Evolutionary Big Optimization (BigOpt) of Signals

Sim Kuan Goh, Kay Chen Tan, Abdullah Al-Mamun
Department of Electrical and Computer Engineering
National University of Singapore,
4 Engineering Drive, 117583, Singapore.
Email: eletankc@nus.edu.sg

Hussein A. Abbass
School of Engineering and Information Technology
University of New South Wales,
Canberra, ACT 2600, Australia.
Email: h.abbass@adfa.edu.au

***Abstract*—Challenging multi-modal optimization problems have been very successfully solved by evolutionary computation (EC) techniques. To date, many methods have been proposed on evolutionary optimization for both single and multiobjective large scale problems. In the age of Big Data, there is an urge to take evolutionary optimization techniques to the next level for solving problems with even larger scales: thousands and millions of variables. These problems arise in many domains ranging from bioinformatics, to neuroscience and social simulations.**

In this paper, we investigate the use of EC to solve Big electroencephalography (EEG) data optimization problems with thousands of variables. The optimization problem attempts to identify maximum information that should be kept from a signal while minimizing the artifact. The high level of epistasis inherent in a signal can slow down the evolution. Therefore, we investigate the advantages of optimizing the problem in the frequency domain with different thresholds as opposed to the time domain. We propose synthetic EEG data sets of various scale and noise level. These data sets were the basis for the Optimization of Big Data 2015 Competition (BigOpt), CEC 2015. Two state-of-art multiobjective evolutionary algorithms (MOEAs) were evaluated. The results of this work suggest that frequency representation of the signals facilitates dimensionality reduction for big scale optimization of time series data, and hence provides faster and better quality solutions for EEG data cleaning. Moreover, the results suggest that existing state-of-art multiobjective evolutionary computation methods are extremely slow. Methods that can optimize the problem faster and with high quality are needed.

I. Introduction

In a time of unprecedented access to affordable and massive amounts of computational power and resources, evolutionary computation (EC) has provided solutions to many real world applications ranging from engineering problem [1], [2], [3], financial strategies [4], [5], rule extraction and data mining [6], [7], [8] to bioinformatics [9], [10]. EC techniques have drawn much attention due to their flexibility in data representation and innate parallelism. In the age of Big Data, there is an urge to push the boundaries of evolutionary optimization beyond the small-scale problems that have been used as benchmark problems in the literature so far. Evolutionary optimization has successfully solved very large scale "simple" problems such as the one-max function, with a million variables. In global and/or multiobjective optimization problems, however, this scale is beyond the capabilities of existing evolutionary optimization algorithms. A Big Data problem is not just about size. The frequency and nature of changes in the underlying trends in the data, heterogeneity of data sources, level of noise, and other factors come into play to define the true complexity of a Big Data problem. One important factor to consider is: to solve Big problems, we need to solve some small ones; that is, to manage 1TB of data per day, we need to be able to manage 12MB of data per second. While 12MB does not sound large, it is the "per second" that makes the problem harder; that is, the time constraint imposed on the size. Optimization of Big Data 2015 Competition (BigOpt2015), a CEC 2015 competition, takes first steps towards bringing EC to the real-world Big Data challenges.

In computational neuroscience and Brain computer interface (BCI), modern EC techniques have been used to parameterize computational models of the brain [11], [12], [13], [14]. EC reveals intrinsic properties of the model and provides automated solutions. Additional knowledge from neurobiology can be incorporated into EC to offer answers to more challenging and difficult problems. Most BCIs make use of electroencephalography (EEG) to capture brain electrical activities using sensors/electrodes. The recorded EEG arises from synchronized neural activities, produced by the firing of neurons in the brain. Some applications require a mesh of electrodes, Quantitative Electroencephalograph (QEEG), to acquire signals from multiple locations spread over the scalp simultaneously. Decoding the QEEG signals into high-order cognitive states, such as emotions, memory and planning, is a key step towards the design of more advanced BCI systems that are able to perform in every day life, not just for patients, but also for normal activities and safety critical decision making environments, such as real-time management of air traffic controllers workload [15], [16].

The performance of real-time BCI depends on how much true brain signal activities are captured by the signal. The primary challenge in capturing true brain activities lies in the existence of non-brain source electrical activities, known as artifacts, that greatly distort EEG recording. Artifacts are induced by normal human voluntarily and involuntarily activities such as eye movements and muscle contractions. EEG signals are passed through an ICA then a human inspects those components that captured the artifact components of the signal. Through visual inspection, the components are removed and the signal is reconstructed through the remaining components [17]. However, visual inspection is a time consuming tasks. With the vast amount of data collected today in QEEG settings, it is near impossible to visually inspect these data. Automated methods to process the data reduce, or eliminate, human involvement.

In this work, the benchmark is based on a multiobjective

978-1-4799-7492-4/15/$31.00

optimization formulation of EEG artifact removal [18], [19]. The proposed idea for dimensionality reduction is compared using MOEA/D [20] and NSGA-II [21]. Synthetic data sets of various scale and noise level are proposed as benchmarks problems with 1024, 3072 and 4864 variables. To overcome the dimensionality of the problem, we do the optimization in the frequency domain and provide detailed comparison with the time domain analysis. Since some bands will have small values (sometimes some domain knowledge can be used to set a threshold for amplitudes that are considered insignificant and therefore can be ignored during the optimization process), they can also be removed from the optimization. This reduces the dimensionality.

The rest of this paper is organized as follows: Section II presents the methodology of this work, it includes an overview of the synthetic data proposed in BigOpt 2015 and EC approaches to the problem. This is followed by Section III, where the results from the proposed method are presented with comparison of various settings. A discussion is presented in Section VI and conclusions are drawn in Section VII.

II. Methodology

In this section, the proposed synthetic date sets will be discussed, followed by the evolutionary computation formulation of the EEG preprocessing problem. Then, the evaluation of the proposed dimensionality reduction techniques are discussed.

A. Synthetic Data Set

Six datasets are presented in this paper. The datasets have different difficulties in term of number of local EEG signals and artifact sources as well as white noise with variance 0.1 as shown in Table I. The dataset D4 was proposed in [22].

TABLE I: Six datasets with different difficulties

	# of Sources	# of Artifact	Noise Free	With Noise
A	4	2	D4	D4N
B	12	6	D12	D12N
C	19	6	D19	D19N

1) Data Set A: It assumes 6 sources of signals. Four sources are EEG data. Two of the sources are assumed to be EMG artifact, which are distorting the EEG signals, operating at high amplitude at high frequencies overlapping with High Beta and Gamma bands. Each source operates with a mixture of two frequencies representative of classical EEG bands as shown in Table II.

The signals are sampled at 256Hz. Besides, the fifth source was activated in the last 250ms of every second, and the sixth source was activated in the last 500ms of every second. All other sources were activated from time 0. The reason for delayed activation of the EMG signals is that in real-world dataset, there is no guarantee that the clock for collecting data is synchronized with the clock of artifacts.

The six sources are mixed into four signals, $\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4$, as follows:

$$\vec{x}_1 = \vec{s}_1 + 0.9 \times \vec{s}_5$$

TABLE II: The synthesis of the six signal sources.

ID	Band	Amplitude	Frequency	Band	Amplitude	Frequency
$\vec{s}_1$	Delta	14	4	Beta	52	22
$\vec{s}_2$	Theta	23	7	Beta	70	19
$\vec{s}_3$	Delta	16	5	Alpha	43	11
$\vec{s}_4$	Alpha	44	9	Gamma	56	47
$\vec{s}_5$	EMG	144	31	EMG	337	51
$\vec{s}_6$	EMG	282	28	EMG	246	49

$$\vec{x}_2 = \vec{s}_2 + 0.9 \times \vec{s}_6$$

$$\vec{x}_3 = \vec{s}_3 + \vec{s}_5$$

$$\vec{x}_4 = \vec{s}_4 + \vec{s}_6$$

2) Data Set B & C: These datasets assume 18(25) sources of signals. 12(19) sources, s_i, are EEG data. 6(6) of the sources, s_k, are assumed to be artifacts as shown in Table III. Similarly, the signals are sampled at 256Hz. Besides, the source 20, 22, 24 were activated in the last 250ms of every second, and the source 21, 23, 25 was activated in the last 500ms of every second. All other sources were activated from time 0.

In these data set, an EEG model based on 10-20 system with 12 and 19 channels will be used for Data set B and C respectively. Similar to Data set A, the signals generated are assumed to have 2 frequencies ranging from 1 to 50 HZ. Figure 1 also shows the topographic map of a scalp data field in a 2-D circular view to show the contribution of each artifact to all electrodes. The topographic map in Figure 1 was drawn using toolbox from EEGLAB [23], the radius of the head model 0.5dm. The synthetic signals generated from the electrode locations are assumed to be local while 6 artifacts are assumed to be generated at labelled position in Figure 1.

The sources are mixed into signals, x_i, as follows:

$$\vec{x}_i = \vec{s}_i + \sum_{k=1}^{N} w_{ik} s_k \tag{1}$$

$$w_{ik} = \exp -r^2 \tag{2}$$

The w_{ik} for i^{th} electrode and k^{th} artifact is proportional to the exponential of negative squared Euclidean $-r^2$ distance between them. The topographic map in Figure 1 was drawn using toolbox from EEGLAB, the radius of the head model 0.5dm. The synthetic signals generated from the electrode locations are assumed to be local while 6 artifacts are assumed to be generated at labelled position in Figure 1.

B. Independent Component Analysis (ICA)

ICA has been widely used to de-mix signals collected from multiple sources. The de-mixing or separation process is based on a linear transformation of the mixed signals into maximally statistically independent components.

The ICA decomposition problem can be formulated as $X = AS + N$, Where X is the measured mixed signals, A is the mixing matrix, N is the noise and S is the Sources. Mathematically, ICA performs linear transformation on mixed

TABLE III: The synthesis of the 25 signal sources. (*) ID will be used in DataSet B.

Type	ID	Amplitude	Frequency	Amplitude	Frequency
EEG	1*	2	21	24	41
	2*	3	42	36	27
	3*	4	33	49	29
	4*	5	44	30	55
	5*	6	20	50	51
	6*	11	24	34	57
	7*	12	46	25	40
	8*	13	17	53	32
	9*	14	37	23	28
	10*	15	16	43	22
	11*	18	35	46	48
	12*	19	34	20	26
	13	22	41	31	39
	14	23	26	37	44
	15	25	27	38	47
	16	32	45	45	33
	17	36	43	35	56
	18	38	47	52	54
	19	39	40	21	42
Artifact	20*	8	29	164	203
	21*	10	30	229	268
	22*	7	49	216	281
	23*	9	52	151	177
	24*	28	51	190	294
	25*	31	50	255	242

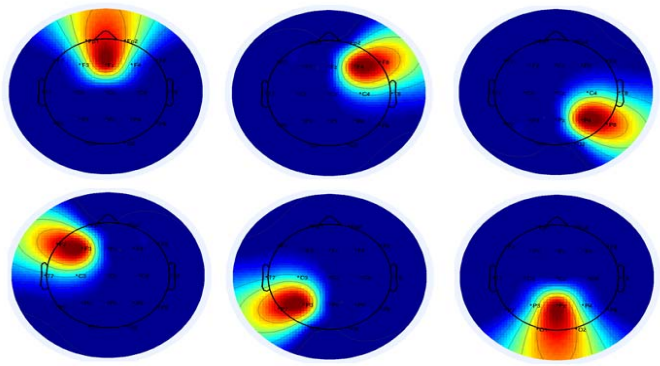


Fig. 1: Topographic maps of scalp data fields in a circular view for each artifact.

signals to components such that the components are statistically independent as well as having non-gaussian distribution. Once we have A and the estimated sources $\hat{S}$, we can find W such that $\hat{S} = WX$. The cleaner signals can be reconstructed from only the useful components. $\hat{X} = W^{-1}\hat{S}$.

C. Multiobjective Problem Formulation

Assume the matrix X is of dimension $n \times m$, where n is the number of inter-dependent timeseries and m is the length of each timeseries. Let S be a matrix of $n \times m$, with n independent timeseries of length m, such that, given A, an $n \times n$ linear transformation matrix,

$$X = A \times S \tag{3}$$

The problem is to decompose the matrix S into two matrices: S_1 and S_2 with the same dimensionality of S; that is $S = S_1 + S_2$ and $X = A \times S_1 + A \times S_2$. Let C be the Pearson correlation coefficient between X and $A \times S_1$; that is,

$$C = \frac{covar(X, A \times S_1)}{\sigma(X) \times \sigma(A \times S_1)} \tag{4}$$

where, $covar(.)$ is the covariance matrix, and $\sigma(.)$ is the variance. The objective is to maximize the diagonal elements of C, while minimizing off-diagonal elements to zeros. At the same time, the distance between S and S_1 should be as minimum as possible; that is, S_1 needs to be as similar as possible as S.

Given, X, A, and S, find S_1 which optimizes the following two objective functions presented in [19]:

$$Minimize f_1 = \frac{1}{N^2 - N} \sum_i \sum_{j \neq i} C_{ij}^2 + \frac{1}{N} \sum_i (1 - C_{ii})^2$$

$$Minimize f_2 = \frac{1}{N \times M} \sum_i \sum_j (S_{ij} - S_{1ij})^2$$

D. Data Representation

We optimize BigOpt data sets in both time and frequency domain. The solutions will be compared.

1) Time Domain: The time series of the signal is encoded into a single individual. The dimension of the solution is 256 multiplied by the number of channels. In this BigOpt, we are solving problems with 1024, 3072 and 4864 variables.

2) Frequency Domain: To overcome the dimensionality of the problem, we do the optimization in the frequency domain instead of the time domain. By using a Fast Fourier Transform (FFT) on the principal components, we at least half the dimensionality of the problem if we use a resolution of 1Hz and work with the magnitude of each spectrum. Since some bands will have close to zero values, they can also be removed from the optimization by thresholding. This reduces the dimensionality even further. Thresholds with value 0, 0.025 and 0.05 of the maximum amplitude of the EEG signals will be studied.

More importantly, the signal in the time domain generates a sequence of interdependent values. This level of interdependency creates extensive level of linkages among the variables that makes the optimization problem very difficult. By transforming the signal to the frequency domains, amplitudes within one epoch can be assumed to be independent variables.

E. MOEAs and Experimental Settings

In this work, two state-of-art MOEAs, namely MOEAD and NSGA-II were used to provide solutions to the formulated multiobjective problems.

1) Parameter Settings: The parameter settings of the problem is summarized in IV. The population size is 20 for all problems. As the scale of the data set increases, the number of generation increases. Pareto solutions were merged over 30 runs.

The setting of control parameters in DE and polynomial mutation: $CR = 1.0$ and $F = 0.5$ in the DE operator, $\eta = 20$ and $p_m = 1/n$ in the polynomial mutation. Other control parameters in MOEA/D-DE: $T = 20, \delta = 0.9, n_r = 2$.

TABLE IV: The parameter settings for MOEAs

Problem Type	D4, D4N	D12, D12N	D19, D19N
Population Size	20	20	20
# of Generation	1000	3000	5000
# of Function Evaluation	20000	60000	100000
# of run	30	30	30

F. Evaluation

1) Score Function: Five solutions will be sampled uniformly from the Pareto set, including the two extreme solutions. The average distance between these five solutions and the corresponding nearest solutions found in the baseline will be used as a measure of quality. Only those solutions found by an entry that dominate the corresponding solution in the baseline will be considered in this metric. For those solutions found by an entry that are dominated by the corresponding solution in the baseline, a fixed distance of value 1000 will be used. The larger this distance, the better.

2) Information Loss and Residual: The advantage of using a synthetic problem is that the ground truth is known. Therefore, we can truly evaluate the efficiency of the proposed method. In this paper, we will use two performance metrics: Information Loss (IL) and Residual (Res) proposed in [19]. The IL metric measures the amount of information loss through the cleaning process of the signal. The second metric, Res, captures the amount of artifact remaining in the reconstructed signal after cleaning.

III. Results and Discussion

We compared the proposed method to the baseline in the BigOpt 2015. The evaluation includes CPU time, Pareto Optimality, Score from BigOpt, Information Loss, and Residual in the solutions. Figure 2, 3 and Table V, VIIb, VIIa, VI summarized the result of this experiment. Solutions P1 to P5 are ordered based on the polar angle in the Pareto front.

A. CPU Time

The optimization was done on Intel Xeon Phi coprocessor 3120 series with OPENMP [24] that provides parallel computing interface to speed up expensive function evaluation. 16 cores were used in this experiment. The time taken of each problem were summarized in Table V. On average, time domain representation requires less time to complete a run. As the threshold th decreases, the number of variables involved increases and hence, the computational time of the EC method. The result suggests that if frequency representation is used, proper thresholding of dimension reduction is important for time efficient optimization.

B. Score from BigOpt

The Score from each problem can be found in Table VIIb, VIIa, VI The best score is achieved by MOEA/D with $th = 0.05$. As the size of problem scales up, frequency domain representation with threshold 0.05 always obtained a better score. This corresponds to better Pareto optimality of the solutions compared to the baseline which cannot be achieved by time domain representation. Negative score was found in time domain optimization because one and more of the solutions were poorer than the baseline. Comparing MOEAD-D with NSGA-II, it is observed that MOEA/D always obtain a better score than NSGA-II in both noise free and with noise cases. th = 0.025 and 0 are not giving promising results as the computational resources are limited in this experiment with small population size and limited number of function evaluation.

TABLE V: The mean and standard deviation of the CPU time (Seconds)

Problem	Threshold	Th = 0	Th = 0.025	Th = 0.05	Time domain
D4	MOEAD	1680.80 (22.26)	296.56 (1.06)	233.13 (1.77)	154.21 (0.30)
	NSGA-II	1723.69 (16.74)	298.60 (0.48)	232.60 (1.35)	158.88 (0.19)
D4N	MOEAD	1702.09 (20.93)	288.46 (0.77)	229.38 (0.52)	153.53 (0.39)
	NSGA-II	1717.59 (10.52)	292.74 (0.65)	234.39 (1.42)	159.21 (0.27)
D12	MOEAD	15126.44 (332.87)	3138.82 (32.85)	1822.11 (8.72)	1256.84 (2.57)
	NSGA-II	15281.03 (61.34)	3138.88 (5.54)	1815.73 (1.53)	1307.58 (115.77)
D12N	MOEAD	15292.75 (174.17)	3215.39 (40.50)	1813.62 (7.62)	1265.91 (2.57)
	NSGA-II	14535.01 (116.54)	3187.63 (3.67)	1803.14 (5.48)	1293.20 (4.81)
D19	MOEAD	-	-	5683.98 (9.61)	5203.60 (47.52)
	NSGA-II	-	-	-	-
D19N	MOEAD	-	-	5807.91 (12.18)	5125.15 (8.71)
	NSGA-II	-	-	-	-

C. Information Loss, and Residual in the solutions

Generally, as observed in Figure 2, 3, the information loss and residual increases as the we move along the solutions from P1 to P5.

TABLE VI: The objective function, Information Loss and Score for Dataset C (D19 and D19N) with MOEA/D

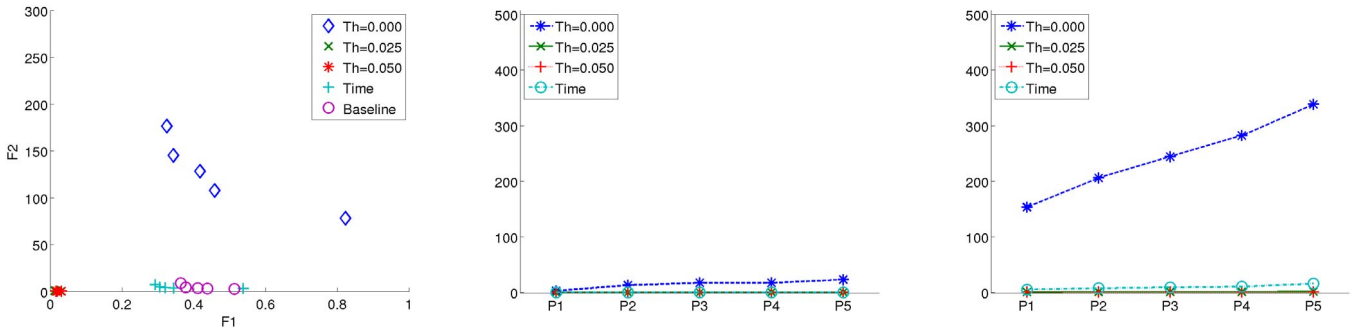
D19, MOEA/D					
Threshold	Sol	F1	F2	InfoLoss	Residue
0	P1	0.211	0.665	1.993	1.069
	P2	0.204	0.726	1.430	1.959
	P3	0.203	0.777	1.288	2.371
	P4	0.203	0.809	1.263	2.580
	P5	0.202	1.009	1.086	3.810
	Score	**19.28**			
Time	P1	3.837	15.288	3.474	13.549
	P2	2.782	20.292	1.451	21.920
	P3	2.672	31.922	0.934	36.001
	P4	2.442	41.791	0.692	47.963
	P5	2.375	53.384	0.837	61.546
	Score	**-140272**			
D19N, MOEA/D					
Threshold	Sol	P1	F2	InfoLoss	Residue
0	P1	0.190	0.644	1.749	1.198
	P2	0.182	0.712	1.198	2.166
	P3	0.181	0.802	1.006	2.862
	P4	0.180	1.029	0.909	4.209
	P5	0.180	1.854	1.612	8.007
	Score	**16.87**			
Time	P1	4.088	17.260	3.682	15.629
	P2	2.850	23.114	1.352	25.206
	P3	2.733	27.714	1.081	30.886
	P4	2.620	46.044	0.643	53.200
	P5	2.531	66.761	0.729	77.394
	Score	**-159886**			

TABLE VIIa: The objective function, Information Loss and Score for Dataset A (D4 and D4N) with MOEA/D and NSGA-II

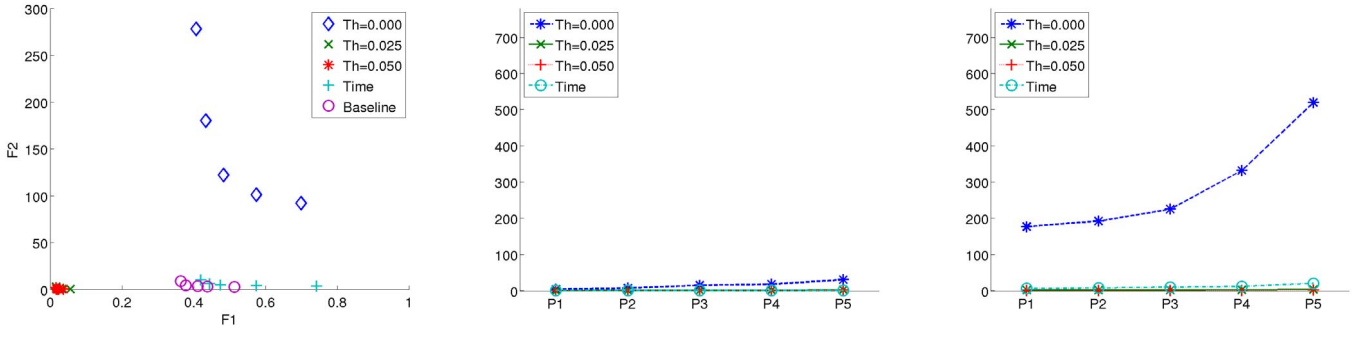
		D4								D4N							
		MOEA/D				NSGA-II				MOEA/D				NSGA-II			
Threshold	Sol	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual
0	P1	0.823	78.170	2.532	153.440	0.699	92.117	3.452	176.593	0.563	73.280	3.377	140.171	0.674	79.394	2.862	147.958
	P2	0.459	107.598	12.949	205.822	0.574	100.874	6.839	191.959	0.371	90.048	10.238	167.380	0.586	81.372	4.311	151.215
	P3	0.418	128.352	17.166	244.010	0.484	121.935	14.899	225.124	0.348	100.668	12.119	186.395	0.520	87.100	8.211	159.465
	P4	0.344	144.877	17.367	282.246	0.434	179.954	17.312	331.923	0.334	110.104	14.180	203.308	0.456	99.957	11.640	183.218
	P5	0.325	176.315	23.244	338.342	0.407	278.070	30.113	520.019	0.325	137.509	18.363	253.605	0.427	150.716	15.358	280.009
	Score	**-304711**				**-750343**				**-70402**				**-476586**			
0.025	P1	0.030	0.168	0.056	0.680	0.056	0.285	0.060	1.238	0.033	0.187	0.046	0.846	0.074	0.368	0.100	1.609
	P2	0.018	0.272	0.062	0.899	0.032	0.394	0.144	1.340	0.018	0.290	0.076	0.898	0.042	0.504	0.141	1.844
	P3	0.016	0.331	0.096	0.948	0.024	0.640	0.308	1.507	0.016	0.378	0.125	0.990	0.030	0.719	0.448	1.858
	P4	0.014	0.403	0.188	0.911	0.018	1.472	0.972	2.024	0.014	0.470	0.171	1.119	0.020	1.589	1.165	2.209
	P5	0.012	0.695	0.377	1.128	0.016	2.817	1.912	3.202	0.014	0.865	0.501	1.368	0.017	3.612	2.775	3.771
	Score	**20.86**				**17.13**				**19.88**				**15.28**			
0.05	P1	0.032	0.114	0.014	0.536	0.037	0.108	0.008	0.601	0.035	0.122	0.014	0.608	0.039	0.124	0.017	0.601
	P2	0.020	0.140	0.009	0.410	0.024	0.135	0.017	0.587	0.019	0.154	0.012	0.426	0.028	0.136	0.017	0.566
	P3	0.017	0.161	0.013	0.464	0.021	0.162	0.029	0.592	0.018	0.172	0.023	0.497	0.021	0.174	0.014	0.550
	P4	0.017	0.179	0.027	0.479	0.017	0.283	0.066	0.706	0.017	0.191	0.030	0.514	0.018	0.338	0.126	0.709
	P5	0.016	0.265	0.075	0.568	0.017	1.878	1.462	1.955	0.017	0.273	0.077	0.628	0.017	2.332	1.896	2.416
	Score	**21.86**				**20.16**				**21.15**				**18.96**			
Time	P1	0.539	3.009	0.344	5.228	0.742	3.680	0.705	5.829	0.623	2.960	0.468	4.849	0.625	3.470	0.696	5.916
	P2	0.345	3.498	0.209	7.599	0.575	3.862	0.481	7.013	0.406	3.899	0.221	8.205	0.521	3.760	0.510	6.978
	P3	0.321	3.942	0.162	8.715	0.475	4.661	0.335	9.129	0.382	4.490	0.197	9.736	0.456	4.763	0.219	9.301
	P4	0.306	4.627	0.141	10.271	0.445	5.895	0.251	11.601	0.364	5.658	0.208	12.464	0.420	6.123	0.233	12.255
	P5	0.293	7.299	0.204	15.819	0.420	9.973	0.335	19.932	0.353	8.687	0.299	17.932	0.400	9.658	0.386	19.585
	Score	**-194.96**				**-5511.35**				**-143.27**				**-5843.39**			

TABLE VIIb: The objective function, Information Loss and Score for Dataset B (D12 and D12N) with MOEA/D and NSGA-II

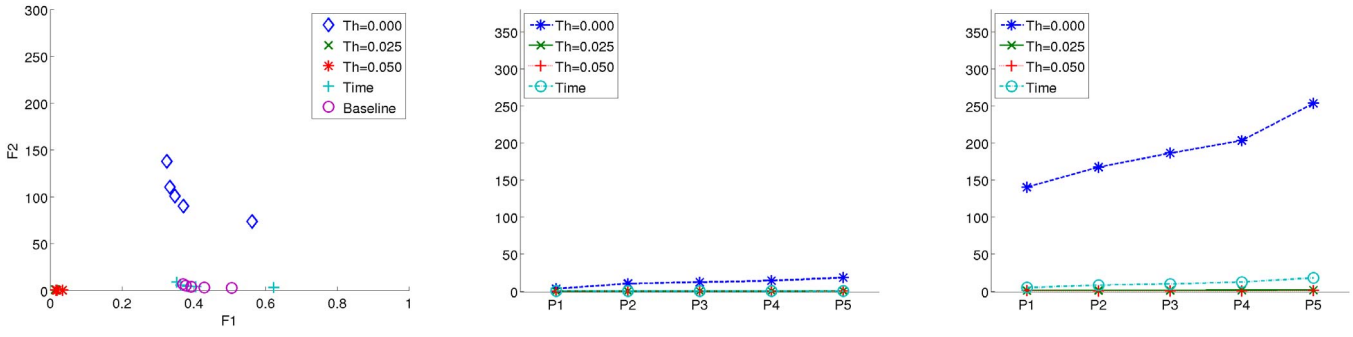
		D12								D12N							
		MOEA/D				NSGA-II				MOEA/D				NSGA-II			
Threshold	Sol	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual	F1	F2	InfoLoss	Residual
0	P1	0.979	176.654	12.308	589.750	0.883	186.250	12.002	625.770	0.918	174.518	9.243	600.345	0.890	204.460	10.117	697.346
	P2	0.605	247.510	45.027	815.266	0.779	206.661	14.704	698.705	0.582	229.012	39.440	765.614	0.794	215.631	13.213	735.573
	P3	0.548	304.343	75.457	984.884	0.670	266.314	28.669	891.938	0.530	266.363	58.399	879.617	0.697	271.039	28.373	916.233
	P4	0.527	356.029	89.850	1133.357	0.636	380.354	54.507	1253.076	0.513	299.203	71.616	982.420	0.639	415.658	53.487	1395.421
	P5	0.510	444.908	118.858	1408.322	0.609	857.066	143.635	2842.643	0.500	421.581	111.868	1372.658	0.612	607.723	97.774	2026.041
	Score	**-173413**				**-648289**				**-171268**				**-680110**			
0.025	P1	0.144	9.064	8.965	22.647	0.386	15.397	6.779	46.724	0.189	9.486	6.966	27.366	0.357	15.475	7.004	47.973
	P2	0.053	11.002	19.532	19.185	0.206	19.220	15.114	55.681	0.054	11.894	20.766	22.554	0.158	24.852	26.319	63.778
	P3	0.039	12.692	25.529	19.175	0.113	26.645	34.067	63.547	0.040	13.916	27.752	22.839	0.077	49.803	83.430	95.984
	P4	0.031	15.049	32.853	20.135	0.059	61.800	115.051	103.023	0.031	16.325	35.542	23.737	0.065	76.402	136.591	137.121
	P5	0.023	28.879	70.484	30.706	0.051	137.439	277.990	205.256	0.022	30.423	74.838	34.611	0.055	151.791	304.596	237.361
	Score	**0.00**				**0.00**				**0.00**				**0.00**			
0.05	P1	0.082	0.443	0.366	0.841	0.126	0.633	0.455	1.389	0.090	0.482	0.525	0.807	0.138	0.764	0.503	1.744
	P2	0.075	0.499	0.272	1.146	0.086	0.955	0.610	2.397	0.076	0.578	0.456	1.226	0.088	1.470	1.559	3.096
	P3	0.074	0.546	0.276	1.309	0.075	2.667	3.549	5.314	0.074	0.671	0.610	1.381	0.075	3.905	6.418	6.622
	P4	0.073	0.662	0.358	1.636	0.073	5.236	8.668	8.895	0.073	0.755	0.729	1.549	0.072	9.022	18.032	12.165
	P5	0.073	0.951	0.693	2.307	0.073	11.090	20.919	16.573	0.071	1.423	1.795	2.792	0.071	21.950	47.637	26.104
	Score	**20.01**				**6.94**				**19.83**				**5.46**			
Time	P1	2.323	9.673	1.980	8.745	2.399	11.857	2.138	11.285	2.486	9.897	2.198	8.887	2.289	12.051	1.837	11.784
	P2	1.737	13.676	0.798	14.867	2.121	12.408	1.545	12.720	1.870	12.972	1.090	13.850	2.109	12.689	1.534	12.937
	P3	1.652	19.436	0.508	21.911	1.854	14.564	0.966	15.839	1.656	15.474	0.725	17.215	1.931	15.095	1.052	16.329
	P4	1.578	30.518	0.655	34.769	1.759	18.953	0.644	21.265	1.597	18.121	0.598	20.531	1.815	18.863	0.833	21.013
	P5	1.508	42.155	0.946	47.920	1.716	29.760	0.424	33.963	1.549	32.169	0.646	37.263	1.761	33.586	0.671	38.482
	Score	**-92899**				**-65090**				**-65545**				**-69196**			



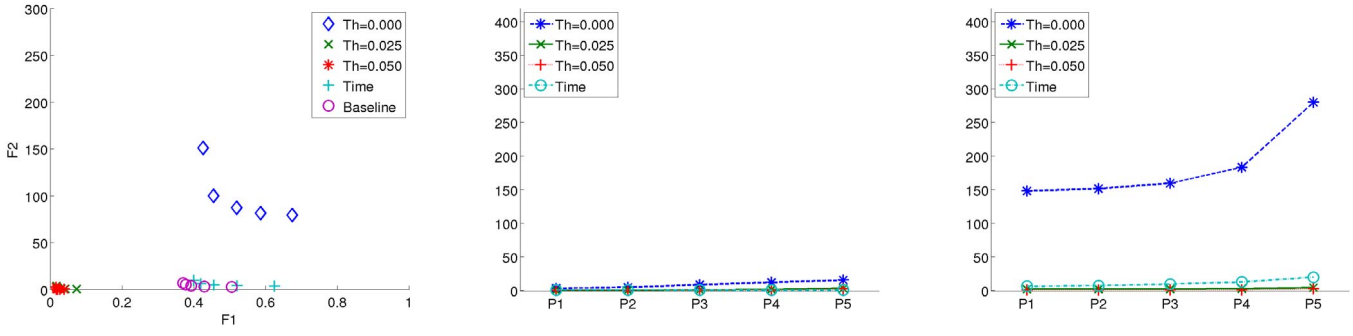
(a) Data set D4 solved by MOEA/D



(b) Data set D4 solved by NSGA-II

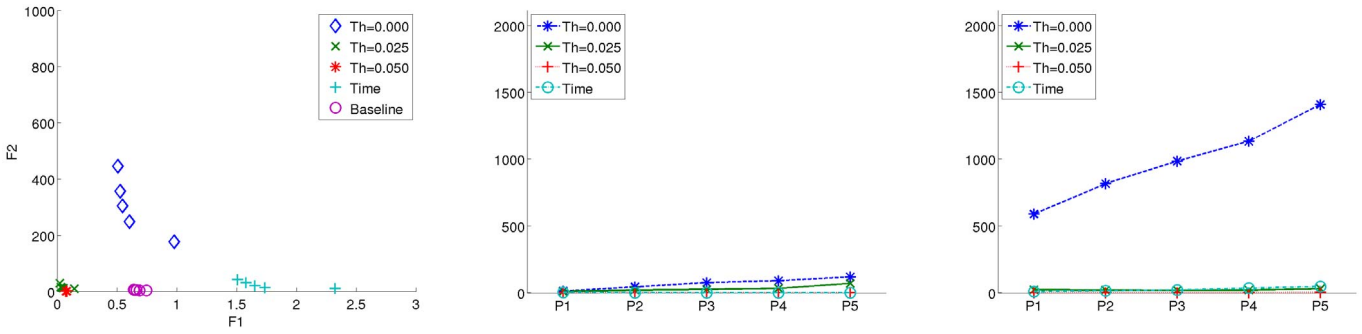


(c) Data set D4N solved by MOEA/D

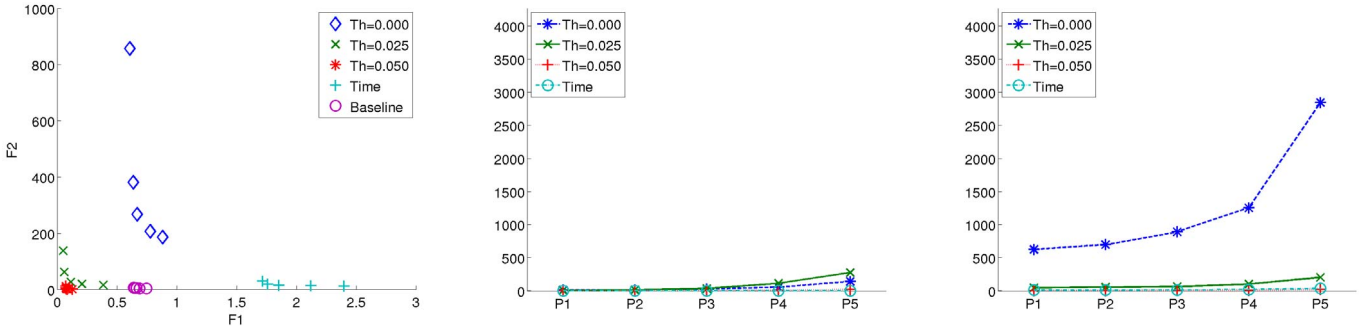


(d) Data set D4N solved by NSGA-II

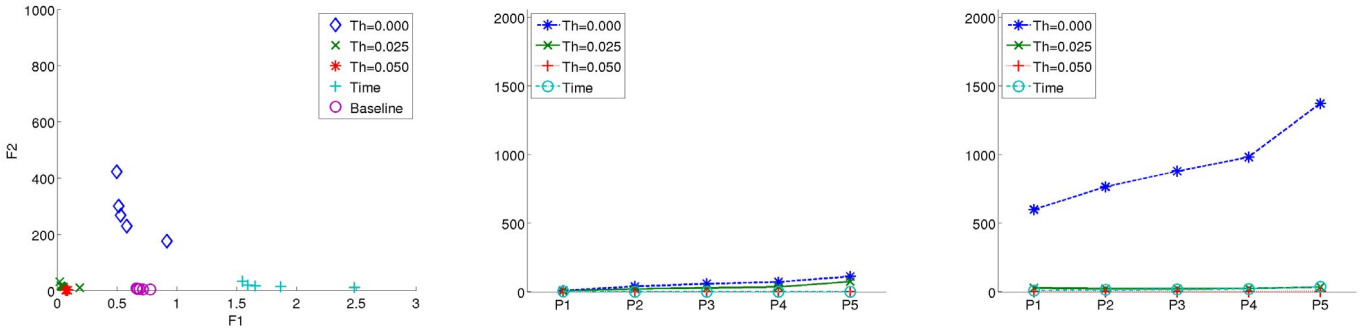
Fig. 2: The Pareto Front (Right Column), Information Loss (Center) and Residual (Left) for Dataset A (D4 and D4N) with MOEA/D and NSGA-II



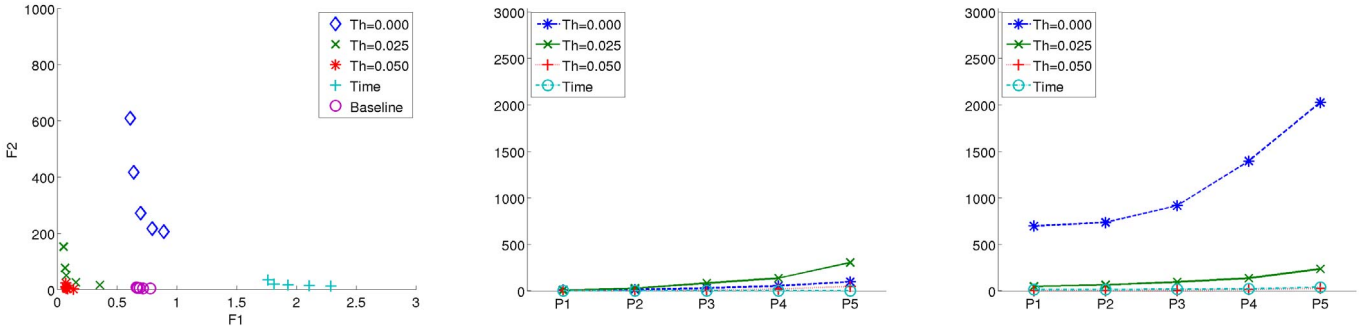
(a) Data set D12 solved by MOEA/D



(b) Data set D12 solved by NSGA-II



(c) Data set D12N solved by MOEA/D



(d) Data set D12N solved by NSGA-II

Fig. 3: The Pareto Front (Right Column), Information Loss (Center) and Residual (Left) for Dataset B (D12 and D12N) with MOEA/D and NSGA-II

As observed in Figure 2, 3 and Table VIIb, VIIa, VI, it is found that P1 with MOEA/D and $th = 0.05$ always achieve the least information loss and residual in the signal. This supports the contribution of solving the MOP using the frequency domain representation with proper dimensionality reduction and MO formulation. Hence, only MOEA/D with $th = 0.05$ was tried on D19 and D19N.

IV. Conclusion

In this paper, we presented a multi-objective formulation of the automated EEG cleaning problem. With the high epistasis in the problem, we proposed to optimize the problem in the frequency domain. This allowed us to solve multi-objective optimization problems with up to 4864 variables. Optimization in the frequency domain facilitates dimensionality reduction, generates better quality solutions for EEG cleaning as measured by information loss, residual, and BigOpt score.

We also presented the details on how the benchmarks for the BigOpt CEC 2015 competition were generated. This work raises the challenges that vast amount of data in EEG experiments (big data) creates very large scale optimization problems. A one second of data took up to 1.5 hours to optimize. This raises many challenges for real-time processing of information. While optimization in the frequency domain had better time performance than the frequency domain, the time to optimize is orders of magnitude greater than the speed data are collected.

In future years, we will increase the size of the dataset to bring it closer to the real-world Big Data challenges for this problem domain. Nevertheless, even with the current smaller scales we used in this paper, existing state-of-art EC methods are extremely slow. Methods that can speed up this optimization problem are well needed. Several aspects of EC can be considered, but not limited to dimension reduction technique, novel operator design and parallel EC for Big Optimization.

Acknowledgment

The second author acknowledges funding from the University of New South Wales, Australia, that allowed him the time to conduct this work in Singapore. This work was also supported by the Singapore Ministry of Education Academic Research Fund Tier 1 under the project R-263-000-A12-112.

References

[1] K. Deb, *Optimization for engineering design: Algorithms and examples*. PHI Learning Pvt. Ltd., 2012.

[2] P. J. Fleming and R. C. Purshouse, "Evolutionary algorithms in control systems engineering: a survey," *Control engineering practice*, vol. 10, no. 11, pp. 1223–1241, 2002.

[3] Y. S. Ong, P. Nair, A. Keane, and K. Wong, "Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems," in *Knowledge Incorporation in Evolutionary Computation*. Springer, 2005, pp. 307–331.

[4] M. G. C. Tapia and C. A. C. Coello, "Applications of multi-objective evolutionary algorithms in economics and finance: A survey." in *IEEE congress on evolutionary computation*, vol. 7, 2007, pp. 532–539.

[5] S.-H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert systems with applications*, vol. 31, no. 3, pp. 652–660, 2006.

[6] J. H. Ang, K. Tan, and A. Mamun, "An evolutionary memetic algorithm for rule extraction," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1302–1315, 2010.

[7] K. C. Tan, Q. Yu, and J. H. Ang, "A coevolutionary algorithm for rules discovery in data mining," *International Journal of Systems Science*, vol. 37, no. 12, pp. 835–864, 2006.

[8] K. C. Tan, E. J. Teoh, Q. Yu, and K. Goh, "A hybrid evolutionary algorithm for attribute selection in data mining," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8616–8630, 2009.

[9] S. K. Pal, S. Bandyopadhyay, and S. S. Ray, "Evolutionary computation in bioinformatics: A review," *Systems, man, and cybernetics, Part c: Applications and reviews, ieee transactions on*, vol. 36, no. 5, pp. 601–615, 2006.

[10] M. P. Wachowiak, R. Smolíková, Y. Zheng, J. M. Zurada, and A. S. Elmaghraby, "An approach to multimodal biomedical image registration utilizing particle swarm optimization," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 3, pp. 289–301, 2004.

[11] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "A hybrid genetic algorithm approach for improving the performance of the lf-asd brain computer interface," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. v–345.

[12] Y. Fan, T. Jiang, and D. J. Evans, "Volumetric segmentation of brain images using parallel genetic algorithms," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 8, pp. 904–909, 2002.

[13] J.-Y. Yeh and J. Fu, "A hierarchical genetic algorithm for segmentation of multi-spectral human-brain mri," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1285–1295, 2008.

[14] C. T. Choi, W.-D. Lai, and Y.-B. Chen, "Optimization of cochlear implant electrode array using genetic algorithms and computational neuroscience models," *Magnetics, IEEE Transactions on*, vol. 40, no. 2, pp. 639–642, 2004.

[15] H. Abbass, J. Tang, R. Amin, M. Ellejmi, and S. Kirby, "The computational air traffic control brain: Computational red teaming and big data for real-time seamless brain-traffic integration," *Journal of Air Traffic Control*, vol. 56, no. 2, pp. 10–17, 2014.

[16] H. A. Abbass, J. Tang, R. Amin, M. Ellejmi, and S. Kirby, "Augmented cognition using real-time eeg-based adaptive strategies for air traffic control," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1. SAGE Publications, 2014, pp. 230–234.

[17] R. Vigário, V. Jousmäki, M. Hämäläninen, R. Hari, and E. Oja, "Independent component analysis for identification of artifacts in magnetoencephalographic recordings," *Advances in neural information processing systems*, pp. 229–235, 1998.

[18] S. K. Goh, H. A. Abbass, K. C. Tan, and A. Al Mamun, "Artifact removal from EEG using a multi-objective independent component analysis model," in *Neural Information Processing*. Springer, 2014, pp. 570–577.

[19] S. K. Goh, H. A. Abbass, K. C. Tan, and A. Al-Mamun, "Decompositional independent component analysis using multi-objective optimization," *Soft Computing*, pp. 1–16, 2015.

[20] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated pareto sets, moea/d and nsga-ii," *Evolutionary Computation, IEEE Transactions on*, vol. 13, no. 2, pp. 284–302, 2009.

[21] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 2, pp. 182–197, 2002.

[22] H. A. Abbass, "Calibrating independent component analysis with Laplacian reference for real-time EEG artifact removal," in *Neural Information Processing*. Springer, 2014, pp. 68–75.

[23] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[24] L. Dagum and R. Menon, "Openmp: an industry standard api for shared-memory programming," *Computational Science & Engineering, IEEE*, vol. 5, no. 1, pp. 46–55, 1998.