Statistical analysis of convergence performance throughout the evolutionary search: A case study with SaDE-MMTS and Sa-EPSDE-MMTS

Joaquín Derrac*, Salvador García†, Sheldon Hui‡, Francisco Herrera* and Ponnuthurai N. Suganthan‡
*Dept. of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology). University of Granada, 18071 Granada, Spain. E-mails: jderrac@decsai.ugr.es,herrera@decsai.ugr.es
†Dept. of Computer Science. University of Jaén, 23071 Jaén, Spain. E-mail: sglopez@ujaen.es
‡School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Ave., 639798 Singapore
E-mails: shel0003@e.ntu.edu.sg, epnsugan@ntu.edu.sg

***Abstract*—Typically, comparisons among optimization algorithms only considers the results obtained at the end of the search process. However, there are occasions in which is very interesting to perform comparisons along the search. This way, algorithms could also be categorized depending on its convergence performance, which would help when deciding which algorithms perform better among a set of methods that are assumed as equal when only the results at the end of the search are considered.**

In this work, we present a procedure to perform a pairwise comparison of two algorithms' convergence performance. A nonparametric procedure, the Page test, is used to detect significant differences between the evolution of the error of the algorithms as the search continues. A case of study has been also provided to demonstrate the application of the test.

I. Introduction

One critical step in the design of new algorithms for tackling computational intelligence problems is to test them among the most representative techniques in its field. Before presenting a new approach, it is imperative to have it compared with the current algorithms in the state-of-the-art in order to demonstrate its usefulness.

Nonparametric statistical tests are a powerful tool for performing these comparisons. In contrast with classical parametric techniques, which are based on the assumptions of independence, normality and homoscedasticity of the data, nonparametric tests are able to perform valid comparisons without the necessity of assuming the aforementioned properties, thus providing better justifications on the correctness of the comparison [1], [2].

The same situation holds when working with optimization algorithms. Valid statistical comparisons have to be performed in order to ensure whether there are significant differences among the algorithms analyzed [3].

Nowadays, most current techniques focus their comparisons of the results of the algorithm only after they have ended. This precludes an experimental setup which has already defined a way of determining a terminate condition for the algorithms (for example, a fixed number of evaluations of the fitness function), and only the results obtained at the end are compared.

However, there are cases in which it may be more practical to complement such comparisons with techniques or analysis that are able to compare the algorithms during the search process, instead of just focusing on the final results. An example of this situation can be seen in the results of the special issue of the Soft Computing Journal on Scalability of Evolutionary Algorithms and other Metaheuristics for Large Scale Continuous Optimization Problems [4], where several of the Differential Evolution algorithms presented performed in a statistically similar manner with respect to the final solutions achieved.

In cases such as the above one, it would be desirable to use another method to evaluate the convergence performance of the algorithms, so as to discern which methods are performing better throughout the search, regardless the final results.

In this work, we present a methodology for performing pairwise comparisons between algorithms, considering its performance along the search process. The Page test [5], a nonparametric procedure, is adopted to perform such analysis, obtaining as a result the confirmation that the convergence performance of one algorithm is better, if significant differences can be found.

Examples of the application of the test are given, involving the computation of the algorithm's performance over several steps of the search process. A case study, including two advanced versions of the Self-Adaptive Differential Evolution algorithm [6], SADE-MMTS [7] and SaEPSDE-MMTS [8] has been carried out, showing how the Page test can be used to complement the results of other statistical techniques (the Wilcoxon test for pairwise comparisons [9], in our case) when the latter will not be able to detect significant differences if the final results are considered.

The rest of this work is organized as follows: Section II presents our proposal for contrasting algorithm's convergence performance, including a description of the original Page test and an explanation about how to employ it to test algorithm's convergence. Section III presents a full case study, showing the potential of the procedure as a complement to the Wilcoxon test. Section IV concludes the work.

978-1-4673-5873-6/13/$31.00

II. Analysis of convergence performance using the Page test

Our proposal is devoted to present a procedure for comparing the convergence performance between two algorithms while solving a given set of optimization problems. This comparison is performed by using the Page test, a nonparametric statistical test originally defined to testing ordered alternatives in multiple classifications. Section II-A provides a brief description of the test, while Section II-B details how it can be used to analyze the convergence performance.

A. The Page test

The Page test for ordered alternatives [5] is a nonparametric test which is analogous to the existing procedures for the two-way analysis-of-variance problem, [10]. The procedure works by considering the null hypothesis of equality between the k treatments analyzed, which can be rejected in favor of an ordered alternative.

When using this test, the practitioner has to provide, before starting the analysis, an order among the k treatments. After such order is provided, the measures of the k treatments on N samples can be analyzed in the same way than the well-known Friedman test [11]: The measures in each sample are ranked from best to worst, giving a rank of 1 to the best measure, a rank of 2 to the second, ..., and a rank of k to the worst. In the case of ties, average ranks are assigned (for example, a tie between the first and the second result would produce an average rank of $(1 + 2)/2 = 1.5$).

After obtaining the ranks, the Page L statistic can be computed using the following expression

$$L = \sum_{j=1}^{k} jR_j = R_1 + 2R_2 + ... + kR_k \tag{1}$$

where r_i^j is the rank of the j-th of k measures on the i-th of N samples, and $R_j = \sum_{i=1}^{N} r_i^j$. With this definition, R_1 represents the sum of the ranks of the treatment predicted to have the smallest sum of ranks, R_2 represents the sum of ranks of the treatment predicted to have the second smallest sum, and so forth. If the data are consistent with the initial ordering stipulated by the practitioner, then the sum of ranks values R_j will follow an increasing order.

The critical values for the L statistic can be computed for small values of k and N (see, for example, Table Q in [12] for values up to $k = 8$ and $N = 12$). In the case that larger values would be needed, a normal approximation should be considered. The normal approximation for the L statistic is given by the following expression

$$Z = \frac{12(L - 0.5) - 3Nk(k+1)^2}{k(k+1)\sqrt{N(k-1)}}. \tag{2}$$

whose estimation, including a continuity correction, is approximately standard normal with a rejection region on the right tail.

B. Analysis of algorithm's convergence performance

In its original definition, the Page test allows the detection of increasing trends within the rankings computed from the data. If a proper order of the treatments is chosen, ranks will follow an increasing order, and thus the hypothesis of equality of ranks will be rejected, in favor of the ordered alternative.

In the context of comparing the convergence performance between two algorithms, the data should consist of the differences between each algorithms best objective value reached by each one, at several points of the search. In continuous optimization problems, where the aim of the algorithms is to minimize an error measure, the differences can be recorded by subtracting the best error of one algorithm to the best error of the second one.

Such differences, if taken at enough points of the search process, would enable tracking of the comparative performance of both algorithms during the search. Using this setup, if differences are computed as *Error(Alg1) - Error(Alg2)*, three different outcomes are possible:

- Differences are increasing: If a consistently increasing trend is found, this means that either the error of *Alg1* is growing faster than the error of *Alg2* or that the error of *Alg2* is decreasing faster than the error of *Alg1*. Since errors are computed as the best value found along the search, the former case is impossible. Thus, it can be assumed that, if an increasing trend is detected, this means that the error of *Alg2* is decreasing faster, which means that has a better convergence performance.
- Differences are decreasing: Following the same reasoning as above, this could only mean that the error of *Alg1* is decreasing faster. Hence, a decreasing trend in the differences means that *Alg1* has a better convergence performance.
- No trend can be identified: If no consistent trend is found, then nothing can be said about the relative convergence performance of both algorithms.

Following these rules, a data matrix can be built by computing the differences between the algorithms at several steps (cut-points). Depending on the number of test functions N available, the number k of cut-points (the treatments) should be chosen accordingly (see, for example, [12] for recommendations).

In this case, the order of the treatments should be increasing, since we are interested to analyze the trends as the search progresses. That is, the first treatment should be the first cut-point, the second treatment should be the second cut-point and so on.

Since the Page test is only able to detect increasing trends in the ranks (decreasing trends in the differences, since lower ranks denote higher values), if differences are computed as *Error(Alg1) - Error(Alg2)* the test will try to confirm that *Alg1* has a better convergence performance. If this is not desired, then differences should be computed as *Error(Alg2) - Error(Alg1)*, whereby the test will try to confirm that *Alg2* has a better convergence performance.

TABLE I. EXAMPLES OF THE PAGE TEST

Algortihms	C1	C2	C3	C4	C5	L	p-value
Alg1 - Alg2	41	49.5	60.5	66	68	925.5	0.000659
Alg2 - Alg1	73	64.5	53.5	48	46	784.5	0.999438
Alg3 - Alg4	54.5	46	73.5	49.5	61.5	872.5	0.217691
Alg4 - Alg3	59.5	69	39	65.5	52	836.5	0.808336

After computing the ranks, the L statistic and its associated p-Value can be obtained. Table I shows several illustrative examples, where C1 to C5 are the ranks computed at 5 different cut-points:

- The comparison of *Error(Alg1) - Error(Alg2)* shows a clear increase in the ranks, which is confirmed by a very low p-value. Moreover, if the comparison is performed in the opposite sense, *Error(Alg2) - Error(Alg1)*, it is also clear that the ranks are not increasing (in fact, they are decreasing), which is reflected by a p-value near to 1.0. This result means that *Alg1* has a better convergence performance that *Alg2*.
- Neither the comparison of *Error(Alg3) - Error(Alg4)* nor the comparison of *Error(Alg4) - Error(Alg3)* show a clear trend in any sense (neither of the p-values depict a significant result). Hence, it can be concluded in this case that no significant trend is detectable between *Alg3* and *Alg4*.

We would like to remark that, as shown by these examples, this application of the Page test is only able to detect significant increasing trends in the ranks (decreasing in the differences). If the opposite trend needs to be detected, the order of the algorithms has to be inverted when computing the differences. Also, note that although the test is symmetric in its definition, the ranks and the p-values computed will not be symmetric if ties are present among the differences.

III. CASE STUDY

A case of study comparing the performances of two evolutionary optimization algorithms has been conducted as an application example of our approach. Section III-A shows the experimental framework with the test functions chosen to evaluate the performance of the algorithms. Section III-B describes the algorithms selected and performs a comparison between them in standard terms. Finally, this comparison is enhanced in Section III-C, with the addition of the testing procedure based on the Page test.

A. Experimental framework

Several test functions have been selected for this case study. They have been described in the *Soft Computing* journal Special Issue on Scalability of Evolutionary Algorithms and other Metaheuristics for Large Scale Continuous Optimization Problems [4], and were recommended as a suitable benchmark for testing the capabilities of evolutionary algorithms and other metaheuristics.

This benchmark consists of 19 functions featuring different traits, the most important of which are the inclusion of unimodal/multimodal, separable/non-separable and shifted functions. The first 11 functions (denoted from F1 to F11) are the following:

- Shifted Unimodal Functions:
  - F1: Shifted Sphere Function
  - F2: Shifted Schwefels Problem 2.21
- Shifted Multimodal Functions:
  - F3: Shifted Rosenbrocks Function
  - F4: Shifted Rastrigins Function
  - F5: Shifted Griewanks Function
  - F6: Shifted Ackleys Function
- Shifted Unimodal Functions
  - F7: Shifted Schwefels Problem 2.22
  - F8: Shifted Schwefels Problem 1.2
  - F9: Shifted Extended f10
  - F10: Shifted Bohachevsky
  - F11: Shifted Schaffer

The other 8 functions (F12 to F19) are obtained as a hybrid composition of the first ones (for detailed information about them, please refer to [4]). In this study, all functions have been considered with 50 dimensions.

B. Algorithms analyzed

In this study, two versions of the Differential Evolution [13], [14], a recent evolutionary, population-based technique for developed for tackling continuous optimization problems, will be analyzed. Both are based in the Self-adaptive Differential Evolution search algorithm [6], which introduces a self-adaptive control mechanism to change relevant control parameters as the search process progress.

The first one, denoted **SaDE-MMTS**, is a Self-adaptive Differential Evolution hybridized with a Modified Multi-Trajectory Search strategy [7]. The Modified Multi-Trajectory Search strategy enhances the search performed by the original SADE algorithm by being applied frequently to refine several diversely distributed solutions at different search stages, satisfying both global and local search requirements.

The second version, denoted **Sa-EPSDE-MMTS**, features an Ensemble of Parameters and mutation Strategies in Differential Evolution with Self-adaption [15], and has been improved with the Modified Multi-Trajectory Search strategy. Again, the Modified Multi-Trajectory Search strategy is introduced in this method to enhance the behavior of the original algorithm, becoming a very accurate search technique [8].

A standard comparison between both approaches would require them both to be executed over a test function benchmark. In this case, 25 runs have been performed on each of the 19 functions included in the experimental framework. A maximum of 5000·D fitness function evaluations (250000) have been considered. Table II shows the average results achieved (the best result in each function is highlighted in **bold**; we also report the number of functions in which each algorithm achieves the best average).

If the final results are analyzed, it is possible to see a small advantage of the Sa-EPSDE-MMTS algorithm. However, differences are too small to state that it improves the

TABLE II. SADE-MMTS VS SA-EPSDE-MMTS: FINAL RESULTS

Function	SaDE-MMTS	Sa-EPSDE-MMTS
F1	**-4.50E+02**	**-4.50E+02**
F2	**-4.50E+02**	**-4.50E+02**
F3	3.94E+02	**3.91E+02**
F4	**-3.30E+02**	**-3.30E+02**
F5	**-1.80E+02**	**-1.80E+02**
F6	**-1.40E+02**	**-1.40E+02**
F7	**1.79E-14**	8.07E+01
F8	**5.46E-07**	2.68E-06
F9	2.10E-01	**1.94E-01**
F10	**0.00E+00**	**0.00E+00**
F11	**1.69E-01**	1.94E-01
F12	**8.52E-31**	9.69E-29
F13	3.19E+01	**1.02E+00**
F14	1.15E-02	**3.68E-14**
F15	1.04E-15	**4.26E-16**
F16	1.26E-31	**0.00E+00**
F17	3.53E+00	**3.39E-01**
F18	**4.06E-02**	7.17E-02
F19	**0.00E+00**	**0.00E+00**
# Best	12	14

TABLE III. RESULTS OF THE WILCOXON TEST

Wilcoxon test	R+	R-	p-value
Sa-EPSDE-MMTS vs SaDE-MMTS	110.5	60.5	0.266769

SaDE-MMTS algorithm. This comparison can be contrasted statistically, by using a proper 1 versus 1 nonparametric test such as the Wilcoxon test [3], [9]. Table III shows the results obtained after the application of the test.

Again, although the $R+$ value representing the ranks of Sa-EPSDE-MMTS is higher, the p-value obtained is too large to reject the null hypothesis. Hence, it cannot be concluded that Sa-EPSDE-MMTS is significantly better than SaDE-MMTS.

However, it would be interesting to see if this conclusion holds in other stages of the search. As a preliminary analysis, it is possible to gather intermediate values corresponding to the best fitness value obtained at several times during the search. If, for example, results are collected after every 10% of the evaluations, averaged, and compared again using the Wilcoxon test, it is possible to somewhat analyze the evolution of the comparative along the search process. Table IV summarize the results of the 10 Wilcoxon tests performed to measure the convergence performance.

In this second analysis, it is possible to see that there are several steps of the search in which the differences are significant: It can be assumed that Sa-EPSDE-MMTS performs better when 40% of the evaluations have been spent, and between a 60% and a 90%. This could be taken as a suggestion that Sa-EPSDE-MMTS is converging better than SaDE-MMTS in general, being only tied at the very end of the run. However, there is no way of summarizing or adding the p-values obtained by the Wilcoxon test to formulate a strong hypothesis stating that Sa-EPSDE-MMTS is showing a better convergence performance.

TABLE IV. WILCOXON TEST WITHIN THE SEARCH

Sa-EPSDE-MMTS vs SaDE-MMTS	R+	R-	p-value
10%	113	77	0.456586
20%	105.5	65.5	0.369362
30%	110.5	60.5	0.266769
40%	123.5	47.5	0.093603
50%	114.5	56.5	0.196438
60%	132.5	38.5	0.03858
70%	166	24	0.004011
80%	160.5	29.5	0.007908
90%	160.5	29.5	0.007908
100%	110.5	60.5	0.266769

C. Convergence performance analysis

After the preliminary analysis, it is advisable to perform a convergence performance analysis to statistically determine whether significant differences exist between Sa-EPSDE-MMTS and SaDE-MMTS along the search. To perform this analysis, the procedure described in this work will be applied.

The first step is to set-up the number of cut-points or steps at which the results will be collected. Since $N = 19$ (there are 19 different functions in the test benchmark), it would be reasonable to choose a number of steps lesser than N. Following that, we have chosen to consider 10 different steps (that is, $k = 10$), which is also in accordance with the analysis performed with the Wilcoxon test in the preliminary analysis.

Next, average differences at each step have been gathered for each function. This means that the differences between the best objective value of every algorithm have been collected at 10%, 20%, ..., 100% of the maximum number of fitness evaluations. Table V shows these differences.

Firstly, the convergence performance of Sa-EPSDE-MMTS - SaDE-MMTS will be tested, which means that we will try to confirm that Sa-EPSDE-MMTS has a better convergence performance. To do so, the cut-points are sorted in increasing order (from 10% to 100%) and the respective differences are analyzed.

Table VI shows the results of this first application of the Page test. It can be observed that the ranks obtained are approximately increasing along the search, which means that the values depicting the differences between both algorithms are decreasing as the search progress. Since these values are computed as fitness(Sa-EPSDE-MMTS) - fitness(SaDE-MMTS), this means that the error of Sa-EPSDE-MMTS is decreasing (converging) faster, yielding lower difference values. This is also contrasted by the p-value obtained, 0.000031, which establishes that there is a significant trend in this sense.

The next step consists on performing the opposite comparison, in order to check if the opposite trend (SaDE-MMTS converges faster) can be found. To do so, differences shown at Table V have to be reversed (that is, multiplied by -1) to reflect that now the test is going to be carried out in the fitness(SaDE-MMTS) - fitness(Sa-EPSDE-MMTS) sense.

Table VII shows the results of this second application of the Page test. In this case, it can be observed that the ranks obtained are approximately decreasing along the search, which means that the values depicting the differences between both algorithms are increasing as the search progress. This result is opposite to the one obtained before, and suggests that the hypothesis of a faster convergence of SaDE-MMTS can be discarded. In fact, the p-value computed this time, 0.999971,

TABLE V. SA-EPSDE-MMTS VS SADE-MMTS AVERAGE DIFFERENCES

Step	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
F1	-3.12E+02	-3.12E+00	-1.49E-02	-6.40E-05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F2	2.63E+01	2.07E+01	9.41E+00	3.44E+00	1.42E+00	5.05E-01	7.28E-03	7.28E-03	7.28E-03	7.28E-03
F3	-1.76E+06	-1.58E+03	-7.49E+01	-4.80E+00	3.67E-01	-7.06E+00	-4.06E+01	-3.12E+01	-2.42E+01	-1.78E+01
F4	2.84E+01	-2.81E+01	-2.70E+01	-1.85E+01	-9.98E+00	-4.61E+00	-1.14E+00	-7.96E-02	3.63E-02	3.98E-02
F5	-3.18E+00	-7.02E-01	-1.27E-01	-4.55E-03	-4.32E-04	-2.96E-04	-2.96E-04	-2.96E-04	-2.96E-04	-2.96E-04
F6	1.13E+00	1.55E-01	-1.46E-02	-1.03E-03	-3.60E-05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F7	6.08E+01	7.95E+01	8.06E+01	8.07E+01	8.07E+01	8.07E+01	8.07E+01	8.07E+01	8.07E+01	8.07E+01
F8	5.17E+03	1.93E+03	6.33E+02	1.56E+02	2.56E+01	2.50E+00	-6.50E-02	-7.55E-03	-8.86E-04	-6.13E-05
F9	-4.64E+00	-3.34E+01	-2.20E+01	-1.46E+01	-7.70E+00	-3.93E+00	-1.86E+00	-6.19E-01	-1.61E-01	-3.26E-02
F10	-2.81E+01	-8.64E+00	-1.27E-01	-2.12E-04	-2.40E-07	-1.86E-10	-1.32E-13	0.00E+00	0.00E+00	0.00E+00
F11	-8.25E+00	-2.63E+01	-1.59E+01	-1.24E+01	-7.22E+00	-3.31E+00	-1.50E+00	-4.98E-01	-1.02E-01	1.51E-02
F12	4.20E+01	1.08E-01	7.18E-05	-1.41E-09	-4.49E-12	-1.27E-15	-2.39E-19	-3.10E-23	-2.28E-27	9.61E-29
F13	-5.90E+05	7.93E-01	-2.21E+01	-2.00E+01	-1.97E+01	-2.82E+01	-4.07E+01	-3.74E+01	-3.40E+01	-3.10E+01
F14	1.93E+01	-2.61E+01	-2.57E+01	-1.94E+01	-1.19E+01	-5.35E+00	-2.00E+00	-4.94E-01	-8.58E-02	-1.52E-02
F15	-1.63E+01	-1.56E+00	-6.87E-02	-2.65E-03	-4.70E-05	-6.82E-07	-1.02E-08	-1.02E-10	-6.09E-13	-6.10E-16
F16	1.08E+00	5.27E-05	2.77E-10	3.40E-17	1.14E-25	-1.26E-31	-1.26E-31	-1.26E-31	-1.26E-31	-1.26E-31
F17	-4.73E+03	5.31E+01	1.10E+01	-5.32E+00	-5.70E+00	-1.74E+00	-9.19E+00	-7.30E+00	-5.57E+00	-3.29E+00
F18	-3.86E+00	-1.75E+01	-1.40E+01	-9.65E+00	-5.55E+00	-2.29E+00	-8.65E-01	-2.31E-01	-4.13E-02	2.24E-02
F19	-2.25E+01	-4.68E+00	-7.41E-02	-1.02E-03	-6.19E-06	-3.03E-08	-1.93E-10	-7.11E-13	-5.24E-16	0.00E+00

TABLE VI. PAGE TEST: SA-EPSDE-MMTS - SADE-MMTS

Step	Ranking
10%	93
20%	84
30%	80
40%	89
50%	101.5
60%	112.5
70%	101
80%	115
90%	128
100%	141
L statistic	6228.5
p-value	0.000031

TABLE VII. PAGE TEST: SA-EPSDE-MMTS - SADE-MMTS

Step	Ranking
10%	116
20%	125
30%	129
40%	120
50%	107.5
60%	96.5
70%	108
80%	94
90%	81
100%	68
L statistic	5266.5
p-value	0.999971

cannot be used to reject the null hypothesis, and thus does not confirm the existence of a significant trend favorable to SaDE-MMTS.

In summary, the joint application of the Page tests in both senses has enabled us to determine that there exists a significant trend (and only one) in the evolution of the differences between Sa-EPSDE-MMTS and SaDE-MMTS. This trend, which depicts a reduction in the error obtained by Sa-EPSDE-MMTS (with respect to SaDE-MMTS), is a result that supports the hypothesis of Sa-EPSDE-MMTS having a better convergence performance than SaDE-MMTS.

D. Some recommendations on the use of the Page test

As we have shown in the previous sections, the Page test can be considered as a new way of comparing the performance of two different algorithms. It enables to find differences within the search process, in contrast with most of the available techniques for comparing different algorithms which are usually limited to analyzing only the final results achieved.

However, it is important to note that, besides its statistical background, the test only identifies trends between the convergence performances, and therefore its results should not be taken as an absolute proof that a certain algorithm performs better than a second one. As with most of the hypothesis testing techniques, the results obtained after the application of the procedure should be carefully analyzed and interpreted by the practitioner.

In this specific case, we recommend its usage along with a competent statistical procedure for detecting differences between pairs of algorithms. For example, the set-up shown in Section III-B would be adequate since the comparisons performed allow to highlight differences between both the final results of the algorithms (using the Wilcoxon test) and along the search process (using the Page test). Note that, although the Page test should be used only for pairwise comparisons, it could be also applied after performing a set of multiple comparisons (for example, by a combination of the Friedman test and the Holm *post-hoc* procedure [3]), for characterizing the performance of pairs of algorithms statistically equivalent at the end of the search (as could be found by the Friedman and Holm procedures). In such cases, it would be desirable to choose, among the best performing techniques, those who show a better convergence performance in the middle stages of the search.

Also, it is important to stress the role of this technique for performing multiple problem comparisons. Although in certain cases single problem comparisons may be desirable, multiple problem comparisons have the advantages of allowing to determine which techniques perform better in a wide range of general situations. Of course, if the specific conditions of an experiment require the analysis to be focused upon a specific kind of problems, what the user must do is to simply choose a proper set of problems of that kind, and performing the analysis considering that restriction on the nature of the problems. This is, indeed, a more suitable option than performing several single problem comparisons and look for a way of, somehow, combining the results obtained (which is not always a straightforward matter).

IV. CONCLUDING REMARKS

In this work, a procedure for analyzing the convergence performance of pairs of algorithms has been proposed. It is based on the use of the Page test, a nonparametric statistical procedure. The usage of the test is demonstrated using a case study featuring a comparison between Sa-EPSDE-MMTS and SaDE-MMTS, two advanced versions of the Self-adaptive Differential Evolution search algorithm.

The case study has shown the potential of the test, demonstrating a key tool for understanding the differences between both algorithms while within the search. It has enabled us to highlight the differences while exist between the algorithms, despite the fact that no significant differences could be found at the end of the search by classical nonparametric statistical techniques.

ACKNOWLEDGMENT

Supported by the Projects TIN2011-28488, P10-TIC-06858 and P11-TIC-9704. J. Derrac holds an FPU scholarship from the Spanish Ministry of Education.

REFERENCES

[1] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability," *Soft Computing*, vol. 13, no. 10, pp. 959–977, 2009.

[2] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the CEC'2005 special session on real parameter optimization," *Journal of Heuristics*, vol. 15, pp. 617–644, 2009.

[3] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.

[4] F. Herrera, M. Lozano, and D. Molina, "Test suite for the special issue of soft computing on scalability of evolutionary algorithms and other metaheuristics for large scale continuous optimization problems," 2010. [Online]. Available: http://sci2s.ugr.es/eamhco/CFP.php

[5] E. P. Page, "Ordered hypotheses for multiple treatments: A significance test for linear ranks," *Journal of the American Statistical Association*, vol. 58, pp. 216–230, 1963.

[6] A. K. Qin and P. N. Suganthan, "Self-adaptive differential evolution algorithm for numerical optimization," in *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, vol. 2, 2005, pp. 1785–1791.

[7] S. Z. Zhao, P. N. Suganthan, and S. Das, "Self-adaptive differential evolution with multi-trajectory search for large scale optimization," *Soft Computing*, vol. 43, no. 1, pp. 1–17, 2011.

[8] S. Z. Zhao and P. N. Suganthan, "Comprehensive comparison of convergence performance of optimization algorithms based on nonparametric statistical tests," in *IEEE Congress on Evolutionary Computation*, 2012, pp. 1–7.

[9] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 14, pp. 80–83, 1945.

[10] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures, 5th edition*. Chapman & Hall/CRC, 2011.

[11] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, pp. 675–701, 1937.

[12] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference, 5th edition*. Chapman & Hall, 2010.

[13] R. Storn, K. V. Price, and J. Lampinen, *Differential Evolution - A Practical Approach to Global Optimization*. Natural Computing Series, Springer, 2005.

[14] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[15] R. Mallipeddi, P. N. Suganthan, Q.-K. Pan, and M. F. Tasgetiren, "Differential evolution algorithm with ensemble of parameters and mutation strategies," *Applied Soft Computing*, vol. 11, no. 2, pp. 1679–1696, 2011.