FP-AK-QIEAR-$\mathbb{R}$ in protein folding application

Josimar Edinson Chire Saire
Research Center in Computer Science
Universidad Católica San Pablo
Arequipa,Peru
josimar.chire@ucsp.edu.pe

Yván Túpac
Research Center in Computer Science
Universidad Católica San Pablo
Arequipa, Peru
Email: ytupac@ucsp.pe

***Abstract*—There are many Evolutionary Algorithms which main features are: population, evolutionary operations (crossover, mate, mutation and others). Most of them are based on randomness and follow a criteria using fitness like selector. The proposal uses probability density function according to best of initial population to sample new population and save better individuals iteratively. Then using centroid criteria sample for every dimension and get better individuals. It had good results with benchmark functions. A real application was performed with experiments in protein folding and it showed good results.**

***Index Terms*—Quantum-inspired evolutionary algorithms; Evolutionary algorithms; Particle filters; Akima interpolation; pdf estimation;**

I. Introduction

Quantum Inspired Evolutionary Algorithm (QIEA) are a populacional metaheuristic based on the concept of multiple universes superposition from quantum physics, then it can be used for optimization process with great intrinsic parallelism. A first approach, the QIEA-$\mathbb{B}$ model [1] uses *q-bits* as the information unit creating quantum individuals behaving as generators of classical individuals and uses an update operator so probabilistically to get better individuals along several iterations.

QIEA-$\mathbb{R}$ [2], [3] was proposed for numerical problems with real codification using the same principle of quantum inspiration. The updating of quantum individuals are based on uniform probability density function to generate better classic individuals for real codification [2]. Using uniform pdf is easy, but it is not the best choice because the adaptability to the probability of data nature (unimodal, multimodal) can be limited.

Particle Filter is a set of estimation algorithms of posteriori probability density function based in equations of bayesian recursion. It is a useful method to generate samples of required distribution with no need assumptions about state distribution. This feature solves the need of any QIEA model: to generate classical individuals.

A combined proposal using QIEA-$\mathbb{R}$ and a mechanism inspired in Particle filter, multilinear regression for pdf estimation using rewarding criteria called FP-QIEA-$\mathbb{R}$[4].

The proposal of this work is to use a technique inspired on Particle Filter , Akima interpolation and rewarding criteria to generate classical population and get the best individuals for global search and use centroid for local search. Then get better results in convergence and a complete domain exploring.

To evaluate this proposal, some benchmark functions(Ackley, Rastrigin, Rosenbrock, Schwefel, Sphere) commonly used in evaluation of optimization algorithms will be used and some experiments with protein folding(perspective of optimization problem).

This paper is organized as follows: Section 2 describes a literature review about QIEA-$\mathbb{R}$, Section 3, describes QIEA-$\mathbb{R}$. Section 4, describe the Particle Filter. Section 5, describes pdf estimation and Akima interpolation. Section 6, describes the FP-AK-QIEA-$\mathbb{R}$ Model. Section 7 shows the experiments and results, and Section 8 states the conclusions.

II. Literature Review

GQA (Genetic Quantum Algorithm) was the first proposal, inspired on the concept and principles of quantum computing using *q-bits* and superposition of states from the quantum mechanics. A (*q-bit*) presents the exploration as much as the exploitation features. In GQA an observation and update stages are presented. The experiments were performed using the Knapsack Problem and shows that QGA has the ability of global search caused by probabilistic representation and show better convergence than CGA (Classical Genetic Algorithm) [1].

QIEA with real codification (QIEA-$\mathbb{R}$), this model has the quantum individual based on uniform pdfs, and the classical individuals are real vectors. An observation and specific operators for this encoding are proposed [2] similarly to QIEA-$\mathbb{B}$. The updating steps are performed using the "1/5 rule" [5] to make wider or narrower the uniform pulse (pdf) and movements of the pulce center. The experiments were performed using benchmark functions (Griewank, Schwefel, Rosenbrock, Sphere, Ackley, Rastrigin and others). The outcomes demonstrate the good performance of the algorithm outperforming the (QIEA-$\mathbb{B}$). It has good results on the most of benchmark functions except in Ackley and Rastrigin function and when the number of evaluations is greater it has the worst results compared with traditional methods in Ackley function, it could be because of the presence of huge number of local optima.

FP-QIEAR-$\mathbb{R}$, this model exchange iterations of QIEA-$\mathbb{R}$ with a mechanism inspired in the concept of Particle filters for classical individual representation with an weight associated(probability) related to fitness function, then uses a rewarding criteria to set greater probability to best individuals

and uses multilinear regression to estimate the probability density function and uses centroid for local search iteratively. It had better results than QIEA-R to get the optimal in all benchmark functions.

Proteins are the basic structures of all living beings [6] and they are composed by a chain of amino acids. Protein folding is the process by chains are transformed in compact structures that perform biological functions. The AB Off-lattice model was introduced by [7] to represent protein structure, there is bond and torsional angles defined in $-180°, 180$ to evaluate the structure one energy function was proposed. There is a set of synthetic protein sequences based in Fibonacci sequence and used for other researchers [8], [9] . Some of them try to solve with deterministic and other with meta-heuristics approaches [10], [11].

III. The QIEA-$\mathbb{R}$ model

The QIEA-$\mathbb{R}$ model, proposed by Da Cruz [2] is composed by the following steps: generation of quantum population, generation of classic population by observing quantum population, and update of quantum population.

Quantum Representation

Quantum individual represent the superposition of possible states. In this model, the set of observable states is continuous.

Once initialized the quantum population $\mathbf{Q}_0$, the main loop of evolutionary process is started and continues as follow:

Observation

Consists in generate classical individuals from the quantum individuals, using the pdf $p_{ij}(x)$, the accumulated probability function and a random numbers generator $\mathbf{U}(0,1)$ to generate several real values.

Updating

The QIEA-$\mathbb{R}$ can be reduce or increase the search space of any quantum individual according to the fitness of classical population obtained during the observation process and using the "rule of 1/5" [5]. If less 20% of classical population from current generation has fitness better than previous generation, the width of gen is reduced, if is greater than 20% the width is increased and if it is equal to 20% the width is not modified, Eq.1, shows the 1/5 rule.

$$\sigma_{ij} = \begin{cases} \sigma * \delta & \varphi < 1/5 \\ \sigma_{ij}/\delta & \varphi > 1/5 \\ \sigma_{ij} & \varphi = 1/5 \end{cases} \tag{1}$$

IV. Particle Filter

Particle Filter (PF) is an approach to solve filtering problem (i.e. estimating the true value of a system from some observations, possibly noisy). The name *particle filters* was introduced in [12] in 1996 inspired in mean field interacting particle methods (Monte Carlo algorithms for sampling from a sequence of probability distributions satisfying a nonlinear equation).

Particle filter uses a set of particles to represent the posterior distribution of some stochastic process, given some noisy observations. Particle filtering provides a way to generate samples from the required distribution without requiring assumptions about the state distributions (non linear) [13]. Samples from the distribution are represented by a set of particles; each particle has a weight that represents the probability of the particle be sampled from the pdf.

V. Pdf estimation

Is a field of numerical analysis focused in the researching of how functions can be approximated for a kind of specific functions (polynomial functions), those has properties of low computational cost, continuity and other. Besides it is not necessary to know the mathematical formula, just a set of point like $(x, f(x))$ and use interpolation, extrapolation, regression and curves fit techniques.

A. Probability Density Function

Definition 1: Be X a random variable, the pdf of X is a function $f(X)$ so for two any numbers a y b with $a \leq b$:

$$P(a \leq X \leq b) = \int_a^b f(X)dx \tag{2}$$

The probability of X in the interval $[a, b]$ is the area under the interval and over graphic of pdf. Besides it must satisfy next conditions:

- $f(x) \geq 0 \quad , \quad$ para todo $\quad x$
- $\int_{-\infty}^{\infty} P_{ij}(x)dx = 1 \quad , \quad P_{ij}(x) \geq 0$

By instance, the next graphic shows pdf of normal distribution:

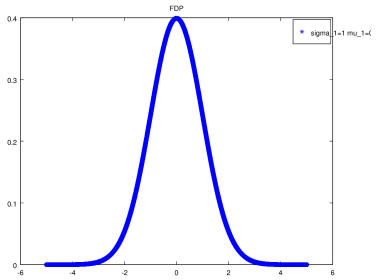

Fig. 1. pdf of normal distribution with σ = 1 , μ = 0

B. Cumulative Density Function

Definition 2: The cdf(cumulative density function) of random variable X is given by:

$$F(X) = P(X \leq x) \tag{3}$$

It can be expressed for the integral of the pdf: $F(x) = \int_{-\infty}^{x} f_X(t)\,dt$

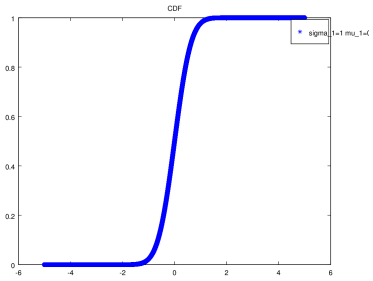

Fig. 2. cdf of normal distribution with σ = 1 , μ = 0

C. Akima interpolation

Akima interpolation is continuous sup-spline interpolation. It is build for pieces from third grade polynomials. Only near neighbors are used to calculate interpolation coefficient polynomial. There is no need of equations and it is efficient. This method uses a reduced number of point, for this reason it avoids strange curves in the resultant curve [14].

By instance it is an approximation of some functions:

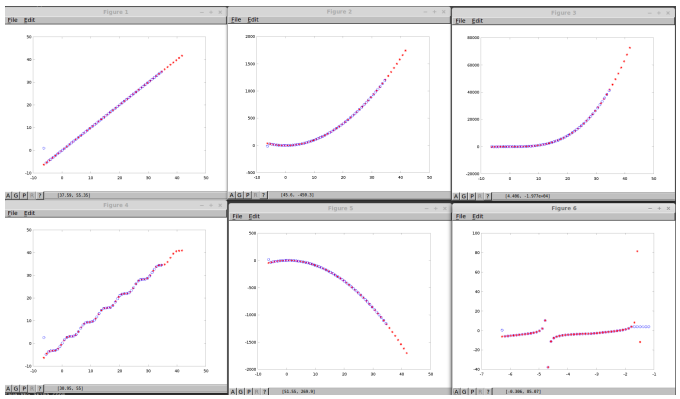

Fig. 3. Function approximation using Akima interpolation

VI. FP-AK-$\mathbb{R}$ MODEL

The proposal uses quantum population for classical generation during the initialization. Then probability density function is generated using fitness function and rewarding criteria for every dimension. The best of every iteration is saved and using centroid explore local individuals.

A. Initialization

For the first classical population it is used the classical generation X similar to QIEA-$\mathbb{R}$. Using the fitness function to evaluate the classical population X, we get the vector Y. Using the criteria of particle filter, it is associated a probability then the vector Y is normalized and reward the best individuals(more probability of being chosen) and Y is normalized to satisfy the sum of probabilities equal to 1. Then this vector Y represents our pdf.

B. Create CDF

This process is repeated T iterations: We have a vector with X_d with a probability associated, we need to order X_d to satisfy the property of cdf(cumulative density function). Generate cdf, then we have the points($X_{ordered-d}$, cdf). A vector of probabilities is generated $prob$, akima interpolation is used to sample a new vector X_d. This process is performed for every dimension until to get a new population X_{est}, X_{est} is evaluated and the best individual is stored in X_{global}.

C. Searching around centroid

Using X_global a centroid for every dimension is obtained. A new individual is calculated sampling around centroid of every dimension.

Algorithm 1 FP-AK-QIEAR using centroids for local search

Initialization

1: Generate quantum population from QIEA-$\mathbb{R}$
2: Sampling using quantum population to get classical population
3: Evaluation and normalization , use rewarding criteria and normalize to get pdf
4: $t \leftarrow 1$
5: **while** $t \leq T$ **do**
6: $d \leftarrow 1$
7: **while** $d \leq D$ **do**
8: Create pdf and calculate cdf , generate probability vector $prob$
9: Using cdf and respective X value to get approximated X values(x_{estim}) for $prob$ by Akima interpolation
10: Save x_{estim} in X_{est}
11: **end while**
12: Evaluation of X_{est} and save the best X_{global}
13: **end while**
14: Calculate centroid for every dimension and save in avg
15: $sigma(radius) \leftarrow 0.01$
16: **while** $d \leq D$ **do**
17: Sample using centroid and sigma radius in the domain [avg-$sigma$, avg + $sigma$] and save in X_{global}
18: **end while**
19: Return the best of X_{global}

VII. PROTEIN FOLDING

Proteins are the basic structures of all living beings [6], they are composed of a chain of amino acids that are linked together by means of peptide bonds. Protein folding is the process by which a polypeptide chain is transformed into a compact structure that performs some biological function. Its known that better understanding the protein folding process can result in important medical advancements and development of new drugs.

The AB off-lattice model was introduced by [7] to represent protein structures. The proposal of this model is: to represent every residue by a single interaction site located at the Cα position. These sites are linked by rigid unit-length bonds ($\widehat{b}_i$)

to form the protein structure. The three-dimensional structure of an N - length protein is specified by the N - 1 bond vectors $\widehat{b}_i$, N - 2 bond angles τ i and N - 3 torsional angles α_i . These angles are defined in the range [-180 , 180] degrees.

The enery function is given by [15]:

$$E(\widehat{b}_i; \sigma) = E_{angles} + E_{torsion} + E_{LJ} \quad (4)$$

There is a set of synthetic protein sequences based in Fibonacci sequence and used for other researchers [8], [9] . Some of them try to solve with deterministic and other with meta-heuristics approaches [10], [11].

VIII. Experiments and Results

A. Benchmark Functions

1) Ackley: This function is characterized by a nearly flat outer region and a large hole at the center. Ackley function poses a risk for optimization algorithms, particularly hill climbing based algorithms, which can be trapped in one of its local optima.

2) Rastrigin: The Rastrigin function has several local minima. It is highly multi modal, but locations of the minimal are regularly distributed.

3) Rosenbrock: This function is unimodal, and the global minimum lies in a narrow, parabolic valley. However, even though this valley is easy to find, convergence to the minimum is difficult.

4) Schwefel: The Schwefel function is very complex, with many local minima.

5) Sphere: It is continuous, convex and unimodal function.

B. Protein folding

To encode the candidate solutions the proposal of [11] was used. Protein folding has many local optimal then to overcome this challenge we use a modification of FP-AK-QIEAR.

To improve this proposal, an updating of pdf in every iterations is performed and for local search a mutation operator.

Algorithm 2 FP-AK-QIEAR with mutation for local search

Initialization

1: Generate quantum population from QIEA-$\mathbb{R}$
2: Sampling using quantum population to get classical population
3: Evaluation and normalization, use rewarding criteria and normalize to get pdf
4: $t \leftarrow 1$
5: **while** $t \leq T$ **do**
6: The same from step 6 to 11 of 1
7: Evaluation of X_{est}
8: Join X and X_{est} to choose the best, update X and evaluate Y
9: Find the best individual and apply mutation according $prob_{mutation} \leq 0.1$ for every dimension and save in X_{mut}
10: Find the worst individual in X and replace with X_{mut}
11: Evaluate Y, normalize, use rewarding criteria and normalize to get pdf
12: **end while**

TABLE I
Benchmark sequences for the 3D-AB off-lattice model

Function	QIEAR
13	ABBABBABABBAB
21	BABABBABABBABBABABBAB
34	ABBABBABABBABBABABBABABBABBABABBAB
55	BABABBABABBABBABABBABABBABBABABBAB BABABBABABBABBABABBAB

C. Experiments

In order to evaluate and compare QIEA-$\mathbb{R}$ and FP-QIEA-$\mathbb{R}$, some experiments were performed under the following benchmark functions: Ackley, Rastrigin, Rosenbrock, Schwefel and Sphere. One feature of these benchmark functions is that can be configured for $n > 2$ dimensional variables. Thus, in this work, the dimensionality used was $n = 30$. The number of iterations in QIEA-$\mathbb{R}$ and FP-QIEA-$\mathbb{R}$ are 75.

A shifting(k) of the optimal is used in the benchmark functions to set a challenge for the models according to: Ackley(k = 20), Rastrigin(k = 5), Rosenbrock(k = 8), Schwefel(k = 50), Sphere(k = 50).

To minimize the effect of randomness inherent in this type of algorithm, 100 experiments for each parametrization were conducted. The curves shown in the results are averaged. Thus, for each parametrization is possible to display behaviors more approximate to the expected value.

For QIEA-$\mathbb{R}$ the parameters used was the same of the experiments of DaCruz [3], and the parameters for the FP-QIEA-$\mathbb{R}$ were the same of Chire [4].

D. Results

The next tables show the results of the models without shifting:

TABLE II
QIEA-R, FP-QIEA-R,FP-AK-QIEAR, MV = minimal value y SD = standard deviation

Function	QIEAR	FP-QIEAR	FP-AK-QIEAR	f*
Ackley	7.2950E-05	4.2923E-06	0.23518	0
Rastrigin	2.566E-08	1.2164E-11	0.0850	0
Rosenbrock	114617	28.7711	41.5253	0
Schweffel	10674.6	10720.8	10660.7	10660
Sphere	1.1956E-07	5.5865E-11	0.1133	0

Function	QIEAR	FP-QIEAR	FP-AK-QIEAR
Ackley	0.6130	0.0005	0.0444
Rastrigin	10.0424	1.0955E-05	1.3742
Rosenbrock	195430	0.0978	27.8806
Schweffel	34.96	25.2524	0.0061
Sphere	3.8527E-07	7.1729E-06	0.168483

The next tables show the results of the models with shifting:

TABLE III
QIEA-R, FP-QIEA-R,FP-AK-QIEAR, MV = MINIMAL VALUE Y SD = STANDARD DEVIATION

Function	QIEAR	FP-QIEAR	FP-AK-QIEAR	f*
Ackley	5.26682e-05	0.00100964	0.233557	0
Rastrigin	29.8488	29.8488	7.84895	0
Rosenbrock	83992.3	1.35092e+06	179.127	0
Schweffel	10673.8	10710.2	10660.4	10660
Sphere	4.2333e-08	7.65062e-05	0.0505994	0

Function	QIEAR	FP-QIEAR	FP-AK-QIEAR
Ackley	0.355471	1.26979	0.0198786
Rastrigin	0.00499164	0.0352158	1.80287
Rosenbrock	3.15257e+06	2.74045e+06	5.6665
Schweffel	15.4077	43.4874	0.0015106
Sphere	33.5176	313.466	0.0104841

For experiments with protein folding we used 24 experiments and we got the best, median and standard deviation to compare with the results obtained using PSO, GSA, BAT, ABC [11], the best performance was using PSO.

TABLE IV
COMPARISON USING PROTEIN FOLDING

N	PSO		FP-AK-QIEAR	
	E_{avg}	E_{best}	E_{avg}	E_{best}
13	-23.102 $\pm$ 0.93	-24.888	-22.1814 $\pm$ 0.813222	-23.9409
21	-43.047 $\pm$ 2.34	-46.611	-41.9036 $\pm$ 2.74061	-46.0356
34	-70.866 $\pm$ 5.95	-80.409	-68.9528 $\pm$ 3.63282	-75.9811
55	-87.715 $\pm$ 19.27	-115.758	-107.276 $\pm$ 8.86613	-119.652

IX. CONCLUSION

In this paper we present the model FP-AK-QIEA-$\mathbb{R}$ using Akima interpolation for pdf estimation, rewarding criteria to sample around the best individual(global search) and centroid to sample for every dimension(local search). FP-AK-QIEA-$\mathbb{R}$ got goof results when the optimal is shifted from the origin and is more stable than QIEA-$\mathbb{R}$ and FP-AK-QIEAR has good results in protein folding problem. It is open for modifications and improve it's performance.

ACKNOWLEDGEMENTS

The authors would like to thank César M.V. Benítes of Bioinformatic Lab in UTFPR(Universidade Tecnologica Federal do Parana), Brasil. The authors would like to thank CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica), by the financial supporting.

REFERENCES

[1] K.-H. Han and J.-H. Kim, "Genetic quantum algorithm and its application to combinatorial optimization problem," in *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, vol. 2, 2000, pp. 1354–1360 vol.2.

[2] A. da Cruz, "Algoritmos evolutivos com inspiração quântica para problemas com representação numérica," Ph.D. dissertation, Departamento de Engenharia Elétrica, 3 2007.

[3] A. da Cruz, M. Vellasco, and M. Pacheco, "Quantum-inspired evolutionary algorithms applied to numerical optimization problems," in *Evolutionary Computation (CEC), 2010 IEEE Congress on*, July 2010, pp. 1–6.

[4] J. Chire and Y. Tupac, "An approach to real-coded quantum inspired evolutionary algorithm using particles filter," in *2015 Latin America Congress on Computational Intelligence (LA-CCI)*, Oct 2015, pp. 1–6.

[5] R. Ingo, "Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution," *Frommann-Holzboog Verlag*, 1973.

[6] L. Hunter, "Artificial intelligence and molecular biology," L. Hunter, Ed. Menlo Park, CA, USA: American Association for Artificial Intelligence, 1993, ch. Molecular Biology for Computer Scientists, pp. 1–46. [Online]. Available: http://dl.acm.org/citation.cfm?id=166459.166462

[7] F. H. Stillinger and T. Head-Gordon, "Collective aspects of protein folding illustrated by a toy model," *Phys. Rev. E*, vol. 52, pp. 2872–2877, Sep 1995. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.52.2872

[8] H.-P. Hsu, V. Mehra, and P. Grassberger, "Structure optimization in an off-lattice protein model," *Phys. Rev. E*, vol. 68, p. 037703, Sep 2003. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.68.037703

[9] D. Kalegari and H. Lopes, "A differential evolution approach for protein structure optimisation using a 2d off-lattice model." International Journal of Bio-Inspired Computation, 2010.

[10] X. Zhang and W. Cheng, *An Improved Tabu Search Algorithm for 3D Protein Folding Problem*, T.-B. Ho and Z.-H. Zhou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[11] R. S. Parpinelli, C. M. V. Benitiez, J. Cordeiro, and H. S. Lopes, "Performance analysis of swarm intelligence algorithms for the 3d-ab off-lattice protein folding problem," *Multiple-Valued Logic and Soft Computing*, vol. 22, pp. 267–286, 2014.

[12] P. D. Moral, "Nonlinear filtering: Interacting particle resolution," *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, vol. 325, no. 6, pp. 653 – 658, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0764444297847787

[13] L. K. B. and P. R. K., "Analysis of linear / nonlinear models using state estimation algorthems for tracking applications – a review," *International Journal of Science and Advanced Technology*, vol. 1, no. 9, November 2011.

[14] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, Oct. 1970. [Online]. Available: http://doi.acm.org/10.1145/321607.321609

[15] C. P. A. Irback and F. Potthast, "Identification of amino acid se- quences with good folding properties in an off-lattice model," 1997.