A Differential Evolution Variant of NSGA II for Real World Multiobjective Optimization

Chung Kwan, Fan Yang, and Che Chang

Electrical and Computer Engineering, National University of Singapore

Abstract. This paper proposes the replacement of mutation and crossover operators of the NSGA II with a variant of differential evolution (DE). The resulting algorithm, termed NSGAII-DE, is tested on three test problems, and shown to be comparable to NSGA II. The algorithm is subsequently applied to two real world problems: (i.) a mass rapid transit scheduling problem and (ii.) the optimization of inspection frequencies for power substations. For both the real world problems, NSGAII-DE is found to have generated better results based on comparative studies.

Keywords: differential evolution (DE), NSGA II, real world problems.

1 Introduction

The non-dominated sorting genetic algorithm II (NSGA II) was developed by Deb et al. [1] to improve the NSGA [2]. The main attraction of NSGA II is its fast non-dominated sorting algorithm that is computationally more efficient than most available non-dominated sorting techniques. In addition, a crowding distance assignment algorithm without a need for a niching parameter for maintaining the diversity among pareto-optimal solutions adds to the attraction of this algorithm. NSGA II is currently one of the most popular multiobjective evolutionary algorithms (MOEAs), with comparisons made with other well known MOEAs like SPEA [3] and PAES [4], achieving competitive superior results for many test problems.

To further improve the algorithm, this paper proposes to replace the crossover and mutation operators of the original NSGA II algorithm using a variant of differential evolution DE [5].Termed NSGAII-DE, the algorithm is tested with three selected test problems. They are based on Schaffer (SCH) [6], Kursawe (KUR) [7], and Zitzler's test problems (ZDT6) [8]. To demonstrate the ability of the proposed algorithm in real world problems, the algorithm is tested on a mass rapid transit scheduling problem improvised from [9] and the optimization of inspection frequencies for substations modeled in [10].

2 Development of Proposed Algorithm

2.1 Background and Related Works

Kwan and Chang proposed a DE-based heuristic in [9] for solution to a simplified train scheduling problem. For that problem, the variant for generating mutant vector V for the $(G+1)^{th}$ generation of the i^{th} candidate solution X_i is:

M. Randall, H.A. Abbass, and J. Wiles (Eds.): ACAL 2007, LNAI 4828, pp. 345–356, 2007.

$$V_{i,G+1} = \lambda(X_{i,G} - X_{r1,G}) + F(X_{r2,G} - X_{r3,G}) \quad r1 \neq r2 \neq r3 \neq i \tag{1}$$

where *r1, r2, r3* $\in$ *[1, N]* are randomly selected solutions from the population of size N. *F* and λ are amplification factors for the bracketed terms. Motivated by initial success, attempts are made to further apply the proposed algorithm to other types of problems. However, subsequent tests revealed that this variant does not converge fast enough for many applications. This motivates us to explore a better variant to suit a wider range of real world applications.

Recently, multi-objective DE based techniques are also reported in works like [11] and [12]. Notably, a similar attempt to replace the crossover and mutation rates by a rotationally invariant DE variant was noted in [13], where the authors reported better performance of their NSDE than the NSGA II for a class of rotated problems. These further support the notion that the common mutation and crossover operators may not be effective in handling certain problems.

2.2 Proposed Variant

The proposed variant presented in equation (2) replaces the term $X_{r1,G}$ in equation (1) with $X_{rBest,G}$ and omit the $X_{i,G}$ term. Thus, instead of generating each mutant vector with respect to $X_{i,G}$, $V_{i,G+1}$ is generated with respect to the term $X_{rBest,G}$.

$$V_{i,G+1} = \lambda X_{rBest,G} + F(X_{r2,G} - X_{r3,G}) \quad r2 \neq r3 \tag{2}$$

where *r2* and *r3* $\in$*[1, N]* are randomly chosen solutions from the population of size *N*. In the single objective case, $X_{rBest,G}$ is merely the best solution for the G^{th} generation. In the multiobjective case, however, the notion of 'best' is no longer a single optimum term, but is chosen from a set of non-dominated or Pareto-optimal solutions. In this work, $X_{rBest,G}$ is randomly chosen from the top 30% of set of solutions of the G^{th} generation. The NSGA II algorithm ranks each candidate solution from '1' onwards depending on how many solutions it is dominated by. Solutions with ranking of '1' denote the non-dominated solutions in the current population. The higher the rank value of a solution, the more it is dominated by other solutions. Besides that, the crowding-distance-assignment assigns each solution based on density estimation, with a higher value representing lesser crowding of other solutions around its vicinity. Solutions at the boundary points are assigned ∞. *rBest* is selected first based on the non-dominated ranking followed by the crowding-distance-assignment. E.g. if more than 30% of solutions are ranked '1', a crowding-distance-assignment of higher value (with the same non-dominated ranking) would be preferred to those of lower values in consideration of whether a solution should be included as one of the possible choices in the set where $X_{rBest,G}$ is chosen from.

The 'crossover' operator of DE generates a trial vector $U_{i,G+1}$ as presented in equation (3):

$$U_{ji,G+1} = \begin{cases} V_{ji,G+1}, \text{if } r(j) \leq CR \text{ or } j = rn(i) \\ X_{ji,G}, \text{ if } r(j) > CR \text{ or } j \neq rn(i) \end{cases} \tag{3}$$

where CR is the crossover rate. j denotes the j^{th} decision variable of the i^{th} candidate solution. $r(j)$ is a randomly chosen number in [0,1]. Thus, if the randomly generated $r(j)$ of the j^{th} decision variable is smaller or equal to the crossover rate CR, $U_{ji,G+1}$ will take the value of the mutated vector $V_{ji,G+1}$, else it will take the value of the original candidate solution $X_{ji,G}$. In addition, to prevent degeneration, the term $rn(i)$ is a randomly chosen decision variable j in the i^{th} candidate solution which will be chosen to be replaced by the mutant vector.

Unlike in the single objective version of DE proposed in [5], $U_{i,G+1}$ does not replace the current $X_{i,G}$ if it is better, but it is treated as a member of the child population candidate as provided in the NSGA II structure, and is involved in non-dominated ranking and crowding distance assignment.

The proposed variant is tested in this paper using theoretical test problems first and subsequently on two real world problems.

3 Comparison with Test Problems

In order to test the effectiveness of our proposed algorithm, we test the algorithm with the three test problems SCH, KUR and ZDT6 detailed in Table 1. All approaches are run for a maximum of 25000 function evaluations (translated to 250 generations with population size of 100).

The real-coded version of the NSGA II is used, with associated parameters, crossover and mutation distribution index (η_c and η_m), both set to 20 as recommended. The NSGAII-DE has crossover settings CR and F set to 0.8 and λ set to 1. Codes are implemented in Matlab. Ten runs are performed for each algorithm.

The three test problems are selected because of the following reasons – SCH is selected to test the convergence for a large range of variable(s). KUR is selected to test convergence to a nonconvex and discontinuous objective space, and ZDT6 is used to test convergence to a nonconvex and nonuniformly spaced objective space. Such features are common in the real-world problems, which we will be applying the algorithm to ultimately. Pareto fronts for all the cases are presented in Fig.1. In Fig.1(a) and (b), the two Pareto fronts from NSGA II and NSGA II-DE coincide with each other. Whereas in Fig.1(c), it is obvious that NSGA II-DE performs better than NSGA II.

In addition, two performance measures proposed in [1] correspond to two goals in a multiobjective optimization, i.e. convergence and diversity: The convergence metric γ measures the extent of convergence of the solutions generated by the algorithms to a known set of pareto-optimal solutions by computing the Euclidean distance between the solutions. The second diversity metric Δ (see equation (1) of [1]) provides a measure comparing the Euclidean distance between each pareto-optimal solution against all the others. Table 2 shows the mean (first rows) and variance (second rows) of the convergence metric γ. A better convergence is obtained by the NSGA II-DE for both KUR and ZDT6, but it performs slightly worse for SCH. Table 3 shows the mean and variance of the diversity metric Δ. Better diversity measures are noted for NSGA II-DE in all three test problems.

With this initial testing on theoretical problems, we established the comparable performance of the DE variant with the original NSGA II, and proceed to applying the algorithms on the two real-world problems.

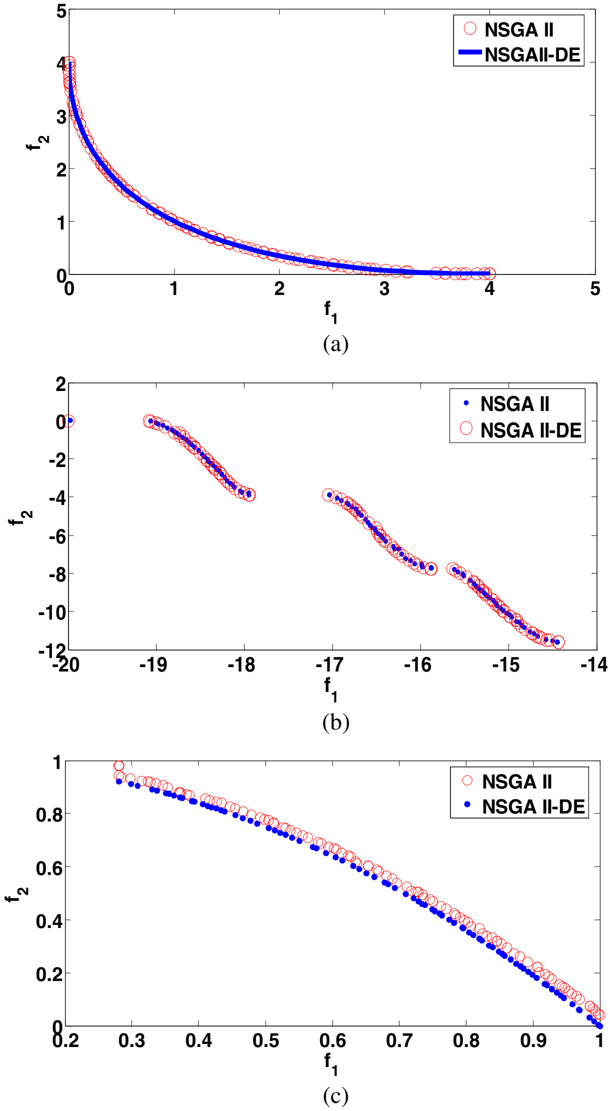

Fig. 1. Pareto fronts of test problems (a). SCH, (b). KUR, and (c). ZDT6

4 Mass Rapid Transit Schedule Optimization

4.1 Background and Formulation

Waiting time and traveling time are important service quality objectives considered in the optimal generation of a mass rapid transit schedule. In this paper, each service

Table 1. Theoretical test problems used in this study

Prob-lems	n	Variable Bounds	Objective functions	Optimall solutions	Com-ments
SCH	1	$[-10^3, 10^3]$	$f_1(x) = x^2$, $f_2(x) = (x-2)^2$	$x \in [0, 2]$	convex
KUR	3	[-5,5]	$f_1(x) = \sum_{i=1}^{n-1}(-10\exp(-0.2\sqrt{x_i^2 + x_{i+1}^2}))$ $f_2(x) = \sum_{i=1}^{n}(\lvert x_i \rvert^{0.8} + 5\sin x_i^3)$	(refer to [16])	Non-convex
ZDT6	10	[0,1]	$f_1(x) = 1 - \exp(-4x_1)\sin^6(6\pi x_1)$ $f_2(x) = g(x)[1 - (f_1(x)/g(x))^2]$ $g(x) = 1 + 9[(\sum_{i=2}^{n} x_i)/(n-1)]^{0.25}$	$x_1 \in [0, 1]$, $x_i = 0$, $i = 2, \cdots n$	Non-convex Non-uniformly spaced

Table 2. Mean (first row) and variance (second row) of the convergence metric γ

Algorithm	SCH	KUR	ZDT6
NSGA II	0.003098	0.025761	0.86599
	0	0.000564	0
NSGA II-DE	0.003262	0.015507	0.86425
	0	0.000005	0.000028

Table 3. Mean (first row) and variance (second row) of the diversity metric Δ

Algorithm	SCH	KUR	ZDT6
NSGA II	0.47102	0.53958	0.77723
	0.004493	0.006099	0.15121
NSGA II-DE	0.4554	0.40955	0.58166
	0.000527	0.000464	0.07022

quality is investigated with respect to the economic objective of operating cost. The work is extended from [9] by increasing the number of variables from 7 to 108 taking into account for instantaneous variables relating to dwell time at each passenger station and run time/ coast level profiles between all track sections. Previously, all the service qualities are combined into a weighted sum that is minimized against the operating cost. This paper investigates the effect that each of the service-related attribute has on operating cost instead.

An intuitive way to incorporate the objective(s) is to minimize the absolute waiting time and traveling time of the whole system. However, [14] has pointed out that the 'expectation' concept of passengers is a more accurate reflection of passengers' attitude towards waiting time and traveling time than the objective(s) above. Adopting that idea, a waiting time dissatisfaction index ($DI_{WT}(t)$) was derived in equation (4) and an actual to shortest journey time ratio (ASJR) defined in equation (5).

$$DI_{WT}(t) = \begin{cases} 0 & \text{for } t \le E_{WT} \\ m_1 t + C_1 & \text{for } E_{WT} < t \le T_{first} \\ \alpha(t)^2 & \text{for } t > T_{first} \text{ and } t > E_{WT} \end{cases} \tag{4}$$

$DI_{WT}(t)$ denotes the average waiting time dissatisfaction incurred by passengers on a particular passenger platform, where E_{WT} denotes the amount of time that a group of passengers are willing to wait before any 'dissatisfaction' is incurred, taken as 60 seconds in this paper. T_{first} is the time interval to the arrival of the 1st train. A linear dissatisfaction measure represents the dissatisfaction of passengers up to the point where the first train arrives (under a periodic planning timetable without real time operational adjustment, this is taken as the dispatch interval). If passengers cannot get onto the first train that arrives due to congestion, the waiting time dissatisfaction index will increase in a quadratic manner (implied by the $\alpha(t)^2$ term).

To reflect the total traveling time 'dissatisfaction', we propose an actual to shortest journey time ratio. The shortest journey time is calculated based on the summation of minimum run times and the minimum dwell time allowable at each station respectively. The *ASJR* definition is:

$$ASJR = \frac{\text{Actual Total Travel Time}}{\text{Shortest Total Travel Time}} \tag{5}$$

where the shortest total travel time can be deemed the 'expected' total traveling time of the passengers.

Under certain operating range, each of the two service qualities is found to be contradictory with the operating cost. Shortening the run time between each station to shorten the traveling time, for example, leads to an increase in electrical energy consumption [15], which in turn increases the operating cost. Reducing the waiting time by increasing the dispatch intervals increases the number of trains in the system and increases the operating cost as well. The aim of applying NSGAII-DE is to discover the range where each service quality is found to have a conflicting relationship with operating cost. This will facilitate the decision making process for implementation considerations.

Detailed definition of operating cost as well as the passenger flow model equations can be found in [9]. Note that the cost coefficients have to be changed to protect the privacy of the study system and the cost values should by no means be interpreted to represent the true values.

4.2 Decision Variables, Constraints and Other Settings

The decision variables to be optimized are dispatch frequencies, dwell times (stopping times) at each station as well as coast levels between each section track relating running time to the electrical energy consumption. Some of the constants relating to maximum train capacity and allowable passenger build-up on each station are summarized in Table 4. Besides the constant parameters, constraints on the bounds on each decision variable and the safety distances allowable between trains apply. Interested readers are encouraged to read [9] for a more detailed description of the problem.

The optimization parameters are presented in Table 5. The population size of 100 and the maximum iterations are determined after trials are conducted. The mutation, crossover distribution index and other parameters are unchanged as in the test case studies.

Table 4. Constant parameters used in the scheduling model

Parameters	Values
Expected waiting time E_{WT}	60 seconds
Maximum train capacity T_{MAX}	1500 passengers
Shortest possible total traveling time (dir1)	3163 seconds
Shortest possible total traveling time (dir2)	3154 seconds
Maximum allowable passenger build-up on each station (P_{MAX})	1000 passengers

Table 5. Optimization parameters

Algorithm	η_m	η_c	F	λ	Coding	Population Size	Max Iteration
NSGA II	20	20	-	-	Real	100	2500
NSGA II-DE	20	20	0.8	1	Real	100	2500

4.3 Pareto-Fronts Generated

Multiple runs are conducted for each algorithm to determine its consistency. The average and best results for each algorithm are presented. The average results are plotted to provide insight in terms of the consistency of the Pareto-optimal solutions.

Two sets of Pareto-fronts generated in Figs.2 and 3 demonstrate the superiority of the proposed NSGA II-DE technique over NSGA II. NSGA II-DE is seen to clearly dominate NSGA II solutions for both situations. Moreover, slight deviations were noted for the average results and best results for NSGA II-DE as compared to NSGA II for both Figs.2 and 3.

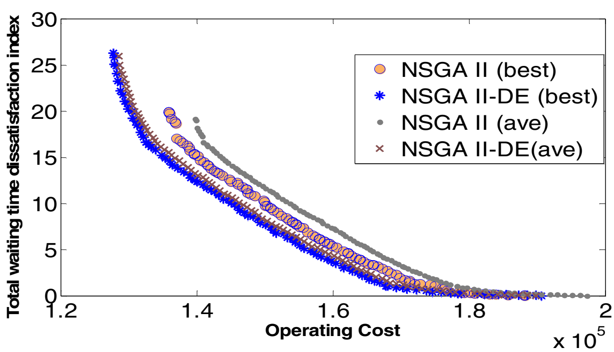

Fig. 2. Pareto plots of operating cost vs. total waiting time dissatisfaction

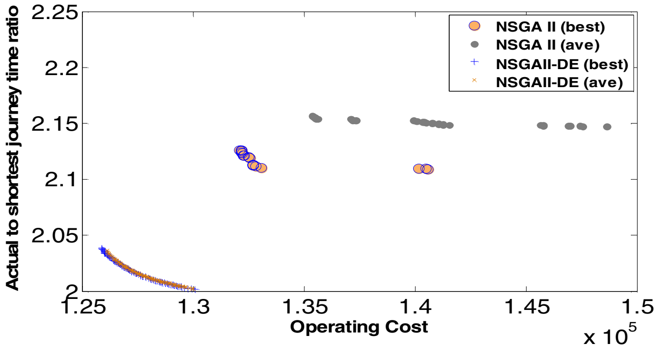

Fig. 3. Pareto plots of operating cost vs. actual to shortest journey time ratio

4.4 Summary of Results and Physical Implications

The two figures display the 'trade-off' relationships between each service quality and the corresponding range of operating cost. Operating cost trading off with the total waiting dissatisfaction index range from approximately $129000 to $190000, while that for actual to shortest journey time ratio is a much shorter range from $126000 to $130000. For the purpose of illustration, this paper has optimized each service quality with operating cost as two objective problems. The establishment of NSGAII-DE as the most appropriate algorithm in this work allows us to extend it to more objectives in the future.

Furthermore, the identification of the trade-off regions for each service quality with respect to operating costs provides greater insight into the problem itself for discovering domain knowledge which can be subsequently employed in more objectives to shorten the computational time.

5 Inspection Frequencies Optimization for Substations

5.1 Background

Improving overall reliability and reducing operating cost are the two most important but often conflicting objectives for substation optimization. A Markov and system reliability model are developed to assess the impact of changing inspection frequencies of individual component on reliability and operating cost for various substation configurations in [10].

At the component-specific level, the deterioration process is modeled with a three-state Markov process shown in Fig. 4. As time progresses, transitions from one state to the next are made. Should proper maintenance be taken after the inspection, the component can be restored from deteriorated condition back to a better one. In this model, $\lambda_{i,i+1}$ is the transition rate from state i to i+1, $\lambda_{i,f}$ denotes transition rate from state i to failure state, and $\mu_{i,j}$ represents the transition rate from state i back to j (j<i).

Generally speaking, with more frequent inspections and subsequent maintenance actions carried out, a component can be restored faster from a more deteriorated state to a better state.

For component n, the expected maintenance cost, $EC_{m,n}$, and the expected repair cost, $EC_{r,n}$, are calculated using the equations (6) and (7):

$$EC_{m,n} = \sum_{i=1}^{N} (EquivalentI_i \times \sum_{k=1}^{3} EquivalentC_{mk}) \ , \tag{6}$$

$$EC_{r,n} = C_r \times FailureFrequency \tag{7}$$

where $EqivalentI_i$ is a product of the state probability and inspection frequency in that state, representing the equivalent inspection frequency in state i, and $EqivalentC_{mk}$ is the equivalent cost of maintenance type k (k=1, 2, 3). C_r is the failure cost each time.

At the system-specific level, the adopted system reliability model was developed to assess the composite reliability of power generation and distribution [16], which views substation configurations as being connected in series or parallel or a combination of both.

The overall cost containing the capital and operating costs and the Loss of Expected Energy (*LOEE*) are the two criteria to evaluate the performance of substation configurations.

The overall cost (*C*) in one substation can be easily calculated by:

$$C = \sum_{n=1}^{M} (EC_{m,n} + EC_{r,n}) + CapC \times Rate \tag{8}$$

where *CapC* is the capital cost, and *Rate* is the interest and depression rate. *M* is the number of components.

LOEE is the reliability objective which measures the reliability worth associated with the cost of the customers due to the failure, which is expressed as:

$$LOEE = \sum_{p=1}^{m} Pf_p \times L_p \times Du_p \tag{9}$$

where *m* is the number of load points in one substation, Pf_p is the probability of failure at load point p. L_p is the loss of load (MW) due to the failure at load point p, and Du_p is the duration of failure at load point p.

5.2 Case Studies

Two basic substation configurations analyzed in this paper are shown in Fig. 5. The capital cost is calculated based on the typical data about the length of bus, number of breakers, transformers and other system equipment. Other parameters are set as: *Rate* = 12%, *N* = 3. The optimization parameters are laid out in Table 6.

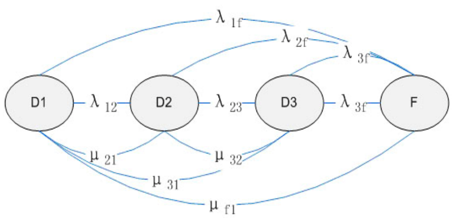

Fig. 4. Three state Markov model

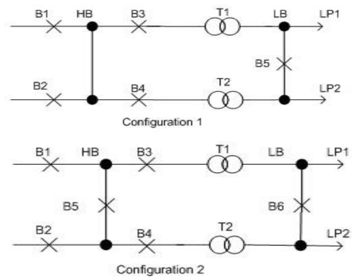

Fig. 5. Typical substation configurations

Table 6. Optimization parameters (c1—configuraiton 1; c2—congfiguration 2)

Algorithm	η_m	η_c	F	λ	Coding	Population Size	Max Iteration
NSGA II	20	20	-	-	Real	110(c1) 120(c2)	65 (c1) 85(c2)
NSGA IIDE	20	20	0.8	1	Real	110(c1) 120(c2)	65 (c1) 85(c2)

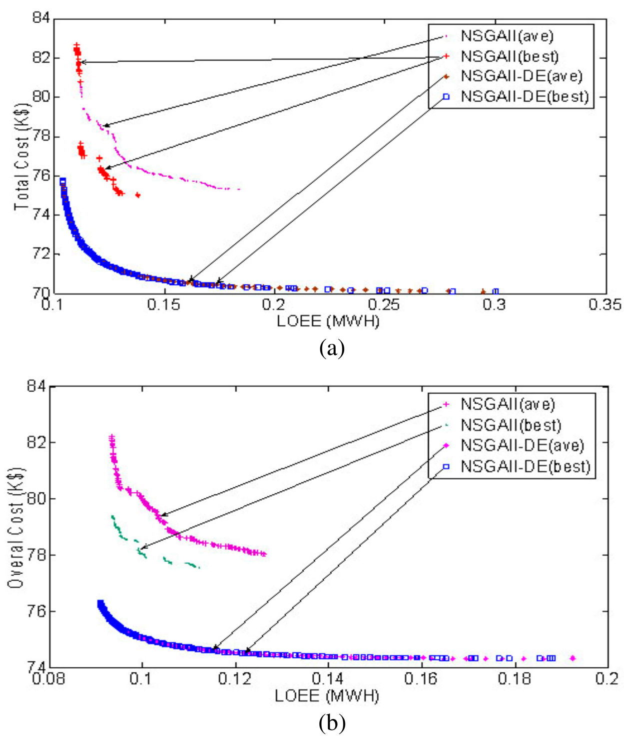

Fig. 6. Pareto plots of LOEE vs. overall cost for (a) configuration 1, and (2) configuration 2

5.3 Optimization Results

The minimization of overall cost (equation (9)) against *LOEE* (equation (10)) is presented for the two configurations depicted in Fig 6. The same model with four configurations has been studied in [10]. The two sets of Pareto-fronts generated in Fig.6 demonstrate the superiority of the proposed NSGA II-DE technique over NSGA II. NSGA II-DE is seen to clearly dominate NSGA II solutions for both situations.

6 Conclusion

This paper has identified a differential evolution (DE) variant of the NSGA II that is well suited to real world applications. The algorithm, termed NSGAII-DE, was tested against three test problems and found to outperform NSGA II in terms of both convergence for KUR and ZDT6 and diversity for all three test problems. The extension of the problem to two real world problems, an optimal train scheduling problem and a substation optimization problem, demonstrate the effectiveness of this algorithm when compared with NSGA II.

References

1. Deb, K., Agrawal, S., Pratab, A., Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197 (2002)
2. Srinivas, N., Deb, K.: Multiobjective function optimization using nondominated sorting genetic algorithms. Evol. Comput. 2, 221–248 (1995)
3. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: A comparative case study and the Strength Pareto Approach. IEEE Transactions on Evolutionary Computation 3, 257–271 (1999)
4. Knowles, J., Corne, D.: The Pareto archived evolution strategy: A new baseline algorithm for multiobjective optimization. In: Proceedings of the 1999 Congress of Evolutionary Computation, pp. 98–105. IEEE Press, Piscataway, NJ (1999)
5. Storn, R., Price, K.: Differential Evolution-A simple and efficient adaptive scheme for global optimization over continuous space. Technical Report TR-95-012, ICSI
6. Schaffer, J.D.: Multiple objective optimisation with vector evaluated genetic algorithms. In: Grefensttete, J.J. (ed.) Proceedings of the First International Conference on Genetic Algorithms, pp. 93–100. Lawrence Erlbaum, Hillsdale, NJ (1987)
7. Kursawe, F.: A variant of evolution strategies for vector optimization. In: Schwefel, H.-P., Männer, R. (eds.) Parallel Problem Solving from Nature - PPSN I. LNCS, vol. 496, pp. 193–197. Springer, Heidelberg (1991)
8. Zitzler, E.: Evolutionary algorithms for multiobjective optimization: Methods and applications. Doctoral dissertations ETH 13398, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland (1999)
9. Kwan, C.M., Chang, C.S.: Application of Evolutionary Algorithm on a Transportation Scheduling Problem - The Mass Rapid Transit. In: IEEE Congress on Evolutionary Computation 2005, Edinburgh, vol. 2, pp. 987–994 (September 2005)

10. Chang, C.S., Yang, F.: Evolutionary Multi-objective optimization of Inspection Frequencies for Substation Condition-based Maintenance. In: Proceedings of 11th Naval Platform Technology Seminar, Singapore (accepted for publication, 2007)
11. Abbass, H.A., Sharker, R.: The Pareto Differential Evolution Algorithm. International Journal on Artificial Intelligence Tools 11, 531–552 (2002)
12. Xue, F.: Multi-objective Differential Evolution and its Application to Enterprise Planning. In: Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA 2003), vol. 3, pp. 3535–3541. IEEE Press, Los Alamitos (2003)
13. Iorio, Li, X.: Solving Rotated Multi-objective Optimization Problems Using Differential Evolution. In: Webb, G.I., Yu, X. (eds.) AI 2004. LNCS (LNAI), vol. 3339, pp. 861–872. Springer, Heidelberg (2004)
14. Murata, S., Goodman, C.J.: An optimal traffic regulation method for metro type railways based on passenger orientated traffic evaluation. In: Proceedings of COMPRAIL 1998, pp. 573–584 (1998)
15. Chang, C.S., Sim, S.S.: Optimizing train movements through coast control using genetic algorithms. In: IEE Proceedings of Electr. Power Appl. vol. 144 (1997)
16. Propst, J.E.: Calculating Electrical Risk and Reliability. IEEE Trans. Industry Applications 31, 1197–1205 (1995)