduction could alter the neighborhood structure, such that the regions corresponding to significantly different functions may overlap.

The remainder of this paper is as follows: in Section II we describe the details of our experimental methodology. The results and their main implications are presented in Sections III and IV respectively.

II. Experimental methodology

A. Black-box continuous optimization instances

We start by defining our notation. Without loosing generality over maximization, the goal in a black-box continuous optimization problem is to minimize the function $f : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X} \subset \mathbb{R}^D$ is the compact input space, $\mathcal{Y} \subset \mathbb{R}$ is the output space, and $D \in \mathbb{N}^+$ is the dimensionality of the problem. A candidate solution $\mathbf{x} \in \mathcal{X}$ is a D-dimensional vector, and $y \in \mathcal{Y}$ is the candidate's cost.

We used the noiseless benchmark set from the COCO package [27] as a representative subset of the problem space and reference set. Instances are generated by scaling and transforming 24 basis functions, with $\mathcal{X} = [-5, 5]^D$. The functions are classified into five categories: Separable (f_1–f_5), low or moderately conditioned (f_6–f_9), unimodal with high conditioning (f_{10}–f_{14}), multimodal with adequate global structure (f_{15}–f_{19}), and multimodal with weak global structure (f_{20}–f_{24}). Transformations include linear translations and rotation, and symmetry breaking through oscillations about the identity. Each instance is uniquely identified by an index. For each basis function, we generated instances $[1, \ldots, 30]$ at $D = 2$, resulting in a total of 720 problem instances. A detailed qualitative description of each basis function is presented in [31].

As a test set we generated instances of the Sphere (Eq. 1), Rastrigin (Eq. 2) and Bent cigar (Eq. 3) functions where $\mathbf{x}^\star \in \mathcal{X}$ and $y^\star \in [-1000, 1000]$ cause translational shifts on $\mathcal{X}$ and $\mathcal{Y}$ respectively, $\mathbf{z} = \mathbf{R}T^{0.5}_{\text{asy}}\left(\mathbf{R}\left(\mathbf{x} - \mathbf{x}^\star\right)\right)$, $\mathbf{R}$ is 60°orthogonal rotation matrix, and $T^{0.5}_{\text{asy}}$ is an asymmetric transformation defined by (4). The values of $\mathbf{x}^\star$ for each instance trace a positive diagonal line through $\mathcal{X}$, i.e., $\mathbf{x}^\star = [x^\star_1, x^\star_2] : x^\star_1 = x^\star_2$, and increase at intervals of 0.05. The values of $y^\star$ increase at intervals of 200. The resulting parametric grid produces 2211 instances per function, for a total of 6633 test instances. Figure 1 illustrates nine of these instances. We aim to observe how well the space generated by the ELA measures responds to these translations of the cost function.

$$f_S\left(\mathbf{x}\right) = \left\|\mathbf{x} - \mathbf{x}^\star\right\|^2 + y^\star \tag{1}$$

$$f_R\left(\mathbf{x}\right) = 20 - 10\sum_{i=1}^{2}\cos\left(2\pi\left(x_i - x^\star_i\right)\right) + \left\|\mathbf{x} - \mathbf{x}^\star\right\|^2 + y^\star \tag{2}$$

$$f_B\left(\mathbf{x}\right) = z_1^2 + 10^6 z_2^2 + y^\star \tag{3}$$

$$T^{0.5}_{\text{asy}} : x_i \mapsto \begin{cases} x_i^{1+0.5\frac{i-1}{D-1}\sqrt{x_i}} & \text{if } x_i > 0 \\ x_i & \text{otherwise} \end{cases} \tag{4}$$

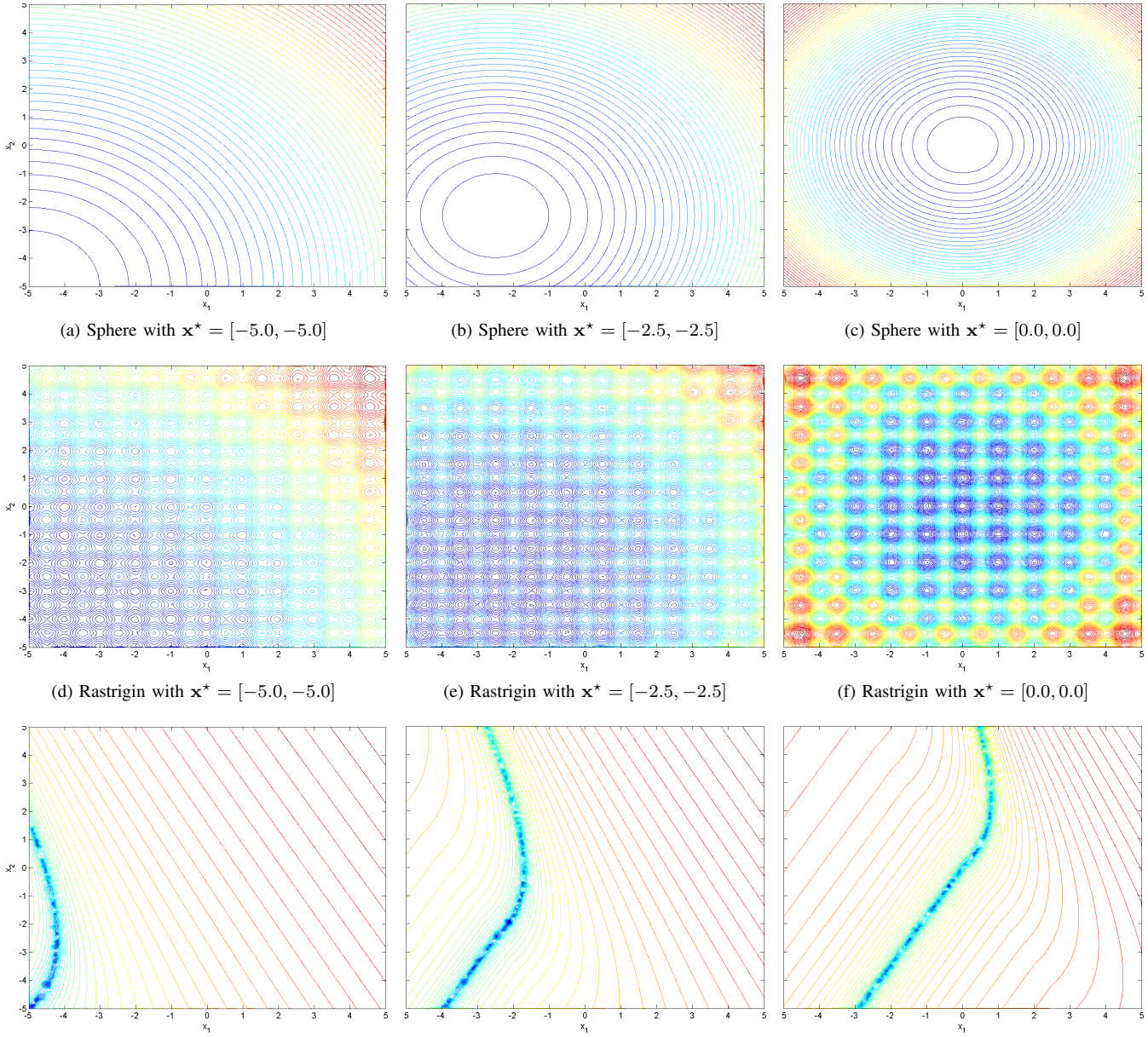
B. Exploratory landscape analysis measures

There are several ELA measures developed for continuous optimization problems, most of which have been adapted from combinatorial optimization [15], [32]. For this work, we implemented the methods summarized in Table I, as they are quick and simple to calculate. Furthermore, these methods can share a sample, guaranteeing that the differences observed on the measures depend only on the instance, and not on sample size or sampling method. For example, the convexity, local search and gradient methods described in [14] cannot share a sample between them nor with the methods in Table I. Besides ensuring a fair comparison, sharing a sample reduces the overall computational cost. Other reported measures with similar sampling simplicity were discarded as they are co-linear with those in Table I [33]. For example: Fitness distance correlation [34], FDC, is co-linear with $\bar{R}^2_Q$ ($\rho = 0.714$). The adjusted coefficient of determination of a linear regression model, $\bar{R}^2_L$, is co-linear with $\bar{R}^2_{LI}$ ($\rho = 0.860$). The adjusted coefficient of determination of a quadratic regression model including variable interactions, $\bar{R}^2_{QI}$, is co-linear with $\bar{R}^2_Q$ ($\rho = 0.871$). The minimum of the absolute value of the linear model coefficients [14], $\min\left(\beta_L\right)$, is co-linear with $H\left(\mathbf{Y}\right)$ ($\rho = 0.865$). The maximum of the absolute value of the linear model coefficients, $\max\left(\beta_L\right)$, is co-linear with $H\left(\mathbf{Y}\right)$ ($\rho = 0.915$). Significance of second order [28], $\xi^{(2)}$, is co-linear with $\xi^{(D)}$ ($\rho = 0.860$). The length scale entropy [35], $H\left(r\right)$, is co-linear with $H\left(\mathbf{Y}\right)$ ($\rho = 0.888$). Settling sensitivity [30], ϵ_S, is co-linear with $H\left(\mathbf{Y}\right)$ ($\rho = 0.884$). Finally, initial partial information, M_0, is co-linear with $H_{\max}$ ($\rho = 0.809$).

We generated an input sample, $\mathbf{X} \subset \mathcal{X}$, of size 2×10^4 using Latin hypercube design (LHD). The output sample, $\mathbf{Y}$, was generated by evaluating $\mathbf{X}$ on an instance. We use the same $\mathbf{X}$ across all the instances from the reference and test sets. After sampling the data, we treat each function as a black-box. We calculate the selected nine measures and normalized them employing the techniques described in Table I. We used principal component analysis (PCA) as a dimensionality reduction method to project the nine dimensional space defined by the measures of the reference set into $\mathbb{R}^2$. The PCA implementation uses the singular value decomposition algorithm to calculate the projection, and centers the data over the mean. The projected coordinates correspond to the two principal eigenvectors, which explain 41.74% and 29.94% of the variance respectively.

III. Results

We used the normalized Euclidean distance, δ, as the measure of dissimilarity between the test instances described in Section II-A. The maximum distance was used as normalization reference, which is equal to 1.0946 for f_S, 1.1189 for f_R, and 3.5013 for f_B. The average δ between instances



(a) Sphere with $\mathbf{x}^\star = [-5.0, -5.0]$ (b) Sphere with $\mathbf{x}^\star = [-2.5, -2.5]$ (c) Sphere with $\mathbf{x}^\star = [0.0, 0.0]$

(d) Rastrigin with $\mathbf{x}^\star = [-5.0, -5.0]$ (e) Rastrigin with $\mathbf{x}^\star = [-2.5, -2.5]$ (f) Rastrigin with $\mathbf{x}^\star = [0.0, 0.0]$

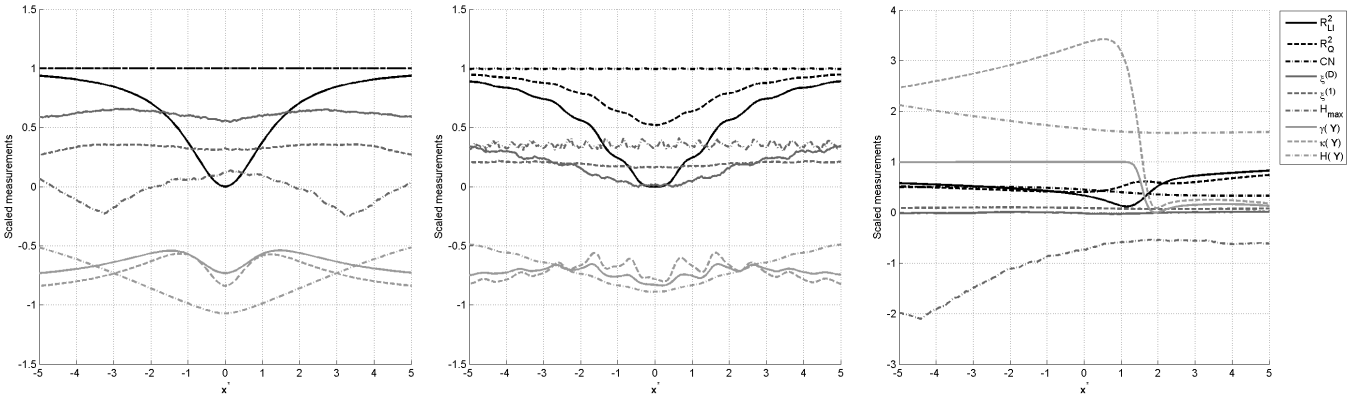
(g) Bent cigar with $\mathbf{x}^\star = [-5.0, -5.0]$ (h) Bent cigar with $\mathbf{x}^\star = [-2.5, -2.5]$ (i) Bent cigar with $\mathbf{x}^\star = [0.0, 0.0]$

Fig. 1. Examples of instances from the test set.

TABLE I

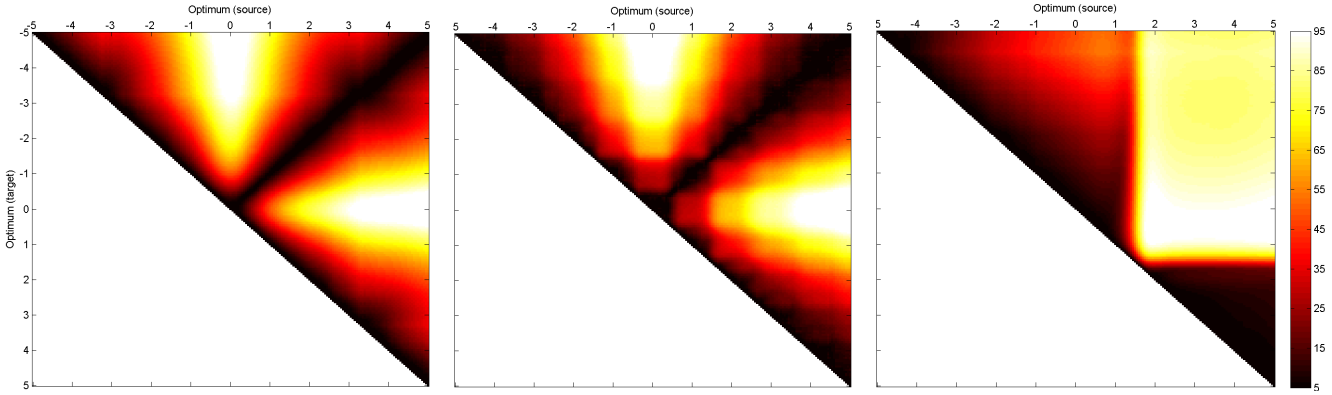
SUMMARY OF THE EXPLORATORY LANDSCAPE ANALYSIS MEASURES EMPLOYED IN THIS PAPER. EACH ONE OF THEM WAS NORMALIZED USING THE TRANSFORMATIONS LISTED.

Method	Feature	Description	Transformations	Reference
Surrogate models	$\bar{R}^2_{LI}$	Adjusted coefficient of determination of a linear regression model including variable interactions	Unit scaling	[14]
	$\bar{R}^2_{Q}$	Adjusted coefficient of determination of a purely quadratic regression model	Unit scaling	
	CN	Ratio between the minimum and the maximum absolute values of the quadratic term coefficients in the purely quadratic model	Unit scaling	
Significance	$\xi^{(D)}$	Significance of D-th order	z-score, $\tanh$	[28]
	$\xi^{(1)}$	Significance of first order	z-score, $\tanh$	
Cost distribution	$\gamma(\mathbf{Y})$	Skewness of the cost distribution	z-score, $\tanh$	[14], [29]
	$\kappa(\mathbf{Y})$	Kurtosis of the cost distribution	$\log_{10}$, z-score	
	$H(\mathbf{Y})$	Entropy of the cost distribution	$\log_{10}$, z-score	
Fitness sequences	$H_{\max}$	Maximum information content with nearest neighbor sorting	z-score	[30]



(a) Sphere (b) Rastrigin (c) Bent cigar

Fig. 2. Values of the measures for the test instances with $y^\star = 0$. The horizontal axes represent the location of $\mathbf{x}^\star$. The measures were scaled using the methods described in Table I.



(a) Sphere (b) Rastrigin (c) Bent cigar

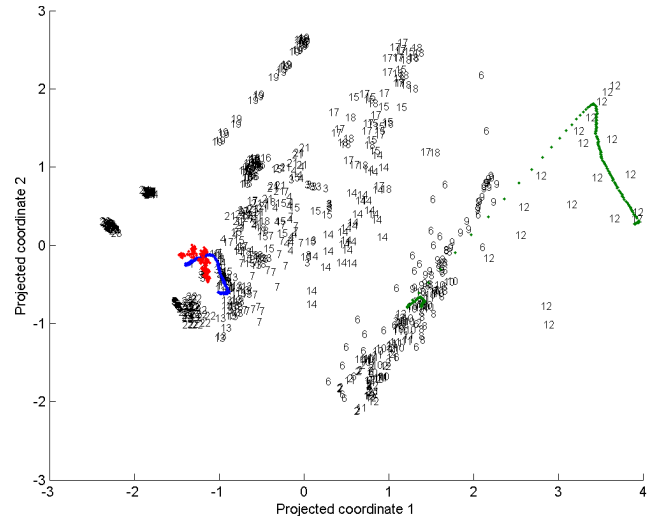
Fig. 3. Normalized euclidean distance, δ, between test instances with $y^\star = 0$ on the measure space. Both the horizontal and vertical axes represent the location of $\mathbf{x}^\star$. The colorbar indicates the magnitude of δ as a percentage.

with the same $\mathbf{x}^\star$ but different $y^\star$ (and its standard deviation) is $10^{-8}\left(1.5 \times 10^{-8}\right)$ for f_S, $8.7 \times 10^{-9}\left(1.3 \times 10^{-8}\right)$ for f_R, and $1.6 \times 10^{-9}\left(3.9 \times 10^{-8}\right)$ for f_B, meaning that the measures in Table I are invariant on $\mathcal{Y}$, i.e., there is no effect on the measures produced by translations on $\mathcal{Y}$, as expected from their definitions. Therefore, we focus on the effect that translations on $\mathcal{X}$ have on the measures.

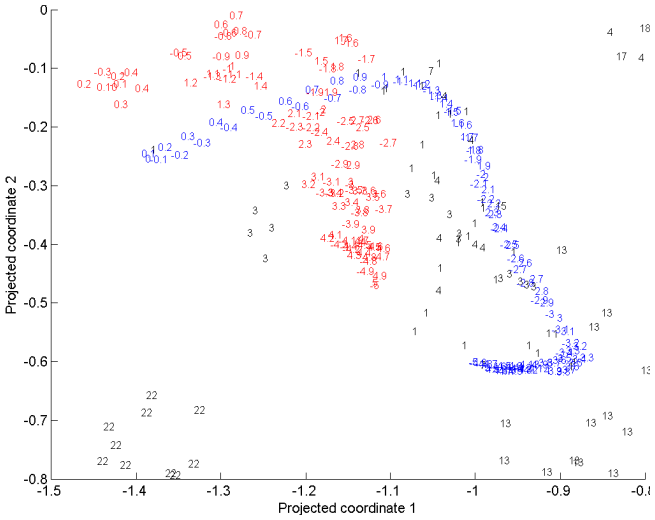
Figure 2 illustrates the values of the measures for the test instances with $y^\star = 0$ after being scaled using the methods described in Table I. The horizontal axes represent the location of $\mathbf{x}^\star$. The figure shows that most of the measures change in value depending on the location of the optimum and the generating function, excepting $\bar{R}^2_Q$ and CN for f_S. Large variations can be observed for $\bar{R}^2_{LI}$, which fluctuates from the top to the bottom of the scale for $\{f_S, f_R\}$. Abrupt transitions can be observed in $H_{\max}$ for f_S, $\left\{\kappa\left(\mathbf{Y}\right), \xi^{(1)}\right\}$ for f_R, and $\left\{\gamma\left(\mathbf{Y}\right), \kappa\left(\mathbf{Y}\right), H_{\max}\right\}$ for f_B. In fact, a phase transition can be observed for f_B with $\mathbf{x}^\star > \mathbf{1}$. This means that a linear translation on $\mathcal{X}$ often results in non-linear changes on the measures.

The effect on the dissimilarity measure is illustrated in Figure 3 as a heat map. The axes represent the location of the $\mathbf{x}^\star$. Black and white represent δ values below 5% and above 95% respectively. Through these graphs, we observe that δ is maximum between $\mathbf{x}^\star = \pm\mathbf{5}$ and $\mathbf{x}^\star = \mathbf{0}$ for $\{f_S, f_R\}$. The phase transition for f_B is also visible in this figure, with the largest difference between $\mathbf{x}^\star = \mathbf{1.95}$ and $\mathbf{x}^\star = \mathbf{0.55}$. These results match our expectations. Both f_S and f_R have axes of symmetry crossing the optimum. Therefore, as $\mathbf{x}^\star$ approaches $\mathcal{X}$ bounds, the global structure of the landscape resembles more to a plane than to a sphere, as the value of $\bar{R}^2_{LI}$ indicates. On the other hand, for f_B with $\mathbf{x}^\star < \mathbf{0}$, the bent ridge dominates the landscape. However, as $\mathbf{x}^\star > \mathbf{0}$, the ellipsoidal part of the function becomes dominant. Eventually the translation takes any measure that would differentiate f_B with the rotated ellipsoidal function (f_{10}). Therefore, we cannot assume that any structure that significantly alters the landscape is included in the function bounds.

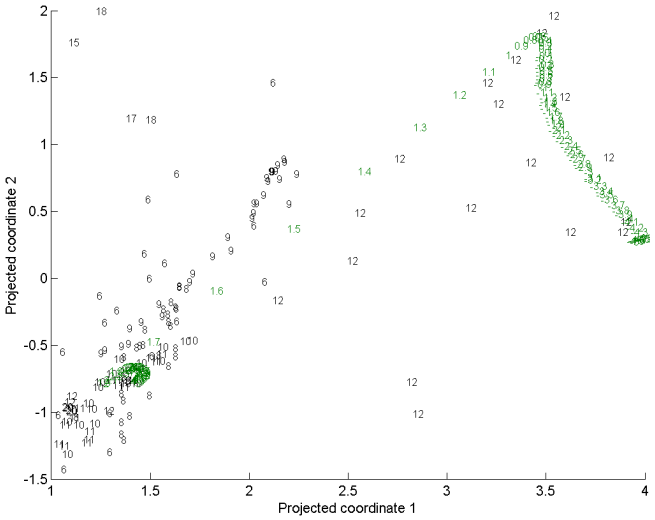
To visualize the path generated by the translations on $\mathcal{X}$, we illustrate the projected measure space in Figure 4a. The numbers in black indicate the basis function from the COCO benchmark from which the instance is derived. The blue and



(a) Reference and test instances

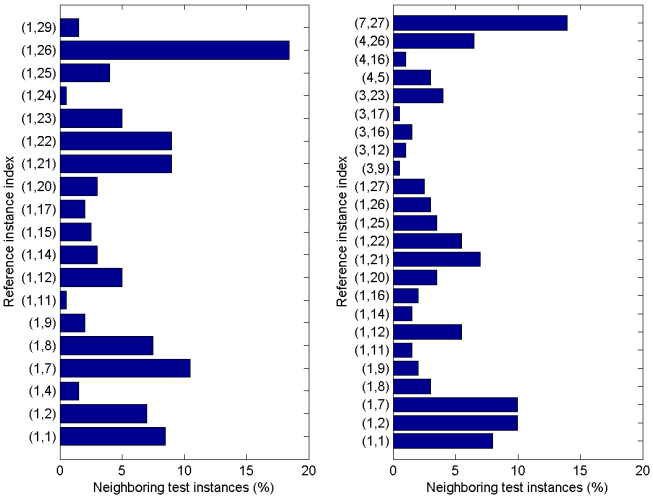


(b) Region containing the Sphere and Rastrigin functions



(c) Region containing the Bent cigar function

Fig. 4. Bidimensional projection of the feature space using PCA.



(a) (b)

Fig. 5. Nearest neighboring reference instances for the f_S instances. Figure (a) shows the results in the measure space, whereas Figure (b) shows the results on the projected space.

red marks on the left side of the space represent the f_S and f_R instances respectively, whereas the green marks at the right side of the space represent the f_B instances. Figure 4b shows a close up on the region containing f_S and f_R, whereas Figure 4c shows a close up on the region containing f_B. The marks now represent the value of $\mathbf{x}^\star$. Since f_S and f_R are symmetric, we expected an overlapping between the instances with equal absolute value of $\mathbf{x}^\star$. However, small variations can be observed over such instances, which may be attributed to numerical and projection accuracy. These figures show that substantial variation on the measure values may be observed due to relatively minor change in the basis function, such as a translation. Therefore, it is possible to translate a function to the point where the same measures can be obtained with a different basis function. For example, Figure 4c shows the phase transition on f_B for $\mathbf{1} < \mathbf{x}^\star < \mathbf{1.7}$, with the instances becoming similar to the rotated ellipsoidal function (f_{10}) with $\mathbf{x}^\star > \mathbf{1.7}$.

Figure 4b also shows that the neighboring instances of f_S includes the Rastrigin (f_3), Büche-Rastrigin, (f_4), linear slope, (f_5), step ellipsoidal, (f_7), and sharp ridge (f_{13}) functions. To verify whether this is a distortion introduced by PCA, we find the nearest neighboring instance from the reference set to each instance of f_S. Figure 5a illustrates the percentage of neighboring instances on the measure space, whereas Figure 5b illustrates the percentage of neighboring instances on the projected space. While the nearest neighbors in the measure space are all instances of the sphere function, f_1, the neighborhood in the projected space has folded to include, among others, instance 27 of f_7. This implies that dimensionality reduction, in this case PCA, alters the results in a potentially deceiving way.

To evaluate which reference instances were substantially transformed by the PCA, we calculate the absolute differ-

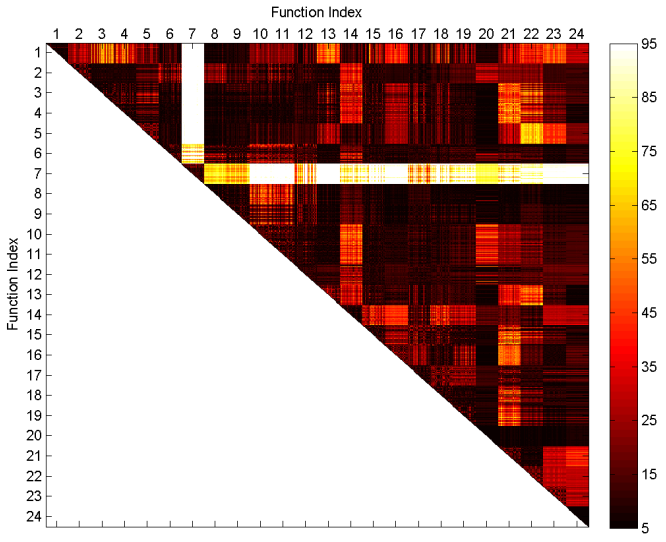


Fig. 6. Normalized absolute difference in δ between the measure space and its bidimensional projection for the reference instances.

ence in δ between the measure space and its bidimensional projection for the reference instances. Figure 6 illustrates the results normalized using the maximum, which is equal to 0.6096. The figure shows that f_7 is the most affected function, loosing contrast against all the functions, particularly with $\{f_1, f_3, f_4, f_{13}, f_{22}\}$. This implies that the discerning characteristic from f_7 —its plateaus— is not effectively preserved in the projected space. Other affected functions are $\{f_1, f_{14}, f_{22}, f_{23}, f_{24}\}$.

IV. Conclusion

At the beginning of this paper, we challenged the assumption that robust measures will capture the fundamental characteristics of a function defined within a bound-constrained region, when they slightly vary across instances. Hence, it may be inadequate to average a measure across all the instances generated from the same function, or to generalize the measure resulting from a single instance. Furthermore, we noted that dimensionality reduction techniques produce a compromised space, which may have altered neighborhood structures. Hence, we proposed two motivating questions: How does the translation of a function affects the landscape measures? How is the instance neighborhood affected by a two dimensional projection? To answer these questions, we followed the experimental methodology outlined in Section II, whose results we presented in Section III. We finalize our paper now by providing answers to the questions according to the evidence presented, and proposing some further lines of inquiry.

The results demonstrate that all the measures under are invariant to translations on $\mathcal{Y}$, while some are invariant to translations on $\mathcal{X}$ on limited cases, e.g., $\bar{R}^2_{LI}$ and CN for f_S. However, translations on $\mathcal{X}$ can produce non-linear fluctuations, even phase transitions, in the measures. Since the significant elements in the landscape may be pushed out from the bound-constrained region, two instances from the same basis function could be located at different parts of the measure space. It is to be expected that rotations of the basis function, under-sampling and increased dimensionality may produce similar effects. As a result, we may question our ability to make critical decisions based on the measures. We should agree that each measure gives an extremely limited picture of the function structure within the bound-constrained region. As such, they should not be interpreted without considering the values of other measures, the size of the bound-constrained region, the sample size, and a measure of algorithm performance. This is not a drawback, as we want measures that reflect different function behavior and not only extreme changes. Nevertheless, we should avoid generalizing any level of robustness to all possible functions and instances after a limited experimental validation, without a rigorous analysis that includes the effects due to the bounds.

The results also show how a dimensionality reduction technique, such as PCA, alters the neighbor relationships. For example, the step ellipsoidal function (f_7), lost contrast against all other functions, particularly against those with whom it shares a global structure, such as the Sphere (f_1), Rastrigin (f_3), and Büche-Rastrigin, (f_4). In other words, f_7 discerning characteristic —its plateaus— was not effectively preserved in the projected space. While this effect is to be expected, as the resulting low dimensional space is a compromise, it has implications on visualization and function generation. For example, bidimensional projections have been used in other optimization branches to identify gaps between instances where new ones could be produced [12]. Given these altered neighborhood structures observed for the COCO benchmark set, we might not be able to obtain a point in such a location nor expect that an instance equidistant from other two may represent a hybrid. Since a translation on $\mathcal{X}$ generates a non-linear path in both the measure and the projected spaces, we should expect that also a non-linear path must be followed to transform an instance into another in either space. The results suggest that each basis function generates a region within the space where all derived instances are located. Such regions may overlap or form a network whose connections are non-linear paths, outside of which an instance cannot be generated.

Arguably, the effects observed may be attributed to our choice of dimensionality reduction technique, given that PCA does not explicitly attempt to retain the neighborhood information. However, PCA continues to be popular and it may be considered a reasonable first choice, even though there are other reported techniques [36]. Besides, the level of topological distortion does not invalidate our conclusions regarding the location of hybrids and transformation between functions. We are examining other techniques to minimize the topological distortion on the bidimensional projection as part of our ongoing research. However, the results so far do not give a single technique radical advantages over all others on this problem, specially since some techniques are equivalent under certain conditions, e.g., PCA is equal to

multi-dimensional scaling (MDS) with Euclidean distances [37]. It is also noteworthy that each dimensionality reduction technique performs depending on the characteristics of the high dimensional data [38], which implies another algorithm selection problem.

Other areas of inquiry are an analysis of the effects that sample size and randomness, as well as noisy functions have on the ELA measures. A better understanding of the space generated by the measures is part of our work into generating a new black-box continuous optimization benchmark set that fills this space.

ACKNOWLEDGMENTS

This work has been funded by the Australian Research Council through grant No. DP120103678. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

REFERENCES

[1] P. Hough and P. Williams, "Modern machine learning for automatic optimization algorithm selection," in *Proceedings of the INFORMS Artificial Intelligence and Data Mining Workshop*, 2006, pp. 1–6.

[2] K. Sörensen, "Metaheuristics— the metaphor exposed," *Int. T. Oper. Res.*, 2013.

[3] D. Weyland, "A rigorous analysis of the harmony search algorithm: How the research community can be misled by a "novel" methodology," *Int. J. Appl. Metaheuristic Comput.*, vol. 1, no. 2, pp. 50–60, 2010.

[4] M. Črepinšek, S. Liu, and L. Mernik, "A note on teaching-learning-based optimization algorithm," *Inform. Sciences*, vol. 212, no. 0, pp. 79 – 93, 2012.

[5] A. D. Corte and K. Sörensen, "Optimisation of gravity-fed water distribution network design: A critical review," *Eur. J. Oper. Res.*, vol. 228, no. 1, pp. 1 – 10, 2013.

[6] A. P. Piotrowski, J. J. Napiorkowski, and P. M. Rowinski, "How novel is the "novel" black hole optimization approach?" *Inform. Sciences*, vol. 267, pp. 191–200, 2014.

[7] A. Auger and B. Doerr, *Theory of Randomized Search Heuristics*. World Scientific, 2011.

[8] K. Tang, F. Peng, G. Chen, and X. Yao, "Population-based algorithm portfolios with automated constituent algorithms selection," *Inform. Sciences*, vol. 279, pp. 94–104, 2014.

[9] C. Blum, J. Puchinger, G. Raidl, and A. Roli, "Hybrid metaheuristics in combinatorial optimization: A survey," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4135 – 4151, 2011.

[10] K. Smith-Miles, "Cross-disciplinary perspectives on meta-learning for algorithm selection," *ACM Comput. Surv.*, vol. 41, no. 1, pp. 6:1–6:25, January 2009.

[11] F. Hutter, L. Xu, H. Hoos, and K. Leyton-Brown, "Algorithm runtime prediction: Methods & evaluation," *Artif. Intell.*, vol. 206, no. 0, pp. 79 – 111, 2014.

[12] K. Smith-Miles, D. Baatar, B. Wreford, and R. Lewis, "Towards objective measures of algorithm performance across instance space," *Comput. Oper. Res.*, vol. 45, no. 0, pp. 12–24, 2014.

[13] L. Kotthoff, "Algorithm selection for combinatorial search problems: A survey," *AI Magazine*, vol. 35, no. 3, 2014.

[14] O. Mersmann, B. Bischl, H. Trautmann, M. Preuß, C. Weihs, and G. Rudolph, "Exploratory landscape analysis," in *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, ser. GECCO '11. New York, NY, USA: ACM, 2011, pp. 829–836.

[15] K. Malan and A. Engelbrecht, "A survey of techniques for characterising fitness landscapes and some possible ways forward," *Inform. Sciences*, vol. 241, no. 0, pp. 148 – 163, 2013.

[16] B. Bischl, O. Mersmann, H. Trautmann, and M. Preuß, "Algorithm selection based on exploratory landscape analysis and cost-sensitive learning," in *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference*. New York, NY, USA: ACM, 2012, pp. 313–320.

[17] M. Muñoz, M. Kirley, and S. Halgamuge, "A meta-learning prediction model of algorithm performance for continuous optimization problems," in *Proceedings of Parallel Problem Solving from Nature (PPSN XII)*, ser. Lect. Notes Comput. Sci., vol. 7941, 2012, pp. 226–235.

[18] K. Malan and A. Engelbrecht, "Characterising the searchability of continuous optimisation problems for PSO," *Swarm Intell.*, pp. 1–28, 2014.

[19] J. Beck and J. Watson, "Adaptive search algorithms and fitness-distance correlation," in *Proceedings of the Fifth Metaheuristics International Conference*, 2003, pp. 1–6.

[20] T. Smith, P. Husbands, P. Layzell, and M. O'Shea, "Fitness landscapes and evolvability," *Evol. Comput.*, vol. 10, no. 1, pp. 1–34, 2002.

[21] C. Müller and I. Sbalzarini, "Global characterization of the CEC 2005 fitness landscapes using fitness-distance analysis," in *Applications of Evolutionary Computation*, ser. Lect. Notes Comput. Sc., vol. 6624. Springer, 2011, pp. 294–303.

[22] T. Jansen, "On classifications of fitness functions," University of Dortmund, Tech. Rep. CI-76/99, November 1999.

[23] M. Tomassini, L. Vanneschi, P. Collard, and M. Clergue, "A study of fitness distance correlation as a difficulty measure in genetic programming," *Evol. Comput.*, vol. 13, no. 2, pp. 213–239, June 2005.

[24] J. He, C. Reeves, C. Witt, and X. Yao, "A note on problem difficulty measures in black-box optimization: Classification, realizations and predictability," *Evol. Comput.*, vol. 15, no. 4, pp. 435–443, 2007.

[25] B. Naudts and L. Kallel, "A comparison of predictive measures of problem difficulty in evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 1, pp. 1–15, April 2000.

[26] M. Muñoz, M. Kirley, and S. Halgamuge, "Landscape characterization of numerical optimization problems using biased scattered data," in *Proceedings of the 2012 Congress on Evolutionary Computation (CEC2012)*, 2012, pp. 1–8.

[27] N. Hansen, A. Auger, S. Finck, and R. Ros, "Real-parameter black-box optimization benchmarking BBOB-2010: Experimental setup," INRIA, Tech. Rep. RR-7215, September 2010.

[28] D. Seo and B. Moon, "An information-theoretic analysis on the interactions of variables in combinatorial optimization problems," *Evol. Comput.*, vol. 15, no. 2, pp. 169–198, 2007.

[29] J. Marin, "How landscape ruggedness influences the performance of real-coded algorithms: a comparative study," *Soft Comput.*, vol. 16, no. 4, pp. 683–698, 2012.

[30] M. Muñoz, M. Kirley, and S. Halgamuge, "Exploratory landscape analysis of continuous space optimization problems using information content," *IEEE Trans. Evol. Comput.*, vol. 19, no. 1, pp. 74–87, February 2015.

[31] O. Mersmann, M. Preuß, H. Trautmann, B. Bischl, and C. Weihs, "Analyzing the BBOB results by means of benchmarking concepts," *Evol. Comput.*, pp. –, June 2014.

[32] E. Pitzer and M. Affenzeller, "A comprehensive survey on fitness landscape analysis," in *Recent Advances in Intelligent Engineering Systems*, ser. Studies Comput. Intell. Springer Berlin Heidelberg, 2012, vol. 378, pp. 161–191.

[33] M. Muñoz, "Decision support systems for the automatic selection of algorithms for continuous optimization problems," Ph.D. dissertation, The University of Melbourne, November 2013.

[34] T. Jones and S. Forrest, "Fitness distance correlation as a measure of problem difficulty for genetic algorithms," in *Proceedings of the Sixth International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers Inc., 1995, pp. 184–192.

[35] R. Morgan and M. Gallagher, "Length scale for characterising continuous optimization problems," in *Proceedings of Parallel Problem Solving from Nature (PPSN XII)*, ser. Lect. Notes Comput. Sc., 2012, pp. 407–416.

[36] L. van der Maaten, E. Postma, and J. van den Herik, "Dimensionality reduction: A comparative review," Tilburg University, Tech. Rep. TiCC TR 2009–005, October 2009.

[37] C. Williams, "On a connection between kernel PCA and metric multidimensional scaling," *Mach. Learn.*, vol. 46, no. 1–3, pp. 11–19, 2002.

[38] G. Wang, Q. Song, H. Sun, X. Zhang, B. Xu, and Y. Zhou, "A feature subset selection algorithm automatic recommendation method," *J. Artif. Intell. Res.*, vol. 47, pp. 1–34, 2013.