K Best Candidates in R-NSGA-III Niche Selection Operator

Sina O. Azartash SAZARTA1@ALUMNI.JH.EDU
A Passion Project

Abstract

We propose the algorithm KBC-R-NSGA-III inspired from the state-of-the-art R-NSGA-III algorithm and evaluate its ability to increase diversity and improve front approximation accuracy. R-NSGA-III is an extension of NSGA-III that uses preference information to guide the evolutionary process and approximates partial sets of the Pareto Optimal Front(POF). Typically in Multi-Objective Optimization(MOO), preferences are incorporated after approximating the POF by selecting amongst a list of Pareto Optimal(PO) solutions. Conversely, R-NSGA-III takes preferences first, and then attempts to find PO solutions that accord with those preferences. The idea is to avoid unnecessary calculation and to generate an increased number of PO points that are diverse and relevant to the specified user interest. Currently, R-NSGA-III has been demonstrated to successfully generate optimal, diverse points in a preferred space, but much additional work remains to empirically assess how it preforms and reacts to different situations. Therefore an additional contribution of our paper is to add to the empirical assessment of R-NSGA-III. Specifically, we refine our focus to evaluate the degree of dependence on the number of objective functions, the number of variables, and data-sets characteristics of MOO standardized test problems. The idea behind KBC-R-NSGA-III is to modify the niche selection operator of R-NSGA-III so that front members are accepted based on stochastic approach rather than deterministic. Our results showed that KBC-R-NSGA-III was unsuccessful in improving accuracy or diversity in the majority of test cases, but was did display superior performance in situations with multiple local optima.

1. Purpose

This passion project serves as an example of my writing ability for AI research. This project started initially as a requirement for the Advanced Machine Learning course at Johns Hopkins University by Dr. John Wilbur Sheppard Spring 2022. Completing the project inspired many further ideas and I became more excited about the field of Multi-Objective Optimization so I greatly expanded on this project. Since then, I have clarified my points, added additional ideas, conducted further experimentation, added extra analysis, and presented more visualizations.

Skills this project is meant to display:

1. Ability to evaluate Machine Learning Research Journal articles
2. Ability to understand state-of-the-art algorithms
3. Ability to code using Latex with Journal of Machine Learning Research Format
4. Ability to modify research code and prototype a new solution
5. Ability to design, conduct, and analyze an experiment

6. Highly specialized knowledge in the field of Multi-Objective Optimization

2. Introduction

Real world decision making situations with multiple stakeholders and conflicting priorities often use Multi-Objective Optimization (MOO) to generate a pool of optimal trade-off solutions. The stakeholders, priorities and constraints of the problem are represented as several conflicting mathematical functions. MOO attempts to maximize the desired value of these functions, but maximizing one function, may reduce the optimality of another. To address this, MOO aims to approximate the Pareto Optimal Front (POF) which is the region of decision space that is both feasible and maximizes the set of all functions. Evolutionary Multi-objective Optimization (EMO) is such an approach that in short, stochastically improves its approximation of the POF over multiple runs. Evolutionary Many-Objective Optimization (EMaO) is then used when the number of objectives is greater than three.

Decision makers should expect that the approximated POF adequately encompasses the breadth of optimal solutions. An abundance of trade-off points in some regions of the decision space and a deficit in others would be improper representation. EMaO algorithms address this by using diversity metrics which attempt to generate a more uniform distribution of points Zhang and Li (2007). One particular strategy is decomposition-based EMaO methods such as MOEA/D, NSGA-II, and NSGA-III which use reference directions to guide the search Deb and Jain (2014); Zhang and Li (2007); Deb et al. (2002a). Multiple reference directions are predefined on a hyperplane representing the objective space. The search is guided by emphasizing trade-off points closer to each reference direction. The idea is that uniformly spaced reference directions on the hyperplane generally lead to a set of better distributed trade-off solutions on the POF.

Once the optimal solutions approximations are obtained, decision makers must then select their desired solutions using preference functions. Often, the preferred solution is a small subset of many optimal solutions. This points to the focus of our paper because in several instances, much computation could have been avoided if non-preferred candidate solutions were not even considered to start with. In our paper, we investigate Reference point Non-Dominated Sorting Genetic Algorithm-III (R-NSGA-III) to solve this problem. RNSGA-III works by reducing the number of trade-off candidate points considered by setting reference search directions in a region of predefined preferred space Vesikar et al. (2018). In our paper, we evaluate the quality of solutions provided by R-NSGA-III and study the effect of modification.

Currently,R-NSGA-III has been shown to successfully approximate optimal solutions in a region of preferred space. R-NSGA-III was introduced in 2018 in the journal article "Reference Point Based NSGA-III for Preferred Solutions" from Kalyanmoy Deb's COIN-laboratory Deb and Jain (2014). In that study, R-NSGA-III was compared to its predecessor R-NSGA-II Deb and Sundar (2006) which showed improved performance on test problems with more than three objectives.

However, many problems remain to be addressed. Specifically, the R-NSGA-III displays a problem with generating diverse points and has a reduced ability to generate PO points. Our proposed solution is K - Best Neighbors - Reference point Non-Dominated Sorting Genetic Algorithm-III (KBC- R-NSGA-III) where we modify the niche selection operator

(NSO) that is responsible for deciding which individuals get emphasized in the search space. The main idea is that NSO emphasizes a best individual by looking a proximity to a reference point. We instead allow multiple K individuals along with the closest individual to be emphasized. We state the the best individual front member is not always the closest member.

Moreover, R-NSGA-III is still very novel, and much information about its capabilities and reactions to different problems is limited. Since we are expanding on the R-NSGA-III method, we need to have a better idea of how it empirically preforms. Therefore, our second contribution is to add to the empirical assessment of R-NSGA-III. We accomplish this by evaluating R-NSGA-III with metrics that measure the accuracy of the approximated front on standard tests problems. We also aims to learn three critical relationships: the dependency of R-NSGA-3 and KBC-RNSGA on the number of objective functions, the number of variables, and various data-sets characteristics.

We expect that KBC-RNSGA will allow more diversity in each front and slightly also improve the accuracy of the approximated POF because the increased diversity will help avoid local optima and encourage more a more various pool of genes in offspring. We compare against compare R-NSGA-III against NSGA-III to accomplish this. Our main hypothesis is that KBC-RNSGA will achieve a statistically significant increase in diversity at least 80% of the time. Our ancillary hypothesis is that KBC-RNSGA, RNSGA-III and NSGA-III all will experience a decrease in approximation accuracy of the POF on more objectives, more data features, and more complex problem sets statistically significantly at least 95% of the time.

The remainder of our paper is organized as follows. In Section 2, we discuss the background and works related to R-NSGA-III. Then in Section 3, we describe our approach and present KBC-RNSGA-III. Next, in Section 4 we present our experimental design and explain how and why it evaluates the relevant aspects of R-NSGA-III. In Section 5, we detail our findings, summarize our results, and highlight major trends. Then, we analyze these trends, explain their significance, explore their ramifications, and address their limitations in Section 6. Finally in Section 7, we draw our conclusions, discuss the importance of our work, and suggest relevant future work.

3. Background and Related Work

R-NSGA-III is an EMaO algorithm that uses preferences to initialize and update reference directions Vesikar et al. (2018). The reference direction then guides the search process by emphasizing reproduction of trade-off solutions closer to a reference point. Both the preference information and the reference directions are represented as a vector that represents a point on the hyperplane representing the objective space. A single preference vector is referred to as an aspiration point. The difference between aspiration points and reference vectors is that aspiration points point to the preferred space, while the reference directions are constructed uniformly around the aspiration point. The background theory and previous work done related to the algorithm is discussed below in the following subsections.

3.1 Pareto Optimality

EMaO algorithms and MOO in general find sets of points that optimize the set of all objective functions in the feasible decision space. An objective function here represents a goal, priority or circumstance in which its maximum or minimum represents the most optimal outcome. A single optimal point produced by MOO is a vector that contains an argument to each objective function. The vector contains sub-optimal input values to each objective function, but the combination of values in the vector represents an optimal input value for the collection of all objective functions. Therefore, a single optimal point can be said to be an optimal trade-off solution that yields the most optimal value for several conflicting objectives. Another way of looking at this is that the optimal point represents a hyper-dimensional coordinate in the decision space. The decision space, denoted by Ω, is the domain of the objective function, denoted as x. The objective space is the co-domain denoted by Rm where m is the number of objectives. Rm is a vector in which each real-valued objective function is denoted as F(x). The attainable objective set is the range $\{Rm(x)|x\epsilon\Omega\}$ which represents feasible decisions.

The term optimal refers to a trade-off point that dominates all other trade-off points except points that are equally as dominant. More formally, consider two trade-off points which are part of the decision space Ω. u dominates v if and only if u $\geq$ v for all objective functions and u v for at least one objective function. Simply put, u dominating v means u is a better solution. If there is no other point in Ω that dominates u, then u is a Pareto Optimal (PO) point. This indicates that if any objective in a PO point is further improved, at least one other objective must deteriorate. Finally, the set of all PO points is referred to as the Pareto Set (PS) which resides on a hyper-surface in the objective space referred to as the Pareto Optimal Front(POF). MOO aims to approximate the POF.

3.2 History

A history of MOO and methods leading to NSGA-III are briefly described Srinivas and Deb (1994). One of the earlier attempts at optimizing several competing objectives was the Weighted Sum Method (WSM). WSM assigns a weight to each objective then tries to find the maximum or minimum sum of all objectives. The disadvantage to WSM is deciding the weights is subjective and difficult, and WSM only provides one optimal solution. Evolutionary algorithms improved on this by using Genetic Algorithms(GA). Briefly, trade-off solutions are represented as an individual that can produce offspring. Individuals that dominate more points are assigned a higher fitness score. Individuals with a higher fitness score then have the highest chance of reproduction. Reproduction combines the vectors of the trade-off solutions to stochastically produce offspring with even higher fitness values. This process of creating individuals with higher fitness values continues iteratively until convergence or termination. Non-dominated Sorting Genetic Algorithm (NSGA-I) was one algorithm that used the GA approach and ranked individuals' fitness scores based on comparing their domination levels using non-dominated sorting Srinivas and Deb (1994). This solved the WSM problem of only single solution per run as NSGA-I could find multiple Pareto optimal solutions in a single run. Furthermore, solutions had their diversity preserved indicating a good spread on the POF. This was done by using a niche formation

technique that counted the number of points in the population within a certain distance from an individual.

3.3 NSGA-II

NSGA-II improves on NSGA-I by reducing time complexity, introducing elitism which improves convergence, and further improving diversity with crowding distance Deb et al. (2002a). Elitism in genetic algorithms refers to ensuring that most fit candidate solutions are carried over to the next generation of solutions. Time complexity of NSGA-II is $O(M*N^3)$ where M is the number of objectives and N is the population size while NSGA-III is $O(M*N^2)$.

The NSGA-II algorithm is summarized Deb et al. (2002a). The process begins with an equal sized parent population and produced offspring population. Using Non-Dominated Sorting, every individual in the population of offspring and parents are compared with every other individual pairwise to check for domination. Each individual contains a list on how many times it has been dominated by another individual and also which individual it dominates. The individuals not dominated by any individuals form the first front 'F1'. The remaining individuals have their dominated count reduced by the number of individuals in F1. In this way, the individuals in F2 now have a dominated count of zero. This process repeats until every individual gets a rank based on the front level. Next, the best fronts are taken to the next generation. To keep a good spread and avoid local optima, the crowding distance decides which individuals are selected to become the next parent population. Individuals with higher crowding distance are added first while the remaining individuals are discarded. The parent population now produces offspring using tournament selection, crossover and then mutation. Tournament selection randomly compares two parents based on their rank. The higher rank gets to reproduce. If there is a tie, the higher crowding distance wins. Crossover randomly selects half of the attributes of an individual and swaps those attributes with another individual. Mutation is the percent chance that an individual attribute is modified. Lastly, this entire process is repeated until a select number of generations or until the fitness fitness values of the population converges.

A problem with NSGA-II is that an increasing number of objective functions causes the algorithm to become computationally intractableDeb and Jain (2014). A linear increase in the number of objectives causes an exponential increase in the proportion of individuals in the population that are non-dominated. In order to continue satisfying elitism, this greatly increases the number of individuals per generation to allow enough room for new high fitness individuals. The consequential number of individuals evaluated slows down non-dominated sorting and also slows down diversity metrics which requires comparing neighboring individuals. An additional challenge is that the hyper-dimensional space becomes very large and individuals can be very spread out in the decision space. This can make reproduction less reliable. Parents can contain very different attributes from each other which results in producing offspring that less reliably has higher fitness scores and is very different from each parent.

3.4 NSGA-III

NSGA-III works similarly to NSGA-II except that it addresses the aforementioned challenges by introducing multiple predefined reference vectors and replacing the crowding distance operator with niche preservation association operator. Compared to NSGA-II, this results in reduced computation and increased diversity when dealing with 3 to 15 objectives. Reference vectors enable focusing on targeted partial sets of the attainable objective hyper-dimensional space. The attainable objective space search is guided in parallel to each reference vector. Individuals closer to a reference point are emphasized for reproduction over other individuals. This lowers the number of candidate solutions considered in each generation and helps enable elitist selection. Also, since the reference points are predefined to be widely distributed on hyperplane, the obtained solutions are more likely widely distributed on the approximated POF and therefore have a higher diversity. Moreover, it is important to note that setting reference vectors requires normalizing the hyperplane. Since the approximation of the POF improves each generation, the hyperplane has to be normalized again each generation. The resulting normalized attainable objective hyper-dimensional space is referred to as the unit simplex.

The niche preservation association operator (niche operator) decides which individuals are emphasized, and it is the basis for our algorithm modification. The niche operator works by results in discretizing distribution of points on the objective space. The end goal is to have each reference vector associated with at least one PO point. The niche operator works by associating the individual(s) with the shortest perpendicular distance to the reference vector. During the calculation of the fronts, each reference point always has at least one trade-off point associated with it at each generation. Therefore, the individual with the highest chance of reproduction is the point with the highest domination, but also the closest to the reference vector. Lower ranking dominated individuals can still be selected for reproduction if there are no other dominating points for that specific reference vector. Finally, in the event that a reference vector is not associated with an individual, the reference vector is eliminated for further consideration in the algorithm. This reduces computation by avoiding searches in fruitless directions.

A popular and effective procedure for generating well spaced reference vectors on unit simplex in NSGA-III and R-NSGA-III is the Normal-Boundary Intersection (NBI) method, also referred to as the Das and Dennis approach Das and Dennis (1998). Reference vectors are constructed by originating from the origin and display identical distance to their nearest neighbor on the unit simplex. Each reference vector's objectives must be positive and their sum must equal to one Blank et al. (2021). The number of reference vectors, denoted by H, is determined by the number of objectives, denoted by M, and deciding how many partitions should discretize the search space, denoted by p.

4. Approach

4.1 Standard Test Problems With Known Pareto Front

Our approach uses standardized test problems in which the POF is already known to better focus on the performance of R-NSGA-III rather than its capability. Specifically we focus on the DTLZDeb et al. (2002b) and WFG Huband et al. (2006) many objective standard-

ized problem sets. Studies on NSGA-3 frequently use DTLZ2, DTLZ3, DTLZ4, WFG5, and WFG6 test problems and therefore we follow suit. We summarized the reasoning for selecting the DTLZ problems below Blank and Deb (2020).

1. DTLZ1: investigates difficulty convergence to local POF
2. DTLZ2: investigates scaled performance with increasing number of objectives
3. DTLZ3: investigates ability to converged to one global POF, but many local POFs
4. DTLZ4: investigates ability to generate a wide distribution of solutions
5. DTLZ5: investigates convergence to a simpler curve, used for visualization
6. DTLZ6: known to be difficult for MOO problems including NSGA-II
7. DTLZ7: investigates ability to maintain sub-populations, disconnected regions.

DTLZ 5,6, and 7 are not relevant. Specifically, we avoid DTLZ5 because we seek challenging problems for evaluating R-NSGA-III. Next, MOO would not perform well with our many objectives and examining their behavior is out of our research scope. Finally, the ability to consider several disconnected optimal regions is not the purpose of R-NSGA-III. Instead we seek to evaluate R-NSGA-III ability to focus on a specified region of space.

The WFG problem collection set is much more complex and customized Huband et al. (2006) to which we specifically focus on WFG5 and WFG6. The main advantage for these two problems is a concave POF shape without any polynomial or constant bias to allow for control. WFG6 has a uni-modal shape, but contains feature inseparability. The uni-modal shape and additional controls allow it to focus more on handling feature interdependence. In contrast, WFG5 has no separability, but has a multi-modal shape that is intentionally shaped to deceive the algorithm to get stuck in local optima.

4.2 Quality Indicators

The R-NSGA-III study did not use metrics and instead focused on displaying proof of concept through visualizations. There are many different performance metrics to use, but we focus specifically on accuracy of the approximated POF which is known as quality indicators. Some of the most popular quality indicators are hyper-volume (HV) Guerreiro et al. (2021), Inverted Generational Distance (IGD) Miyakawa et al. (2019) , and IGD+ Ishibuchi et al. (2015).

HV calculates the volume of the dominated portion of the objective space by manually deciding a reference vector that bounds the objective. Normalized from 0 to 1, a higher HV indicates higher quality POF approximation. The main advantages of HV is that the true POF does not need to be known and HV is Pareto-Compliant Bezerra et al. (2017). A quality indicator is Pareto-Compliant if and only if the ranking of POF approximation does not contradict Pareto Optimally. For example, two PO points should not be ranked differently by quality indicators. The main disadvantages with HV is exponential time complexity and a decreased measurement accuracy with higher objectives Guerreiro et al. (2021).

IGD is based on Generational Distance which calculates the distance between each objective vector in the approximation front and the closest objective vector in a reference front. IGD reverses the order of the fronts by instead calculating the distance between each objective vector in the reference front and the closest objective in the approximation front Miyakawa et al. (2019). Lower IGD values are associated with higher quality. The problem with IGD is that it is not pareto compliant and it is sensitive to the selection of reference vectors. However, IGD is arguably the most popular MOO metric as it is fast to compute and is more resilient against many objectivesMiyakawa et al. (2019).

The advantage of IGD+ is that it is weakly pareto compliant. IGD+ modifies IGD by using the euclidean distance on objective vectors dominated by the reference front. For the objective vectors that are non-dominated, only the dominated objective functions in the objective vector have their distance calculated Ishibuchi et al. (2015).

Recent research on IGD has consistently shown similar behavior between IGD and IGD+ Bezerra et al. (2017) which suggests using only IGD+ in our experiment. However, several state-of-the-art research continues to use IGD and therefore we include these values for proper comparison to other research. Finally, we decide to exclude HV from our research because we focus on test problems with 4 to14 objectives to which it will perform less ideally. We expect IGD and IGD+ to show decreased performance with more dimensions and features for all compared algorithms because of the difficulties associated with managing higher data complexity.

4.3 Riesz s-Energy

A recent discovery by K. Deb's lab found success in modifying Minimal Riesz Energy Point Configurations (Riesz Energy) to address the problems of many dimensions Blank et al. (2021). We include this in our paper to ensure our empirical assessment is up to date. We update all three algorithms(NSGA-III, R-NSGA-III, and KBC-RNSGA-III) with Riesz s-Energy for comparison.

The R-NSGA-III paper uses NBI to generate well spaced points on the unit simplex. Some problems arise with NBI when with an increasing number of objectivesDeb and Sundar (2006). As the number of M dimensions increases, the number of reference vectors explodes exponentially.A solution to this problem could be to reduce the number of partitions, but this results in lower resolution. Also, NBI develops a bias as it begins to generate an increasing number of reference vectors on the exterior of the objective hyperspace and struggles with generating interior vectors.

Riesz Energy defines energy as the proportional inverse distance between two bodies. The minimum energy solution then corresponds to a diverse distribution of bodies in a three-dimensional space Hardin and Saff (2003). Riesz Energy was then modified to become Riesz s-Energy which generalizes to more dimensions. This is done by obtaining a matrix which minimizes the energy to every other vector on the objective space hyperplane.

4.4 Proposed Algorithm: KBC-RNSGA-III

NSGA-III switched from crowding distance to the niche association operator to reduce computation effort in deciding which points are emphasized for reproduction and promoting diversity when many objectives. Since R-NSGA-III limits the search space based on pref-

erence, it makes less member evaluations. Also, since R-NSGA-III is exposed to a smaller subset of points in the subspace, it makes sense it should work harder in determining the ranking association of those points.

In the current niche operator in NSGA-III, the closest individual member is associated with a reference vector. Then all individuals within ϵ from that single selected member are eliminated algorithm. This is done for each reference vector in which each of them has a corresponding selected member. Consequently, all individual member fronts have to have their distance measured to each selected member. If they are too close to any selected member they are cleared. If there are any remaining population members, then the 2nd member to each reference vector is selected. Then again, any point within ϵ from a selected point is eliminated from the algorithm. This whole process continues until all points are eliminated or selected. In order to associate an individual member with a reference vector, argmin of the distance matrix of individuals is taken to the reference vectors.

The problem with this approach is that very good candidates may be eliminated. Also, eliminated candidates may contained very similar distances to each reference vector. For example, member A and member B of the current front are 0.5 and 0.501 from the reference vector. If member B is also within ϵ distance of member A, then it is inappropriately discarded.

We propose stochastic acceptance of the members being selected by the niche association operator for each reference point. Our idea is akin to soft clustering where we select the k Nearest Neighbors to a reference point in which closer members have a higher probability, but farther members still have chance to be accepted. The downside of this approach is that we are setting a parameter. We need to select K number of points because assigning a probability to all members would be too computationally expensive. However, we reason this parameter will still work well with small k values and it will likely not be necessary to consider a large number of candidates. In order to avoid an increase in computation time, we insert this soft clustering approach after the distance of each member has been compared to each reference vector. So we assign a probability only to the k Best Neighbors.

The motivation for this change is to promote the diversity of points. The author of the R-NSGA-III study had mentioned that the diversity of points is more of a concern for R-NSGA-III than NSGA-III because R-NSGA-III has less points to consider. This is because R-NSGA-III considers a preferred subset of the approximated POF instead of the fullest of the approximated POF. However, this also largely due to the introduction of the shrinking linear constant in R-NSGA-III denoted as μ. R-NSGA-III shrinks the spacing of the the reference vectors around the user selected aspiration point. This is done so the reference vectors encompass the preferred search space. However, the shrinking constant is significant. A μ of 0.02 was used by the author in their study. This indicates that the projected perpendicular distance from the reference vectors to the central aspiration vector was reduced by 98%. This results in a strong tightening effect of the search space, but we hypothesize results in reduced opportunities of for an offspring to develop more different traits and results in a dense clustering around the aspiration point.

The larger concern for KBC-R-NSGA-III is the impact on the the approximation quality of the POF. Sometimes considering points that are farther away from a reference point may introduce problems with reproduction particularly inefficient crossover. The high shrinking value used by the author suggests that this is a significant concern. Consequently, we aim to

Algorithm 1 KBC-R-NSGA-III(K)

B is Front Member Set
R is Reference Vector Set
P is Probability Matrix
S is Sorted Matrix
$D^{RxB}$is Distance Matrix
for b = 1,2. . . in B **do**
    for r = 1,2. . . in R **do**
        $d_{b,r}$ = Normal Distance(b,r)
        $D \leftarrow d$
    end for
end for
$S \leftarrow \emptyset$
for k = 1,2. . . in K **do**
    s $\leftarrow$ argmin(D)
end for
for s = 1,2 . . . in S **do**
    $P \leftarrow p_{b,r} = 1/d_{b,r}^2$
end for
Random Select p in P

determine if we can increase diversity without sacrificing approximation ability of the POF. We evaluate this by using quality indicators and comparing our proposed modification to the quality indicators of NSGA-III and R-NSGA-III.

5. Experimental Design

The purpose of this paper was to evaluate R-NSGA-III in various circumstances, and to study the effect of algorithmic modifications and genetic operator extensions to R-NSGA-III. We compare NSGA-III, R-NSGA-III, and our proposed k-best niche neighbors modification on DTLZ1, DTLZ2, DTLZ3, DTLZ4, WFG5, and WFG6. We explain why we specifically choose these three comparisons. First, we exclude MOO algorithms such as R-NSGA-II and NSGA-II because they are already well documented to perform lower on many objective problems. Conversely, we do keep NSGA-III even though it does not use preference based search. NSGA-III approximates the POF while R-NSGA-III approximates a preferred partial set. This should result in consistently higher quality indicator values for NSGA-III because it has a larger vantage point for the developing approximated front. However, it is a valid comparison here because it allows us to specifically quantify the addition of the preference based search component as R-NSGA-III directly builds on NSGA-III. Therefore, other EMaO algorithms are not relevant because they do not help show how R-NSGA-III uses aspiration points to generate reference vectors.

We investigated the performance of our algorithms on a combination of 4, 9, 14 objectives with 14, 28, and 42 variables (also referred to as features in our study). The minimum number of objectives in our study is 4 because we are interested specifically in a many-

objective analysis. We set our maximum number of objectives at 14 because problems are typically observed in EmAOs after 15 objectives, and we just wanted to be a little more safe. In regards to the number of variables, we wanted to be sure that the number objectives never exceed the number of variables to avoid the curse of dimensionality. We chose 28 and 42 as multiples of 14 for the purposes of being methodical.

Other studies evaluating EMaO including NSGA-III show a strong relationship between problem complexity and the number of generations provided to approximate the POF. For this reason, we designed the number of generations to depend on a linear combination of the number of objectives, features, and the difficulty of the problem. For 4 objective problems, we ran 100 generations; for 9 objectives, we ran 200 generations; and for 300 generations, 14 objectives. To account for the number of variables, we added 25 generations for 28 variables, and 50 extra generations for 42 variables. Furthermore, certain DTLZ problems are more difficult than others. By observing other EMaO studies, we assign an extra 25 generations for DTLZ1, WFG5, and WFG6. Also, we allow 50 extra generations for DTLZ4, and 75 for DTLZ3. All experiments we ran 3 times each and then averaged.

5.1 Experimental Design Limitations

Our study lacked computing power relative to professional research. In the R-NSGA-III paper, the number of generations used is between 300 to 500 Vesikar et al. (2018), and studies on NSGA-III show around 300-2000 generations Deb and Jain (2014). We were forced to limit our study to 225 generations for our most complex problems because of consistent long wait times for computation to complete. However, we do not believe this to be a very significant issue. Our results still demonstrate how well our algorithms approximate the POF with a limited budget.

Although our approach for selecting parameters is systematic, it is also somewhat arbitrary. We approximated the ratio of objectives, features, and generations by observing general trends in other papers that evaluate EMaO. We attempted to calculate a unique linear combination of these parameters to predict the number of generations used in other studies. However, our calculation did not reliably converge to any relationship, so we decided to go with a simple uniform linear relationship. While there is likely a guide that dictates the optimal parameters for evaluation, we were unable to find it.

6. Results & Empirical Assessment

We ran each algorithm on 4, 9, 14 objective functions, 10, 25, 14 different numbers of features, on six different data-sets. Lower IGD values are associated with a higher approximation of the POF.Table I represents a sample on the DTLZ2 problem set. The results of other tables DTLZ1, DTLZ3, DTLZ4, WFG5, WFG6 were too large too include in this paper, so we summarized them in Table 3.

Consistent with our expectations, IGD become worse with an increasing number of objectives and an increasing number of features. Also, for the majority of the time, we observed increasing by 4 objectives had a stronger increase in IGD/IGD+ than increasing 25 additional features. This is expected because the number of objectives causes a exponential increase in the search space.

In Table I. The left hand-side values are IGD and the right-hand side values are IGD+. NSGA-III was consistently top performer on all combinations of objective numbers and number of features. This was expected because NSGA-III has the advantage of a better vantage point to see the developing front. The improved performance of NSGA-III over R-NSGA-III varied widely. NSGA-III was often reported 0.5 to 10 times more effective than R-NSGA-III in approximating the POF. Furthermore, contrary to our expectations, our proposed method under-performed on all combinations of objectives and numbers of variables. At best, the IGD/IGD+ values of our proposed method was nearly equivalent to R-NSGA-III. At worst, the values were around 25% worse.

Table.I DTLZ2

M	F	Gens	NSGA-III	R-NSGA-III	KBC-RNSGA-III
4	10	125	0.01186 , 0.00996	0.34228 , 0.14134	0.53114 , 0.47704
4	25	150	0.03292 , 0.03163	0.35069 , 0.16642	0.56897 , 0.52319
4	50	175	0.08514 , 0.08149	0.38301 , 0.25268	0.58454 , 0.55262
9	10	225	0.02378 , 0.01798	0.60236 , 0.35109	0.86170 , 0.72298
9	25	250	0.04434 , 0.04343	0.65642 , 0.45104	0.93655 , 0.77764
9	50	275	0.21692 , 0.21369	0.71911 , 0.56134	0.95032 , 0.83877
14	10	325	0.46717 , 0.32658	0.79123 , 0.50684	1.00057 , 0.82979
14	25	350	0.02943 , 0.02931	0.70169 , 0.48201	1.00605 , 0.81725
14	50	375	0.11564 , 0.11564	0.73969 , 0.52576	0.99359 , 0.80926

We attempted to tune number of neighbors to and capture any k best dependent relationships in Table II. Results show a slight worsening of IGD+ values with increasing values of k. Interestingly, IGD increased a little more strongly than IGD. A likely explanation for this is that an candidate individuals became too far from the reference point which also caused the space between parents to be too far. Parents far away from each other are less reliable to produce high fitness children Deb and Jain (2014). Contrary to our expectations, k worsened approximately linearly. K did not show a bell curve.

An additional function of this table is to show the difference between R-NSGA-III and KNN-RNSGA-III our proposed approach. The percent difference row at the top show by how worse our method preformed. The percent difference was calculated as the difference between R-NSGA-III and KNN-RNSGA-III divided the sum.

Table.II DTLZ2 with 250 generations, 9 Objectives, 25 Features

	k=2	k=3	k=5	k =10	k=20
%diff	-7.42 , -20.4	-7.95 , -19.0	-10.9 , -20.5	-14.6 , -23.7	-17.6 , -26.6
R-NSGA3	0.66 , 0.45	0.66 , 0.45	0.66 , 0.45	0.66 , 0.45	0.66 , 0.45
KBC-RNSGA3	0.76 , 0.68	0.77 , 0.66	0.82 , 0.68	0.88 , 0.73	0.94 , 0.78

Table III essentially contains six copies of Table I for each problem set. The easiest problem set for all algorithms was DTLZ2 and DTLZ4. Performance highest on DTLZ2 suggest the algorithms did not experienced much difficulty with an increase in objectives alone. Objectives certainly increase IGD/IGD+, but it seems more problematic when its paired with additional complexity. High performance on DTLZ4 suggests that the algorithms did not have difficulty generating a wide distribution. This was the intended effect of our proposed modification that we were hoping to see. By allowing individuals father away from a reference point to be associated in a probabilistic manner, this is a greater di-

versity of points around the reference point. The hardest problems was DTLZ1 and DTLZ3. Difficulty on DTLZ1 suggest that all algorithms had more difficulty with problems that had more local optima and more variable inseparability.

Below, KBC-R-NSGA-III out-preforms R-NSGA-III on the DTLZ3 problem. This is the only example of success DTLZ3 measured the ability to converge to one global optimal, but many local optima.

Table.III Dependence on Problem Sets

IGD				
Problem	Extra Gens	NSGA3	RNSGA3	KBC-RSGA3
DTLZ1	+25	5.67	14.7	16.63
DTLZ2	+0	0.04	0.66	0.94
DTLZ3	+75	33.93	58.74	47.73
DTLZ4	+50	0.03	0.56	0.75
WFG5	+25	1.82	2.08	2.94
WFG6	+25	1.96	7.85	11.38
IGD+				
Problem	NSGA3	RNSGA3	KRNSGA3	
DTLZ1	+25	5.64	14.69	16.64
DTLZ2	+0	0.04	0.45	0.78
DTLZ3	+75	33.92	58.74	47.73
DTLZ4	+25	0.03	0.26	0.62
WFG5	+25	0.51	1.70	2.47
WFG6	+25	0.88	7.38	10.71

7. Discussion

KBC-R-NSGA-III displayed reduced performance on all of the data-sets, but did show improved performance on the data-set with multiple local optima. Multiple local optima can be viewed as a specific niche. Through modifying algorithm, we learned many clues on how to improve our algorithm. We graph if individuals had moved too far away from each other based on real time IGD values of the front. We trigger KBC-R-NSGA-III only if individuals had very similar distance values set by a small threshold distance. We expect this to give noticeable improvements.

7.1 Importance

There are three reasons that make the evaluation of R-NSGA-III particularly important and pertinent to MOO processes and decision making science in general. Preference based EMaO methods present an opportunity for high utility to industrial and real life optimization problems in which the computational complexity is already high. Preference guided search avoids extra computation and still generates optimal, relevant solutions. Evaluating R-NSGA-III helps better understand its niche and pitfalls, and we hope inspire further modification to the algorithm.

First, the use of preference information provides an opportunity for NSGA-III to include domain expert knowledge in the decision making process. Decision makers may already have

an idea of what they are searching for based on intuition and experience. This experience and intuition is invaluable knowledge as the objective functions can only approximate a real world situation. Therefore, the human element in focusing on a specific preferred region on the Pareto front represents an opportunity to model a closer simulation of a real life optimization problem.

Second, R-NSGA-III provides a more succinct candidate pool. Decision making research has shown that as the number of candidate solutions increase in a problem, the decision maker may become more indifferent and less engaged in their reasoning Council (1983). R-NSGA-III provides a relevant and targeted approach to finding optimal solutions that decision makers are interested in.

Third, R-NSGA-III can provide additional points that are similar to an optimal point. In the case that preference information is not available, a low resolution POF can be approximated to quickly obtain optimal points. Then, the preferred optimal optimal points can be used to decide a preferred region of interest for additional exploration. Alternatively, the decision maker may already have access to an approximation of the POF from a third party, but seeks to further investigate a specific region. The current POF approximation may contain irregularities and counter-intuitive results, or a specific region may need to be validated for security purposes.

7.2 Experimental Limitations

NSGA-III consistently out-preformed R-NSGA-III. We believe the superiority of NSGA-III is likely inflated because we experienced great difficulty in deciding an aspiration vector. During our experiment, we found that IGD and IGD+ values can be highly dependent on the aspiration point chosen. If aspiration values are facing towards the edges or away from the shape, the IGD and IGD+ will explode indicating a drop in performance. We also observed that there is sometimes a relationship with the number of aspiration vectors chosen. Adding aspiration points sometimes caused IGD and IGD+ values to increase especially when the aspirations were close to each other.

Although we the Pareto front for these problems were known, visualizing an appropriate aspiration point in a hyperplane was very difficult. We attempted to use hyper-dimensional visualization techniques to select an aspiration point, but this poorly generalized to situations with 14 objectives. Even with 9 objectives, our aspiration points frequently points to the edge of the true Pareto front. Reference vectors that worked well for 4 dimensions did not work well with 9 dimensions and reference vectors in 9 dimensions were especially unpredictable in generalizing to 14 dimensions. We believe unpredictable drops in quality indicator performance occurred when we picked an aspiration point far away. We believe this likely happened because the multi-modal natures of the some of the shapes and the multiple dimensions makes it very difficult to predict what coordinates will be align with a shape. In the end, we instead settled to choose the origin point as an aspiration point because it was well represented in all true Pareto front hyper-planes.

However, placing aspiration points at the origin of a shape in not optimal. For all shapes except DTLZ1, the number of optimal points in the center were the least dense and the number of points near the edges of the hyperplane the most dense. Therefore, we gave a large disadvantage to R-NSGA-3 and KNN-RNSGA-III. We chose only one reference vector

and chose it at a location where the number of optima are much less. NSGA-III is more superior in this regard because the reference vectors do not depend on the placement of an aspiration point.

Another likely source of error was the number of generations chosen. More generations reliably decrease IGD (increase performance). While used a systematic approach to assign generations to more difficult problems, we are not sure if we gave specific problems too many generations and some others too little. This would not increase errors between algorithms because each algorithm was given the same amount of generations However, we believe that NSGA-III benefits much more from a higher number of generations because it is able to utilize a much larger, and coarse distributed search space.

7.3 Superior Performance with Multiple Local Optima

KBC-R-NSGA-III showed superior performance on the DTLZ3 test data-set on 30 runs with 4, 9, and 14 objectives and 10, 25, and 50 variables per objective. This was repeated 30 times. DTLZ3 tests the EMaO algorithm's ability to escape local optima. The improved performance of KBC-R-NSGA-III here makes sense because the stochastic niche selection prevents settling to any local optima. The major contribution of this work would be that a stochastic niche selection operator could be beneficial when facing a data-set with multiple local optima. A Boolean hyper-parameter could be added to R-NSGA-III to activate a stochastic niche selection operator to better handle multiple local optima.

7.4 Future Work

There is much more useful knowledge we can learn by preforming further experiments on this algorithm. Since R-NSGA-III avoids much computation by approximating a partial set of the POF, we wanted to know by how much of an advantage this is. We planned to do this with three measurements. First, we would determine the number of average number of extra generations it would take for R-NSGA-III to match NSGA-III.Second, we were planning on counting the number of population evaluations for each method. Third, we wanted to plot average number of extra generations needed against population evaluations avoided. Next, we were planning on investigating the role of MCDM making algorithms in selecting an aspiration vector. MCDM algorithms typically have much less time complexity because the preform many objective optimization constraining the problem to a discrete structure. We wanted to investigate a partnership between MCDM and R-NSGA-III by examining how much benefit we could get by using an MCDM algorithm to output an aspiration vector to be used by R-NSGA-III. Lastly, we also wanted to compare their combined population evaluations of MCDM algorithms and R-NSGA-III against NSGA-III to see if this is a warranted approach.

If this project is pursued further, generating an adaptive shrinking procedure would be the best use of time. We believe a significant clue is examining the relationship of the niche association operator influencing the amount of shrinking. The R-NSGA-III author uses a very small shrinking factor which reduces the reduces the perpendicular distance of reference vectors around the aspiration by 98%. We believe this should not be always necessary because there may already be an abundance of dominant individual front members around a reference point. This achieves good exploitation of high fitness individuals around

a reference vector, but, limits exploration too much which likely prevents finding of PO points in the long-term. In other words, adaptive shrinking is the concept of exploitation vs exploitation. These two ideas were the original inspiration for our proposed algorithm KNN-RNSGA-III modification. As the front approximation evolves, due to the randomness of the reproduction process and the unknown complex hyperplane shape, different regions of the objective space should require different degrees of shrinking. Shrinking should be triggered on a spectrum in response to a widening spread of individuals around a reference vector. In cases where individuals are already near a reference vector, shrinking should be slightly turned off to allow additional exploration.

References

Leonardo C. Bezerra, Manuel López-Ibáñez, and Thomas Stützle. An empirical assessment of the properties of inverted generational distance on multi- and many-objective optimization. In *9th International Conference on Evolutionary Multi-Criterion Optimization - Volume 10173*, EMO 2017, page 31–45, Berlin, Heidelberg, 2017. Springer-Verlag. ISBN 9783319541563.

J. Blank and K. Deb. pymoo: Multi-objective optimization in python. *IEEE Access*, 8: 89497–89509, 2020.

Julian Blank, Kalyanmoy Deb, Yashesh Dhebar, Sunith Bandaru, and Haitham Seada. Generating well-spaced points on a unit simplex for evolutionary many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 25(1):48–60, 2021.

National Research Council. *Research Needs for Human Factors*. The National Academies Press, Washington, DC, 1983. ISBN 978-0-309-07867-2.

Indraneel Das and J. E. Dennis. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002a.

K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, volume 1, pages 825–830 vol.1, 2002b.

Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4): 577–601, 2014.

Kalyanmoy Deb and J. Sundar. Reference point based multi-objective optimization using evolutionary algorithms. In *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, GECCO '06, page 635–642, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595931864.

Andreia P. Guerreiro, Carlos M. Fonseca, and Luí s Paquete. The hypervolume indicator. *ACM Computing Surveys*, 54(6):1–42, jul 2021.

D. P. Hardin and E. B. Saff. Minimal riesz energy point configurations for rectifiable d-dimensional manifolds, 2003.

S. Huband, P. Hingston, L. Barone, and L. While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki, and Yusuke Nojima. Modified distance calculation in generational distance and inverted generational distance. In António Gaspar-Cunha, Carlos Henggeler Antunes, and Carlos Coello Coello, editors, *Evolutionary Multi-Criterion Optimization*, pages 110–125, Cham, 2015. Springer International Publishing. ISBN 978-3-319-15892-1.

Minami Miyakawa, Hiroyuki Sato, and Yuji Sato. A study for parallelization of multi-objective evolutionary algorithm based on decomposition and directed mating. In *Proceedings of the 2019 3rd International Conference on Intelligent Systems, Metaheuristics amp; Swarm Intelligence*, ISMSI 2019, page 25–29, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450372114.

N. Srinivas and Kalyanmoy Deb. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evol. Comput.*, 2(3):221–248, sep 1994. ISSN 1063-6560.

Yash Vesikar, Kalyanmoy Deb, and Julian Blank. Reference point based nsga-iii for preferred solutions. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1587–1594, 2018.

Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.