

Surrogate Infeasible Fitness Acquirement FI-2Pop for Procedural Content Generation

Roberto Gallotta
Araya Inc.
Tokyo, Japan

Kai Arulkumaran
Araya Inc.
Tokyo, Japan

L. B. Soros
Cross Labs
Kyoto, Japan

Abstract—When generating content for video games using procedural content generation (PCG), the goal is to create functional assets of high quality. Prior work has commonly leveraged the feasible-infeasible two-population (FI-2Pop) constrained optimisation algorithm for PCG, sometimes in combination with the multi-dimensional archive of phenotypic-elites (MAP-Elites) algorithm for finding a set of diverse solutions. However, the fitness function for the infeasible population only takes into account the number of constraints violated. In this paper we present a variant of FI-2Pop in which a surrogate model is trained to predict the fitness of feasible children from infeasible parents, weighted by the probability of producing feasible children. This drives selection towards higher-fitness, feasible solutions. We demonstrate our method on the task of generating spaceships for Space Engineers, showing improvements over both standard FI-2Pop, and the more recent multi-emitter constrained MAP-Elites algorithm.

Index Terms—Procedural Content Generation, Constrained Optimisation, Evolutionary Algorithms, Quality-Diversity

I. INTRODUCTION

Designing content is still one of the most time-consuming steps in developing video games. Many games nowadays include some kind of procedural content generation (PCG) to reduce the time spent in developing content, be it levels, interactive items or even entire quest lines [1]. One usual PCG approach is to search within a space of possible content for solutions with good quality. This search process can be seen as an optimisation problem, where the quality of the content is defined by an utility function. In some cases there are restrictions imposed upon the content, which can be expressed as constraints in the optimisation problem. The feasible-infeasible two-population (FI-2Pop) genetic algorithm (GA) [2] has been widely used in constrained optimisation tasks due to its simplicity and intuitiveness. FI-2Pop has been used for PCG extensively, as it allows the programmer to specify domain-specific constraints and fitness functions [3].

FI-2Pop is able to freely explore the solution space in its entirety, regardless of feasibility (constraint satisfaction), so that feasible solutions that would otherwise be unreachable due to the feasibility region’s topology can be reached via infeasible solutions. In the original algorithm, infeasible solutions are selected for evolution in inverse proportion to the number of constraints broken, but this selection mechanism does not directly optimise for feasible solution fitness. In this work we propose a novel variant of FI-2Pop, training a surrogate

fitness function that predicts the fitness of infeasible solutions’ offspring, weighted by the probability of generating a feasible child, which can accelerate the optimisation process. We apply our algorithm to PCG for creating novel spaceships in the video game Space Engineers, improving upon prior work based on standard FI-2Pop [4]. We then show our offspring fitness prediction method’s effectiveness in a mixed-initiative setting, solving the spaceship generation task with the constrained multi-dimensional archive of phenotypic-elites (CMAP-Elites) algorithm [5]. Finally, we extend the latter application to include an optimising emitter [6], which uses our novel infeasible fitness function to balance both fitness and coverage of solutions.

II. BACKGROUND

A. Procedural Content Generation

PCG is the process of algorithmically creating components for an application, either independently or in tandem with a human designer. PCG has been widely used in video games, creating everything from assets to entire levels [1].

Evolutionary algorithms (EAs; including GAs), are a family of population-based search algorithms well-suited for PCG: by assigning a fitness value to solutions, EAs can be used to search for desirable content. To do so while retaining a wide set of solutions to be chosen from, one can use EA-based “quality-diversity” (QD) algorithms, e.g., MAP-Elites [7], [8].

B. FI-2Pop

The FI-2Pop GA [2], [9] evolves two separate populations of solutions to follow a fitness signal in a constrained domain. One population is comprised solely of feasible solutions and the other solely of infeasible solutions. Each feasible solution has no constraint violations and its fitness is determined by a domain-specific utility function, whereas each infeasible solution violates at least one constraint and its fitness is the inverse of the number of constraints violated.

The GA loop of FI-2Pop is as follows: one or more solutions are picked from each population (based on their fitness) as parents to generate new offspring solutions via crossover and mutation. Each parent may generate either a feasible or infeasible offspring, regardless of its own feasibility. All offspring are then added back to the populations based on their feasibility. If the size of a population exceeds a set limit,

solutions are removed from it based on their fitness. This loop is then repeated for a given number of generations.

Selecting infeasible solutions inversely to the number of constraints violated makes selection pressure towards feasibility a generic repair mechanism. While simple and effective, this heuristic could be replaced by a more information-rich function for optimisation, as we propose in section IV.

C. Surrogate-Assisted Evolutionary Algorithms

Surrogate-assisted EAs (SAEAs) [10] use a surrogate model to approximate the fitness function in computationally-expensive problems. SAEAs use a surrogate model that is trained on the fitness of solutions collected from previous runs. The model then interacts with the EA by estimating the fitness of newly-generated solutions. SAEAs typically include a re-evaluation step where some or all solutions in the current generation are evaluated with the true fitness function, to be used in subsequent training of the model. This online training helps ensure convergence of the EA by improving the modelling of relevant areas of the search space.

There are many classes of fitness models; in our work we use a regression-based absolute fitness model, where the estimator takes features of the solution and is trained to predict the fitness directly [10]. However, in contrast to most uses of SAEAs, we use our model to create a different fitness function than the original algorithm’s (FI-2Pop’s). Our proposed fitness function is similar to “fitness inheritance” [11], [12], where the estimation is based on the parents’ fitness, although we assign fitness based on the fitness of the parents’ *prior offspring*. We call this novel estimation “acquirement”, and give more details in section IV.

D. MAP-Elites

MAP-Elites [7] is an “illumination algorithm” that searches for a diverse set of high-performing solutions, as defined by behaviour characteristics (BCs) of interest. MAP-Elites projects and subdivides the search space into BC-based niche bins, each containing one or more solutions. In standard MAP-Elites, the underlying search process is unguided and unbiased: at each step, a random bin is selected amongst all non-empty bins with uniform probability. New solutions are then generated from the selected bin and then assigned to their respective bins. Bins have a fixed capacity, so new solutions can take the place of older or less-fit solutions.

MAP-Elites requires an optimisation algorithm in order to produce new solutions, and EAs are typically used for this in practice. For PCG with constraints, the most common choice is FI-2Pop. There are multiple variations on MAP-Elites, such as letting a bin contain more than one solution (elite), or allowing near-bin crossover. Other variations have focused on the bin selection process itself: *emitters* [6] can be used to more directly optimise for average fitness or coverage. Recent work [13] extended this to a meta-optimisation process, whereby a bandit algorithm selects between different types of emitters to optimise for given metrics. MAP-Elites has been augmented with surrogate models before, using

Gaussian processes [14] and neural networks [15]; however, these simply model the original fitness function instead of replacing it. In Section V we demonstrate FI-2Pop with our novel fitness function extended to the CMAP-Elites setting, and with the use of emitters.

III. DOMAIN

Space Engineers is a 3D sandbox game with realistic graphics and physics, played by approximately 5,000 players daily, where the objective is to build spaceships and other structures to collect resources, navigate through space, and interact with other characters. Structures are formed out of interacting blocks, each with their own properties and functionalities. In order to be usable, these structures must satisfy various functional constraints. For example, a spaceship needs a cockpit, an engine, and thrusters.

Procedurally generating structures for Space Engineers involves creating solutions that both respect constraints defined by the game itself, as well as any additional, hand-engineered hard or soft constraints, whilst simultaneously striving to find aesthetically-pleasing results. We follow previous work [4], which introduced a hybrid EA consisting of multiple modular L-systems [16] and FI-2Pop, to combine domain knowledge with constrained optimisation in order to procedurally generate spaceships. We use the same feasible fitness function and constraints. The hard constraints are, firstly, no intersections between blocks, and secondly, the structure must contain all required blocks. The soft constraint is to have symmetry along any axis. The feasible fitness function is based on a probability density estimate of various metrics, derived from human-designed spaceships made available on Steam Workshop. The 4 metrics used are the ratio between the number of functional blocks and the total number of blocks, the ratio between the filled and total volume, the ratio between the major and medium axis of the ship, and the ratio between the major and minor axis.

IV. SURROGATE INFEASIBLE FITNESS ACQUIREMENT

As discussed previously, the infeasible fitness for standard FI-2Pop selects infeasible parents likely to result in feasible offspring (by virtue of having relatively few constraint violations), but this fitness calculation does not account for offspring fitness. We propose using a surrogate model to estimate the fitness of infeasible parents’ children, weighted by the probability of the children being feasible, driving selection towards high-performing, feasible solutions.¹ This can be seen as encouraging evolvability [17] in the infeasible population, where parents are chosen directly based on their probability of generating children with higher fitnesses.

The main FI-2Pop loop involves selecting parents to create children using crossover and mutation, after which children are placed into their respective populations. Selection and evolution happen separately for the feasible and infeasible populations. In our variant, we keep a buffer for each

¹A heuristic analogous to a look-ahead search with depth 1.

infeasible parent that contains the fitnesses of any feasible children it produces; we also keep a tally of the empirical probability of the infeasible parent producing a feasible child. We then train a surrogate model—in our experiments, a fully-connected neural network—to directly predict a statistic of the children’s fitness from the phenotypic representation of the parent, weighted by the sample probability of producing a feasible child, using the mean-squared objective. The estimated fitness is therefore $S_{c_f}[F(c_f)] \cdot \frac{n_f}{n_{all}}$, where S is a statistic (e.g., mean) calculated over all feasible children c_f , $F(c_f)$ is the fitness of the feasible children, and n_f and n_{all} are the number of feasible and total offspring, respectively.

In preliminary experiments, we tested different statistics of the children’s fitness, including the mean (denoted “ μ ”), the maximum (denoted “M”), the minimum (denoted “m”), as well as the upper and lower confidence bounds (which have been used as selection heuristics in SAEAs [18]). We found that the confidence bounds have properties similar to the maximum and minimum, so we retained only the first 3 statistics for the rest of our experiments. At initialisation, the fitness of all infeasible solutions is set to a small, positive value ϵ , causing the algorithm to select infeasible solutions uniformly. The surrogate model is trained as soon as one datapoint is available, and is updated online whenever new data is available. Whenever the model is updated, we use it to reassess fitnesses for recently-generated infeasible solutions.

FI-2Pop has previously been combined with MAP-Elites to form CMAP-Elites [5], a quality-diversity algorithm that can find solutions that satisfy given constraints. In our current work, we demonstrate our novel variant of FI-2Pop within CMAP-Elites. The most benefit comes from using an optimising emitter (denoted “E”), which selects bins with the highest fitness per population. As the optimum choice of statistic and/or emitter may change per domain, or even over time, we use the ϵ -greedy bandit algorithm [19] (denoted “B”) to select both in order to maximise fitness and coverage.

V. RESULTS

We demonstrate the performance of our proposed FI-2Pop variant over 3 experiments using the previously proposed hybrid EA in the Space Engineers domain [4]: FI-2Pop, CMAP-Elites (which contains FI-2Pop), and CMAP-Elites with emitters. For all experiments we ran each method for 50 generations, and provide average results with standard error (SE), collected over 50 random seeds.

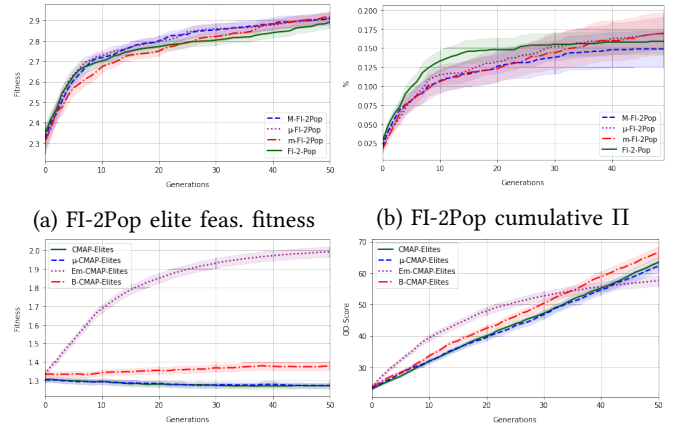
In our first experiment, we compared the standard FI-2Pop algorithm and our variant with different statistics. All statistics result in a higher elite fitness (Figure 1a), which is the final solution given by FI-2Pop. We also tracked the percentage of offspring with infeasible parents that are feasible (denoted Π), introduced at each generation. We plot the cumulative sum of this metric in Figure 1b, where it demonstrates that our variant can introduce more feasible solutions with infeasible parents, as compared to standard FI-2Pop which plateaus after 30 generations.² Finally, we also

²The exception is M-FI-2Pop, where the fitness may be too optimistic.

tracked the estimator loss. Although the loss jumps when new datapoints are added, it otherwise steadily decreases over time, indicating that it continues to capture the current data distribution.

In our second experiment, we tested CMAP-Elites with the standard FI-2Pop GA and our variants. We set CMAP-Elites to operate on a fixed 32×32 grid, using axis ratios (Section III) as the BCs.³ Under this setting, we found that our variants perform similarly to CMAP-Elites.

In our third experiment, we replaced the standard MAP-Elites selection rule (a “random emitter”) with the optimising emitter (E). We found that this resulted in a much higher feasible fitness overall, at the cost of a lower coverage. To compensate for this, we used an ϵ -greedy bandit (B) to optimise the percentage increase of feasible fitness plus the percentage increase of coverage. This setup produced high-fitness solutions (higher than standard CMAP-Elites, lower than E) at a high coverage (comparable to standard CMAP-Elites, higher than E). We show some of the metrics of the most salient experiments in Figures 1c and 1d.



(c) CMAP-Elites avg. feas. fitness (d) CMAP-Elites feas. QD-score
 Fig. 1: Different metrics for FI-2Pop and CMAP-Elites (mean \pm SE) for the feasible population.

We report final metrics for all experiments in Table I. From these results, we demonstrate that our variant can find solutions at a higher fitness, as compared to the traditional FI-2Pop algorithm. A similar conclusion can be drawn for the application of our variant to CMAP-Elites, when combined with the optimising emitter. Additionally, we note that using a bandit to select the emitter and statistic allows us to find an advantageous tradeoff between feasible fitness and coverage, and can also be adapted to optimise different metrics, or combinations thereof. We quantify this tradeoff via the QD-score [20], which is calculated as the sum of elite fitnesses in all archive bins (with zero fitness if the bin is empty). Our B-CMAP-Elites achieves the highest QD-score overall.

Finally, we give a qualitative overview of the spaceships generated by the different approaches. We found that our

³We note that this choice of BCs restricts the maximum coverage to 50%.

Method	Feas. fitness		Infeas. fitness		Coverage		QD-score (Feas.)	Π
	Top	Avg	Top	Avg	Feas.	Infeas.		
FI-2Pop	2.8915	2.7777	-	-	N/A	N/A	N/A	0.0032
M-FI-2Pop	2.9088	2.8074	0.1610	0.1554	N/A	N/A	N/A	0.0030
μ -FI-2Pop	2.9103	2.7949	0.1677	0.1639	N/A	N/A	N/A	0.0034
m-FI-2Pop	2.9166	2.7867	0.0977	0.0953	N/A	N/A	N/A	0.0034
CMAP-Elites	2.4637	1.2730	-	-	0.0477	0.0420	63.4168	0.0179
M-CMAP-Elites	2.3903	1.2699	0.0857	0.0443	0.0482	0.0425	64.0300	0.0154
μ -CMAP-Elites	2.4851	1.2737	0.1115	0.0533	0.0462	0.0401	62.2313	0.0138
m-CMAP-Elites	2.4221	1.2581	0.0713	0.0375	0.0480	0.0421	63.8382	0.0192
EM-CMAP-Elites	2.6915	1.9298	0.0238	0.0138	0.0327	0.03290	57.0036	0.0080
E_{μ} -CMAP-Elites	2.6359	1.9125	0.0326	0.0144	0.0342	0.0331	58.784	0.0079
Em-CMAP-Elites	2.7284	1.9925	0.0604	0.0392	0.0323	0.0328	57.6187	0.0085
B-CMAP-Elites	2.4860	1.3792	N/A	N/A	0.0471	0.0431	66.5877	0.0207

TABLE I: Final metrics for all experiments. **Bold** values indicate the best result within each setting.

FI-2Pop variant is able to generate more complex solutions than those produced by standard FI-2Pop (Figure 2b versus Figure 2a). Spaceships generated by standard CMAP-Elites are relatively simple (Figures 2c and 2d), whilst our variant with the optimising emitter can find more interesting solutions (Figures 2e and 2f), while retaining comparable diversity if used in combination with the bandit algorithm.

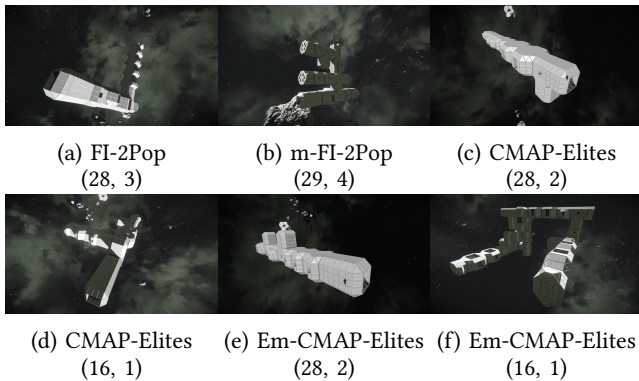


Fig. 2: In-game screenshots of elite solutions with comparable BCs (major/middle/minor axes ratios, given in parentheses) from different experiments.

VI. DISCUSSIONS

In this work we presented a variant of FI-2Pop that uses a surrogate model to compute a different fitness for the infeasible population, based on the ability of an infeasible solution to generate high-performing feasible solutions, more directly targeting final (elite) fitness. We also tested CMAP-Elites with both standard FI-2Pop and our variant. When combined with an optimising emitter, our method outperforms standard CMAP-Elites. Finally, we showed how adding a simple bandit algorithm to automatically choose between emitters and metrics can help our variant achieve both high fitness and coverage of the search space.

There are several considerations to using our method. Firstly, training the surrogate model requires additional time and resources. However, the tradeoff would be favourable for domains where evaluating the feasible fitness is expensive. Secondly, our choice of fitness function performs better in domains where population hopping (in particular, when infeasible parents produce feasible offspring) is frequent,

as more data is available to train the model. Finally, our method does not explicitly take into account the number of constraints violated, which may be a more important signal in other domains. Nevertheless, the concept of using a surrogate model to construct a more informative fitness function for FI-2Pop’s infeasible population has potential benefits for PCG, and even wider applications requiring constrained optimisation.

ACKNOWLEDGEMENTS

This project was partly funded by a GoodAI research grant.

REFERENCES

- [1] M. Hendriks, S. Meijer, J. Van Der Velden, and A. Iosup, “Procedural content generation for games: A survey,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, pp. 1:1–1:22, 2013.
- [2] S. O. Kimbrough, G. J. Koehler, M. Lu, and D. H. Wood, “On a Feasible–Infeasible Two-Population (FI-2Pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch,” *EJOR*, vol. 190, no. 2, pp. 310–327, 2008.
- [3] A. Baldwin, S. Dahlskog, J. M. Font, and J. Holmberg, “Mixed-initiative procedural generation of dungeons using game design patterns,” in *IEEE CIG*, pp. 25–32, 2017.
- [4] R. Gallotta, K. Arulkumar, and L. B. Soros, “Evolving Spaceships with a Hybrid L-system Constrained Optimisation Evolutionary Algorithm,” in *GECCO Companion*, 2022.
- [5] A. Khalifa, S. Lee, A. Nealen, and J. Togelius, “Talakat: bullet hell generation through constrained map-elites,” in *GECCO*, pp. 1047–1054, ACM, 2018.
- [6] M. C. Fontaine, J. Togelius, S. Nikolaidis, and A. K. Hoover, “Covariance matrix adaptation for the rapid illumination of behavior space,” in *GECCO*, pp. 94–102, ACM, 2020.
- [7] J.-B. Mouret and J. Clune, “Illuminating search spaces by mapping elites,” *arXiv:1504.04909*, 2015.
- [8] D. Gravina, A. Khalifa, A. Liapis, J. Togelius, and G. N. Yannakakis, “Procedural Content Generation through Quality Diversity,” in *CoG*, 2019.
- [9] “Introducing Distance Tracing of Evolutionary Dynamics in a Feasible-Infeasible Two-Population (FI-2Pop) Genetic Algorithm for Constrained Optimization,” 2004.
- [10] H. Tong, C. Huang, L. L. Minku, and X. Yao, “Surrogate models in evolutionary single-objective optimization: A new taxonomy and experimental study,” *Information Sciences*, vol. 562, 2021.
- [11] R. E. Smith, B. A. Dike, and S. A. Stegmann, “Fitness inheritance in genetic algorithms,” in *SAC*, pp. 345–350, ACM, 1995.
- [12] K. Chellapilla and D. B. Fogel, “Fitness distributions in evolutionary computation: motivation and examples in the continuous domain,” *Biosystems*, vol. 54, pp. 15–29, 1999.
- [13] A. Cully, “Multi-emitter MAP-elites: Improving quality, diversity and data efficiency with heterogeneous sets of emitters,” in *GECCO*, pp. 84–92, ACM, 2021.
- [14] A. Gaier, A. Asteroth, and J.-B. Mouret, “Data-Efficient Exploration, Optimization, and Modeling of Diverse Designs through Surrogate-Assisted Illumination,” *arXiv:1702.03713*, 2017.
- [15] Y. Zhang, M. C. Fontaine, A. K. Hoover, and S. Nikolaidis, “DSA-ME: Deep surrogate assisted MAP-elites,” in *GECCO*, 2022.
- [16] A. Lindenmayer, “Mathematical models for cellular interactions in development I. Filaments with one-sided inputs,” *J. Theor. Biol.*, vol. 18, no. 3, pp. 280–299, 1968.
- [17] L. Altenberg, “The Evolution of Evolvability in Genetic Programming,” in *Advances in Genetic Programming*, Complex Adaptive Systems, pp. 47–74, M.I.T. Press, 1994.
- [18] B. Liu, Q. Zhang, and G. G. E. Gielen, “A Gaussian Process Surrogate Model Assisted Evolutionary Algorithm for Medium Scale Expensive Optimization Problems,” *IEEE TEVC*, vol. 18, no. 2, pp. 180–192, 2014.
- [19] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. The MIT Press, 1998.
- [20] J. K. Pugh, L. B. Soros, and K. O. Stanley, “Quality Diversity: A New Frontier for Evolutionary Computation,” *Front. Robot. AI*, vol. 3, 2016.