Generalizing Diversity with the Signature Transform

Arno Feiden*
Fraunhofer SCAI
Sankt Augustin, Germany
arno.feiden@scai.fraunhofer.de

Biagio Paparella
Institut für Numerische Simulation
Universität Bonn
Bonn, Germany
paparell@ins.uni-bonn.de

Jochen Garcke†
Institut für Numerische Simulation
Universität Bonn
Bonn, Germany
garcke@ins.uni-bonn.de

ABSTRACT

Defining differences is a necessary prerequisite for finding diversity in solutions to a given problem. Diversity, in turn, is a property that is difficult to quantify, but is expected to bring favourable properties.

We introduce the signature transform as a general behaviour descriptor to distinguish solutions to control problems specifically in the context of Quality-Diversity and MAP-Elites. We define a robustness score and profile, a structured analysis of the behaviour of agents and populations of agents under a changing environment, as an instance of a beneficial property of diversity.

The signature transform integrates well with the CVT-Elites approach and offers a functional generalized diversity measure. The robustness analysis substantiates the abstract diversity induced by distance in signature space through tangible effects and enriches the usual evaluation criteria of diverse populations.

The generalization of diversity from handcrafted behaviour descriptors opens the possibility to utilise Quality-Diversity techniques for problems where the alignment of the problem with diversity measures is not obvious.

CCS CONCEPTS

• **Computing methodologies → Evolutionary robotics**; *Continuous space search*; • **Computer systems organization → Redundancy**; • **Mathematics of computing** → *Time series analysis.*

KEYWORDS

Quality-Diversity, Robustness, MAP-Elites, Signature transform, Time series analysis, Behavioural diversity

ACM Reference Format:
Arno Feiden, Biagio Paparella, and Jochen Garcke. 2024. Generalizing Diversity with the Signature Transform. In *Genetic and Evolutionary Computation Conference (GECCO '24 Companion), July 14–18, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3638530.3654295

*also with Universität Bonn
†also with Fraunhofer SCAI

GECCO '24 Companion, July 14–18, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0495-6/24/07.
https://doi.org/10.1145/3638530.3654295

1 INTRODUCTION

Quality-Diversity (QD) concerns problems, where not one single optimum, but rather a range of different options can legitimately be called a solution. These solutions to the original problem differ among each other along the lines of newly introduced dimensions of diversity. In practice they are a choice of the user. The usefulness of the results may be defined completely by the user. Otherwise, the user should be attentive to align the dimensions of diversity with the goal they try to achieve.

Cultivating diverse populations holds certain intrinsic advantages: diversity breeds innovation. This innovation can be seen and measured in complicated exploration scenarios as in mazes [10], and through robustness scenarios, as when a robot has been damaged and tries to find in its behaviour repertoire, created by a QD algorithm, a way to move that still works with one less limb [1].

While these ideas acknowledge the intrinsic value of diverse populations, they still rely on the alignment of the diversity in anticipation of the problem the population shall overcome. It could be the implicit information given by the diversity measure that improves exploration and robustness.

In contrast to a specific behaviour characteristic that aims at alignment to a specific purpose, we pursue the idea of a **generalized diversity measure**. A diversity measure can be called general if it builds on a behaviour descriptor that indiscriminately uses the available information and is neither limited to the environment nor tied to the intended application of the population.

We suggest the signature transform on rollouts of a learned agent as a general behaviour descriptor: it will harmonize the length of episodes and is generally contained in $[-1, 1]$, which facilitates defining a QD-archive structure in which behaviour characteristics are categorized. It is feasible to compare signatures with an L^2-norm and to aggregate signatures of multiple rollouts with a mean, which we interpret as the numerical approximation of the expected signature. This helps to comprehensively represent the behaviour of the agent to avoid the phenomenon of (un)lucky individuals. Also, save truncation, the results of the signature transform still contain the complete information of the rollouts.

To substantiate this kind of general diversity has intrinsic value, we turn to robustness: a diverse population should contain members that react differently to changes of the environment, some better some worse. But from a global viewpoint, diversity should make a population more robust. We formalize and generalize the way to define a concrete **measure of robustness** of a population with a robustness score and profile. To this end an agent is developed with a standard environment. We then take variables that define the behaviour of the environment and change them to values on a grid, and roll the policy out against this changed environment.

Parallel to a QD score [10], we define a robustness score of an agent and a population. This problem setup itself may be interesting to look at: how can we optimize the robustness score of a single agent or multiple agents? What approaches are beneficial or detrimental to robustness, specifically when allowing or denying the possibility to align learning with the robustness challenge?

The generalized diversity measure introduced here boasts a tangible advantage in comparison to simple diversity measures required by classical approaches, demonstrating two things. First, it is possible to fuel a QD algorithm with a general behaviour characteristic, which is not just the trajectory, to define diversity. Second, the resulting population has the type of useful properties typically expected. There is more to gain and to find within the idea of general diversity.

2 DIVERSITY OF BEHAVIOUR

Quality-Diversity encompasses a set of evolutionary algorithms that aim to output a population of high performing solutions against a posed problem that are at the same time diverse in a predefined measure of their behaviour. MAP-Elites [8] is a specifically prevalent subset of this paradigm that cultivates a population of solutions in predefined niches. Competition between individuals happens only in these niches.

The classic approach divides a low-dimensional behaviour space into a Cartesian grid. High-dimensional spaces can be tackled by evenly dividing the space in a given number of cells using Centroidal Voronoi Tessellations ("CVT Elites") [15]. The evolutionary loop, given a random starting population, consists of sampling the population, creating offspring based on that sample, observing the behaviour of the offspring and putting them in niches depending on their behaviour. If there is more than one member of the resulting population in one niche, the population is reduced to at most one member per niche by local competition.

Many applications of these strategies aim at solving optimal control tasks that are also solved with reinforcement learning techniques. Typically, optimization problems come with a measure of quality through the optimization variable. A measure of diversity is not necessarily obvious and is usually handcrafted based on the intended application of the diverse subpopulation.

Whenever some type of more general diversity is considered, applications are looked for and found in anticipated intrinsic values of diversity. Robustness against damage [15], exploration even when considering seemingly unaligned measures of diversity [10], or coverage of the analytically understood underlying behaviour space of a problem [4].

The most significant aspect of behavioural diversity is the definition of the **behaviour characteristic** or **behaviour descriptor**, a function operating on the space of rollouts of an agent. The whole concept of behavioural diversity further requires a way to compare two of these behaviour characteristics, mostly the Euclidean norm.

Often, behaviour characteristics are picked with a certain problem in mind, that may either directly or indirectly be solved by this choice. The last position of a robot arm or of a maze walker is a direct way to encode the actual objective of a problem into the diversity of the population. When trying to find different gaits in a locomotion task, the ratio of contact with the floor by different parts of the body turns out to be a nice behaviour descriptor. However, especially in that last case, a more abstract idea of different gaits, is not necessarily fully covered by that descriptor, nor is it guaranteed that such a computationally simple, low-dimensional representation can be found for any abstract problem.

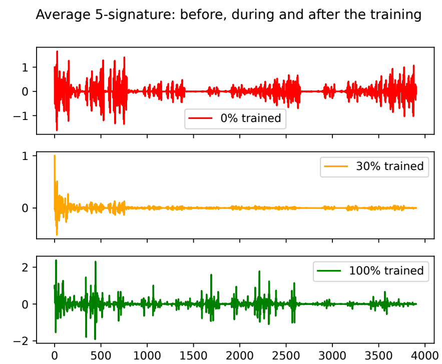

Figure 1: Signatures of depth 5 of the behaviour of an agent before, during and after training of CartPole.

3 THE SIGNATURE TRANSFORM

We make use of the *signature transform* in order to compress the statistical information contained in multiple time series into finite collections of controllable coefficients to quickly quantify statistical differences between random paths.

Given a one dimensional random variable, it is possible to compute its *moments* by averaging its powers. Similarly, on a whole time series, the time series statistical behaviour is compressed by taking averages of appropriate "crossed-powers": the signature transform [3]. It has seen successful application in time series classification [7], especially in form of preprocessing steps, which is closely related to the question on how to distinguish policies according to their behaviour.

Let $f : [0, 1] \to \mathbb{R}^d$ be a piecewise linear function with $f(0) = (0, ..., 0)$. For general functions, preprocess the data accordingly. Here, we always consider its *augmentation* $X(t) = (t, f(t))$ with a time component t.

A multi-index I_n of depth $n \geq 1$ is a sequence of n digits, $I_n = (i_1, \ldots, i_n)$, each digit being an integer between 1 and $d + 1$, included. The k-dimensional canonical simplex is the set $\Delta^k = \{(u_1, \ldots, u_k), u_i \in [0, 1] : 0 < u_1 < \cdots < u_k < 1\}$.

Definition 3.1 (Signature coefficient). The signature coefficient of X corresponding to the multi-index $I_n = (i_1, \ldots, i_n)$ is the real number s_{I_n} defined as:

$$s_{I_n} = \int_{\Delta^n} \dot{X}^{i_1}(u_1) \cdots \dot{X}^{i_n}(u_n) du_1 \cdots du_n, \tag{1}$$

where $\dot{X}^k$ is the derivative of the k-th component of the path X.

For each depth n, we have $(d + 1)^n$ possible multi-indices. If we consider all the indices until depth n, for a d dimensional path, we obtain a total of $\sum_{i=1}^{n} (d + 1)^i$ coefficients.

The *signature transform* $S(X)$ is the infinite sequence of all the possible signature coefficients. The *truncated signatures* $S^N(X)$ is the finite sequence of the signature coefficients up to and including depth N.

It is possible to prove that the truncation error follows an asymptotic decay in the Euclidean norm [5], essentially depending on the *length* of the curve $X(t)$. Any agent in a given environment is a stochastic process, its sampled paths are the rollouts. It can be shown that two stochastic processes have the same probability distribution if and only if the expected values of the corresponding signature transforms are equal ([9], section 3.3). Consequently, the difference of two agents is represented by the Euclidean norm between the average values of the truncated signatures of their rollouts.

4 ROBUSTNESS

Robustness has been suggested as a use case for diverse populations. A repertoire of different behaviours can be cycled through or selected from by a hierarchical system or intelligence to confront the challenge of a change of the problem setup. This approach has been successful where the behaviour characteristic and the change in environment align: a repertoire of gaits that differ by how much a certain leg touches the ground, turns out to be useful in a situation when one of the legs fail [1].

Robustness in a single agent presented here can be seen as a type of flexibility, of how well an agent is able to generalize to adjacent out of distribution problems. This is a sought after property, because it facilitates lifting a policy from incomplete training data to applications where different data points may be encountered [6].

We introduce a method to evaluate the robustness of a trained policy and the joint robustness of a population of trained policies. Choose two parameters in a given environment, we used gravity g and pole length l of the CartPole environment C. CartPole has by default $g = 9.81$ and $l = 1$. We vary them equidistant in a certain range which results in a grid $\{(g_i, l_j)\}_{i=1,\ldots,N, j=1,\ldots,M}$.

This results in a two-dimensional array of modified environments. Now, the same agent is repeatedly rolled out on each environment modified corresponding to the given values. We call the mean value of the optimization variable, the accumulated reward during an episode in the case of CartPole, the *fitness*.

A visualization of these results in a heat map, a two dimensional matrix, where the hue at position (i, j) represents the fitness of the agent given g_i and l_j, gives a visible profile of the robustness that unveils hidden properties of the agent, see fig. 2. Also, a simple quantifiable measure for robustness of an agent is given by adding up the results to a robustness score or averaging the reward over all cells. For a population of agents, for each cell we roll out every agent and take for that cell the performance of the most successful agents as the (joint) robustness of the population.

5 ROBUSTNESS THROUGH DIVERSITY

Elites algorithms allow the evolution of a population that is diverse along certain predefined dimensions. They are defined through the accompanying behaviour descriptor and archive. We claim that specific behaviour descriptors create specific diversity, whereas the intrinsic rewards of general diversity can only be unlocked with general behaviour descriptors. We present an instance of this claim here, by comparing the robustness of a population created with a simple low-dimensional behaviour descriptor and a population created with diversity in signature space.

Table 1: Comparison of the CVT-Sig and ME-Last populations

Population	ME-Last	CVT-Sig	RL-All
Number of Solutions	612	127	48
Contributing Solutions	34	20	16
QD Score	**312486**	122101	n/a
Robustness Score	58655	**102199**	101459
Average Reward	146.6	**255.5**	253.6

Robustness is tested against on values of gravity and pole length from the centre points of a Cartesian grid of $[0, 100] \times [0, 10]$ with $20 \times 20 = 400$ cells and 10 rollouts per modified environment to find the fitness. We use the gymnasium [14], pyribs [13], stable baselines 3 [11], and iisignature [12] as reference implementations.

We trained two populations of solutions for the CartPole problem: one, called ME-Last, with MAP-Elites [8] using a 25×25 grid archive on the physical last position of the cart and the angle of its pole in the space $[-4.8, 4.8] \times [-0.42, 0.42]$. We choose the last position of the cart as a behaviour descriptor here, as it is a simple and often used characteristic [1, 4, 10]. All things considered the last position of the trajectory is the result of all the previous states and the decisions made.

The other population, called CVT-Sig, is trained with CVT-Elites [15] using the signature transform of trajectories of states with depth 3 as a behaviour descriptor. CartPole has a 4-dimensional state space, so each signature has length 155, so we initialize the CVT-archive in $[-1, 1]^{155}$.

Both archives are equipped with 625 cells, the number of cells in the CVT archive set to match those of the grid archive. The CVT archive divides the space in the given number of cells by sampling from it and using k-means clustering to find the same number of centroids. An individual falls into the cell associated with the closest centroid. Both methods use fully connected feed forward neural networks with two hidden layers of size 16 and bias as agents, which are initialized with 0 and mutated by Gaussian noise with $\sigma_0 = 0.1$ for 3000 generations of new 30 offspring, where every new individual is evaluated 3 times. To avoid premature convergence, as CartPole is quickly solved with the maximum reward of 500, we use for both setups annealing [2] with initial value 0 and learning rate 0.5.

Additionally, we trained agents with different single-objective methods to solve the CartPole problem. PPO (Proximal Policy Optimization), A2C (Advantage Actor Critic), and TRPO (Trust Region Policy Optimization) are classic reinforcement learning algorithms that rely on gradient steps, while ARS (Augmented Random Search) is a high performing instance of random search, the same mechanism that ultimately develops the solutions in MAP-Elites. We trained 12 agents per method, each of them with the default hyperparameters of stable baselines 3, except that the networks have layers with reduced size to 16 nodes each, until they always solve the problem setup with the full 500 points in the original environment.

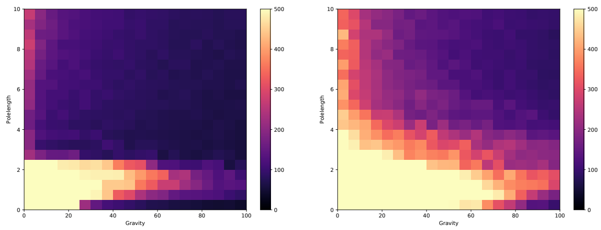

Figure 2: Heat maps showing the joint robustness of the populations ME-Last (a) and CVT-Sig (b) against changing gravity and pole length in the CartPole problem.

We find in table 1 that while the CVT-Sig has a much smaller QD-Score as only about a fifth of the cells are occupied by solutions to the problem, while almost all members of the ME-Last population solve it. However, the joint robustness of these 127 members of the CVT-Sig population is almost double that of the 612 solutions coming from the ME-Last population. Just a small amount of solutions contribute to this maximum. All other solutions of the population are dominated by 34 or 20 agents of the populations ME-Last and CVT-Sig respectively.

This constitutes a clear example of an advantage having a generalized diversity in a population. The abstract diversity implied by the signature transform outperforms the tangible diversity of a last position characterization of behaviour, see also fig. 2. The robustness challenge presents a task that is both concrete and not directly related to the diversity imposed by the learning algorithm. The comparison against the population RL-All of 48 reinforcement learning agents from the four different reinforcement learning algorithms shows the robustness score and average reward of CVT-Sig is on-par to that obtained with sophisticated RL-algorithms.

6 DISCUSSION

It is sensible and potentially advantageous to consider generalized behaviour descriptors of reinforcement learning agents. While the examples delivered here only offer anecdotal evidence, it is still clear, that it is possible to formulate generalized diversity and benefit from it. They constitute a proof of concept for the generalization of behaviour descriptors. Signature transform turns out to be a great fit for the Elites setup, as it normalizes episode length, clearly defines a space for the archive, allows aggregation of multiple rollouts, and is computationally viable with one signature transform per rollout.

By further formalizing the robustness challenge we validated this idea. The associated robustness challenge opens an enormous wealth of benchmark problems, for example to find very robust agents, or to identify algorithms that output robust agents, or to find the conceptual building blocks of robustness both for single agents and for populations.

A robustness profile opens a way to differentiate between multiple solutions to the same problem that seem otherwise indistinguishable. It allowed an extensive review of populations to a problem as simple as the CartPole problem. This may open a path to better understand how solutions of problems develop and evolve, even after apparent convergence to a solution. The examined parameters of robustness of the robustness challenge will still vary from environment to environment, but they are somewhat naturally induced by the problem, which discourages cherry picking and makes it harder to align learning dimensions of diversity with the dimensions of robustness.

Canonically there have been some specific robustness challenges posed, such as the adaption to damage to a robot [15], which require a review using signature transform. Next to robustness, potential challenges to test the ideas of generalized diversity are exploration [10] and coverage of behaviour [4], problems that have been shown to benefit from selecting for diversity.

Similarly, representatives of diverse populations can be selected as stepping stones for further evolution, even with gradient-based methods. This may consolidate robustness in even smaller populations and improve overall robustness, as gradient-based methods like PPO are apparently superior in developing robust agents than the random search represented by ARS.

Generalized diversity is a lot harder to pin down than diversity specified to concrete characteristics. Following ideas from evolutionary computing seems to the correct path to find a better grasp on generalized diversity, which conceptually and evidently can unlock new favourable results that are hard or even impossible to engineer with specified, guided diversity.

REFERENCES

[1] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. 2015. Robots that can adapt like animals. *Nature* 521 (05 2015), 503–507.
[2] Matthew C. Fontaine and Stefanos Nikolaidis. 2022. Covariance Matrix Adaptation MAP-Annealing. https://doi.org/10.48550/ARXIV.2205.10752
[3] Peter K. Friz and Nicolas B. Victoir. 2010. *Multidimensional Stochastic Processes as Rough Paths.* Cambridge University Press.
[4] Luca Grillotti and Antoine Cully. 2022. Unsupervised Behavior Discovery With Quality-Diversity Optimization. *IEEE Transactions on Evolutionary Computation.*
[5] Lévy Lyons, Caruana. 2007. *Differential Equations Driven by Rough Paths.* Springer.
[6] Sebastian Mayer, Léo Françoso Dal Piccol Sotto, and Jochen Garcke. 2022. The Elements of Flexibility for Task-Performing Systems. *IEEE Access* 11 (2022), 8029–8056.
[7] James Morrill, Adeline Fermanian, Patrick Kidger, and Terry Lyons. 2020. A Generalised Signature Method for Multivariate Time Series Feature Extraction. *CoRR* abs/2006.00873 (2020). arXiv:2006.00873v2 [cs.LG]
[8] Jean-Baptiste Mouret and Jeff Clune. 2015. Illuminating search spaces by mapping elites. arXiv:1504.04909
[9] Hao Ni, Lukasz Szpruch, Magnus Wiese, Shujian Liao, and Baoren Xiao. 2020. Conditional Sig-Wasserstein GANs for Time Series Generation. (2020). arXiv:2006.05421
[10] Justin K. Pugh, L. B. Soros, Paul A. Szerlip, and Kenneth O. Stanley. 2015. Confronting the Challenge of Quality Diversity. *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation* (2015).
[11] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research* 22, 268 (2021), 1–8.
[12] Jeremy Reizenstein and Benjamin Graham. 2020. Algorithm 1004: The iisignature Library: Efficient Calculation of Iterated-Integral Signatures and Log Signatures. *ACM Transactions on Mathematical Software (TOMS)* (2020).
[13] Bryon Tjanaka, Matthew C Fontaine, David H Lee, Yulun Zhang, Nivedit Reddy Balam, Nathaniel Dennler, Sujay S Garlanka, Nikitas Dimitri Klapsis, and Stefanos Nikolaidis. 2023. Pyribs: A Bare-Bones Python Library for Quality Diversity Optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference.* 220–229.
[14] Mark Towers, Jordan K. Terry, Ariel Kwiatkowski, John U. Balis, Gianluca de Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Arjun KG, Markus Krimmel, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Andrew Tan Jin Shen, and Omar G. Younis. 2023. Gymnasium.
[15] Vassilis Vassiliades, Konstantinos Chatzilygeroudis, and Jean-Baptiste Mouret. 2016. Using Centroidal Voronoi Tessellations to Scale Up the Multidimensional Archive of Phenotypic Elites Algorithm. *IEEE Transactions on Evolutionary Computation* 22 (2016), 623–630.