Revealing Domain-Spatiality Patterns for Configuration Tuning: Domain Knowledge Meets Fitness Landscapes (Appendix)

Anonymous Author(s)

1 Landscape Metrics

1.1 Fitness Distance Correlation (FDC)

FDC is a widely used metric for evaluating the relationship between the fitness (e.g., runtime and throughput) of solutions (i.e., configurations) and their distance to the global optimum in a given landscape [1]. It helps to characterize the global structure of the fitness landscape and evaluate the ease or difficulty of navigating toward the global optimum. Formally, FDC is calculated as:

$$\varrho(f, d) = \frac{1}{\sigma_f \sigma_d} \frac{1}{s} \sum_{i=1}^{s} \left(f_i - \bar{f}\right)\left(d_i - \bar{d}\right), \tag{1}$$

where s is the number of solutions considered in FDC. Within such a solutions set, d_i denotes the distance of ith solution to the nearest global optimum. $\bar{f}$ ($\bar{d}$) and σ_f (σ_d) are the mean and standard deviation of fitness (and distance).

1.2 Local optima

Local optima capture the underlying local structural properties of a problem's landscape. Formally, a configuration $\boldsymbol{c}^\ell$ is considered as a local optima if $f(\boldsymbol{c}^\ell)$ is better than $f(\boldsymbol{c})$, $\forall \boldsymbol{c} \in \mathcal{N}\left(\boldsymbol{c}^\ell\right)$, where $\mathcal{N}\left(\boldsymbol{c}^\ell\right)$ is the neighborhood of $\boldsymbol{c}^\ell$. Two key aspects are considered when analyzing local optima: (1) the number of local optima; and (2) the quality of local optima.

1.2.1 The number of local optima. The number of local optima reflects landscape multimodality, indicating the presence of multiple high-performing regions and complexity of the problem. To quantify this, we define ℓ_p as the proportion of local optima among all sampled configurations:

$$\ell_p = \frac{|\mathcal{L}|}{|\mathcal{S}|}, \tag{2}$$

where $\mathcal{L}$ and $\mathcal{S}$ denote the sets of local optima and all sampled configurations, respectively.

1.2.2 The quality of local optima. The quality of local optima particularly their relative performance compared to the global optima—indicates the prominence of local peaks and the susceptibility of the landscape to trapping suboptimal solutions. For minimization problems, we define the relative quality of local optima as:

$$\ell_q = \frac{\frac{1}{|\mathcal{S}|} \sum\limits_{\boldsymbol{c} \in \mathcal{S}} f(\boldsymbol{c}) - \frac{1}{|\mathcal{L}|} \sum\limits_{\boldsymbol{c}^\ell \in \mathcal{L}} f(\boldsymbol{c}^\ell)}{\frac{1}{|\mathcal{S}|} \sum\limits_{\boldsymbol{c} \in \mathcal{S}} f(\boldsymbol{c}) - f(\boldsymbol{c}^g)}, \tag{3}$$

where $\boldsymbol{c}^g$ represents the global optimum, and $\boldsymbol{c}^\ell$ denotes a local optima from $\mathcal{L}$. A higher ℓ_q indicates that local optima are of higher quality, being closer to the global optimum, while a lower ℓ_q suggests an abundance of low-quality local optima.

1.3 Basin of attraction

The basin of attraction [2] of a local optimum $\boldsymbol{c}^\ell$, denoted as $\mathcal{B}(\boldsymbol{c}^\ell)$, is the set of all configurations that converge to $\boldsymbol{c}^\ell$ under a local search process. Formally, it can be defined as:

$$\mathcal{B}(\boldsymbol{c}^\ell) = \{\boldsymbol{c} \in C \mid \text{LocalSearch}(\boldsymbol{c}) \to \boldsymbol{c}^\ell\}. \tag{4}$$

Here, we consider hill climbing as the local search algorithm, and its step size follows the neighborhood structure defined in our main manuscript. In other words, if neighbors differ by d options, the local search needs to take d options per step, ensuring consistency in identifying basin of attraction.

1.4 Autocorrelation

Autocorrelation [4] quantifies the ruggedness of a fitness landscape by measuring how fitness values change across neighboring configurations. Formally, it is defined as:

$$r(d) = \frac{\sum_{i=1}^{N-d} \left(f\left(\boldsymbol{c}_i\right) - \bar{f}\right)\left(f\left(\boldsymbol{c}_{i+d}\right) - \bar{f}\right)}{\sum_{i=1}^{N-d} \left(f\left(\boldsymbol{c}_i\right) - \bar{f}\right)^2}, \tag{5}$$

where $f(\boldsymbol{c}_i)$ is the fitness of ith visited configuration $\boldsymbol{c}_i$ in the random walk [3], $\boldsymbol{c}_{i+d}$ is the next visited neighbor at step size d, and N is the walk length. A high $r(d)$ indicates a smooth landscape, where similar configurations yield comparable fitness values, while a low autocorrelation suggests a rugged landscape with abrupt fitness variations. Notably, to accommodate analysis on sampled datasets, our random walk starts with a step size of $d = 1$. If no neighbor is found, d is incrementally increased until a valid neighbor is located, ensuring continuity. The walk proceeds until all configurations in the dataset are visited, recording the total walk length N. Finally, autocorrelation is computed using Equation (5).

References

[1] Terry Jones, Stephanie Forrest, et al. 1995. Fitness distance correlation as a measure of problem difficulty for genetic algorithms.. In *ICGA*, Vol. 95. 184–192.

[2] Katherine Mary Malan. 2021. A survey of advances in landscape analysis for optimisation. *Algorithms* 14, 2 (2021), 40.

[3] Jorge Tavares, Francisco B Pereira, and Ernesto Costa. 2008. Multidimensional knapsack problem: A fitness landscape analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38, 3 (2008), 604–616.

[4] Edward Weinberger. 1990. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological cybernetics* 63, 5 (1990), 325–336.