

Multi-objective Estimation of Distribution algorithm based on Voronoi and Local search

Elham Mohagheghi
Student of Master of Control Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
E_mohagheghi@stu.um.ac.ir

Mohammad-R. Akbarzadeh-T.
Professor Control Department
SCIIP
Ferdowsi University of Mashhad
Mashhad, Iran
akbazar@um.ac.ir

Abstract In this paper we propose an Estimation of Distribution Algorithm (EDA) equipped with Voronoi and local search based on leader for multi-objective optimization. We introduce an algorithm that can keep the balance between the exploration and exploitation using the local information in the searched areas through the global estimation of distribution algorithm. Moreover, the probability model in EDA, receives special statistical information about the amount of the variables and their important dependency. The proposed algorithm uses the Voronoi diagram in order to produce the probability model. By using this model, there will be a selection based on the area instead of selection based on the individual, and all individual information could use to produce new solution. In the proposed algorithm, considering the simultaneous use of global information about search area, local information of the solutions and the Voronoi based probability model lead to produce more diverse solutions and prevent sticking in local optima. Also, in order to reduce the data dimension, the principle component analysis is proposed. Several benchmarks functions with different complexity like linear and non-linear relationship between the variables, the continuous-discontinuous and convex/non-convex optima fronts use to show the algorithm performance.

Key Words

Estimation of distribution algorithms, Multi-objective optimization, Particle swarm optimization, Voronoi diagram.

I. Introduction

The real world problems usually have a lot of criteria that could formulate as a multi-objective problem. Problems that have some disproportionate and often competitive cost functions usually recognize as the multi-objective optimization problems. In most of the optimization problems there are four challenges: a) The multiple conflicting goals, b) The very complex search area, c) Uncertainty, d) The

dynamic area. Efficient optimization strategies need to face these challenges.

The success of population based algorithm in multi-objective optimization problems reported in many papers. But random recombination in these algorithms may disturb the good building blocks, thus, forecasting to move to the optimize area may be difficult. To solve this problem, the estimation of distribution algorithms, with the idea of exploiting the relationship between variables, introduce as a new computing paradigm in the population based algorithms. In estimation of distribution algorithms crossover operator and mutation are not implemented. Instead, reproduction is done by making a representative probability model, and then new solutions produce by the sampling of this probability model. These algorithms could work with noisy information. Their parallel natures help them to optimize different areas of the search space simultaneously in order to find some promised solutions. But their weakness is that they model the solutions distribution even if they stuck in local optima, and use this probability information as a reference model to produce new generation. To avoid being stuck in local optima, the algorithm should be able to search and keep a diverse set of solutions. According to these algorithms that estimate the probability distribution of the best solutions and using just global information, whenever the kept solutions tend to special area, the algorithm may consider the kept solutions as desired and makes the probability distribution based on them. Thus it has a great importance to keep diversity, specially the ability to produce a set of diverse solutions. In this regard, inspired from local search based on leader in particle swarm algorithm, global search in estimation of

distribution algorithm and partitioning by Voronoi diagram, propose an algorithm that able to find more diverse solutions, better approximation of the optimal Pareto front and avoiding the algorithm to stick in local optima. Due to simplicity and it's consumedly performance in the principal component analysis, the use of this method proposed in order to reduce variables dependency and data dimension.

II. Literature Review

The estimation of distribution algorithms performance is naturally depends on statistical modeling and sampling method. There is a good wealth of literature on probability modeling in estimation of distribution algorithms.

Bayesian network is one of these probability models, which could consider the dependency between variables and use to learn parameters relationship and problem structure. These networks approximate the conditional probability. Each node and connection in these networks relate to the conditional probability. Some of the multi-objective estimation of distribution algorithms that use Bayesian networks includes BMOA, mohBOA, DT-MEDA, DP-MrBOA, [1-4]. MBN-EDA is a new modeling method in multi-objective estimation of distribution algorithm that uses multi dimensions Bayesian networks (MBN) to learn a common model of aims and variables. This model not only, like other estimation of distribution algorithms, considers the variables relationship, but also could consider the goals relationship [5].

Restricted Boltzmann machine is another statistical model that uses a hidden layer to model its variables distributions. This network includes hidden and visible units. The probabilities of the joint configuration over the hidden and visible units in network are trained until that distribution over the global state achieves a certain level of thermal equilibrium. Then, probability model is made by using network energy function. In RBM-EDA, Restricted Boltzmann machine model is introduced as the estimation of distribution algorithm in the multi-objective optimization filed [6]. In [7], EDA based on RBM combined with PSO for multi-objective optimization in noisy environment.

VEDA [8] and mohBOA ignore the regularity in making probability models. Since the probability modeling technics check widespread under special discipline and regularity in statistical

learning fields, thus it is possible to use discipline in designing estimation of distribution algorithms in continues multi-objective algorithms. In RM-MEDA, the regularity feature of continues multi-objective optimization problems use as a base for an estimation of distribution algorithm in facing variables relationship. RM-MEDA, model promised area in search space by a probability model that its gravity center is a piece continuous manifold. Local PCA algorithm uses to make this model [9].

Some methods combine the estimation of distribution algorithm with other heuristic search methods in order to increase the performance. The work presented in [10] is a combination of EDA and particle swarm algorithms. Gao et al. present an algorithm that combine EDA based on univariate distribution with particle swarm optimization (PSO) algorithm [11]. In [12], the combination of EDA and particle swarm optimization algorithm presented to control reservoir flood. This algorithm divides the population to some sub-population and makes the probability model for each of them. Based on the probability model, each sub-population reproduces new members with the estimation of distribution and particle swarm optimization algorithm. In [13], use univariate Gaussian model to estimate the solutions distributions. Some parameters of this model have the learning ability. In this method an EDA with comparative learning rate combine with chaos disorder search method.

While most of the multi-objective estimation of distribution algorithms are based on Pareto front, there are some study in other multi-objective optimization frame. In [14] the main idea is that any information from solutions near the weight vector should be helpful. Thus, making a probability model from all neighborhood solutions may have more efficient information than crossover operator [15]. In [16-19] cited some of estimation of distribution algorithm based on decomposition method. Shim et al. use a selection method based on decomposition in a combination of EDA based on univariate distribution and some local search methods [20]. In some of EDA a combination of some of probability models are used [21]. However, existing models could not reflect the problem completely. when the amount of variables and components increase, the optimization results become invalid [22]. Additionally, when consider all the relations between variables the computing cost increase.

III Proposed Method

A. Reducing Data Dimension

One of the reasons that could increase problem complexity is increase in data dimension. There are different methods to reduce problem complexity like correlation, the principal component analysis, linear programming and etc. Due to simplicity and performance of the principal component analysis, at first in the proposed method this procedure used to reduce the data dimension and variables transfer to coordinate system with fewer dimensions. In new coordination system problem solved and finally data return to the first coordinate system.

B. Principal Component Analysis

Principal component analysis is a statistical method that is frequently used to evaluate a group of correlated variables. Uses an orthogonal linear transformation to takes data to a new coordinate system and can be used to reduce the data dimension. Select fewer factors as the principal components from primary factors, by delete some unimportant information. The first extracted principal component, consider the most scattering over all data. That means the first component correlate at least to some of variables. The second extracted component has two important features, first this component consider the most set of data that are not compute by the first component, second it has no correlation to the first component. In other words, ignore the pervious component, pass the first to the final components, each component explain less variance. It means that always the first principal component explains the greatest amount of variance and the final components explain the least amount of variance, thus by ignoring the final components not lose many information.

C. Global search with estimation of distribution algorithm based on Voronoi

Voronoi partitioning has several advantages. Voronoi diagram has the ability to adapt the problem structure. Cover all the search space. Although unselected individual have low rank, but they may include well information, therefore, ignoring them may lose part of information. By partitioning, all individuals (selected and unselected) information could use to lead the search. After transferring the variables to new system, partition the search space by Voronoi diagram.

D. Voronoi partitioning

The discrete Voronoi diagram presented in [8] is used for simplicity. At first, all the search space divided to small grid, in which at most a member is existed in each grid. Then allocate the member rank to the related grid. The next step is to allocate the rank to the neighbors, till all the grids have rank. When a grid places near some members, it achieves the worse rank. All the grids with the same rank form a Voronoi mesh. Therefore, each mesh has its own rank. This rank use to calculate that mesh selection probability. To calculate individual rank, non-dominated sorting algorithm used [24] and for calculate probability geometric distribution is used. Mathematically, the geometric distribution is calculated as (1).

$$P = P_G(1 - P_G)^{r-1} \quad (1)$$

In (1) P_G is a number in $[0, 1]$ and r is the related rank. In this paper $P_G=0.8$.

E. Local search by particle swarm optimization algorithm

Each particle in particle swarm optimization algorithm, use local information of the best solutions that it has ever achieved (local optima) and the best answer that all particles have ever achieved (global optima). Using local information to exploitation could be helpful. In the proposed method, a mesh selected due to related probability. The control parameter $r(it)$ set the running algorithm steps. In the first steps due to keeping diversity and new areas exploration, the probability model sample and new solutions in selected mesh and produce by uniform distribution. But in final steps, after exploration of the best area, the local information use to produce new solutions by particle swarm algorithm. At the end, new solutions compare to pervious one and bests (non-dominated solutions) are kept. The proposed algorithm steps are shown in “Fig. 1”.

IV. Evaluations

The functions that choose for the test has complexity like linear/non-linear relationship between variables, continues/discontinues and convex/non-convex Pareto fronts. The first six functions are F1-F6 and F7-F9 are F8-F10 [9]. F10-F12 are respectively FON2, KUR1 and SCH1 functions in [8], F13-F15 are POLONNI,

ZDT3 and VIENNET functions, F16, F17 are respectively 2D and 3D DTLZ1 functions and F18 is DEB function.

1. Initialize population.
2. The principal component analysis and transferring data to the new coordinate system with fewer dimension
3. Making probability model
 - 3.1. Dividing the search space to grids and producing Voronoi diagram
 - 3.2. Determining the probability of each Voronoi mesh selection: $P = P_G(1 - P_G)^{r-1}$
4. Producing new population
 - 4.1 Computing control parameter: $r(it) = 1 - \frac{it}{maxit}$
 - 4.2 Producing random number (rand)
 - 4.3 Comparing random number to control parameter
 - 4.3.1 $rand < r(it)$
 - 4.3.1.1 Selecting a Voronoi mesh based on determined probability (3.2)
 - 4.3.1.2 Producing n_{sample} by uniform distribution in selected mesh
 - 4.3.1.3 Returning new data to the primary coordination system
 - 4.3.2 $rand > r(it)$
 - 4.3.2.1 Grid search space
 - 4.3.2.2 Selecting leader
 - 4.3.2.3 Updating particle position and velocity
 - 4.3.2.4 Adding non-dominated particles to archive and deleting non-dominated members
 - 4.3.2.5 Updating grid
 - 4.3.2.6 Returning new data to the primary coordinate system
5. Merging new population and pervious one
6. Evaluating the population and keeping non-dominated members in archive
7. If the terminate condition is met, end otherwise return to (2).

P: selecting Probability of each mesh $P_G = 0.8$: Geometric distribution parameter, r: Each mesh rank, $r(it)$: Control parameter, it: Algorithm iteration, maxit: The max iteration, n_{sample} : Number of new samples

Fig. 1. Proposed algorithm steps

A. Pareto front drawing

For sample in “fig. 2-4.” the True and approximation Pareto fronts by proposed, RM-MEDA [9] and MO-PSO-EDA [12] are shown.

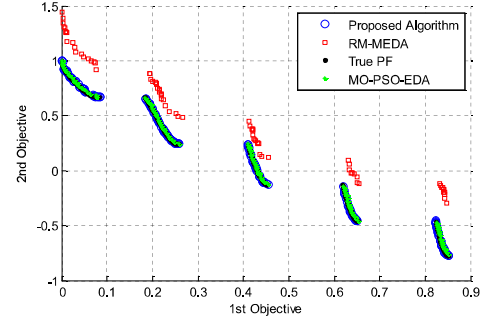


Fig. 2. True Pareto front (black), proposed (blue), RM-MEDA (red) and MO-PSO-EDA (green) for ZDT3 function (discrete, non-convex)

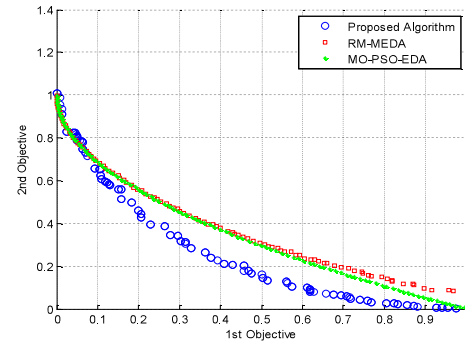


Fig. 3. Proposed, RM-MEDA and MO-PSO-EDA for ZDT1 function that x change to x^2 (non-linear relationship)

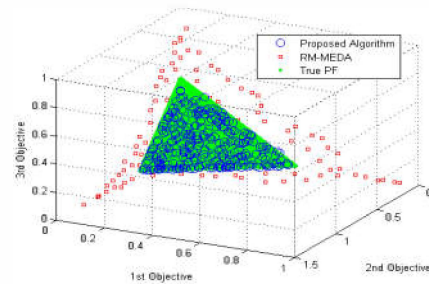


Fig. 4. True Pareto front and proposed and RM-MEDA for DTLZ1 function

B. Numerical criteria comparison

B.1 Quality criteria

C (A, B) between two approximations A and B of the Pareto front, is percent of B that dominate by A, and define as “(2)”:

$$C(A, B) = \frac{(|\{b \in B | \exists a \in A: a \text{ dominates } b\}|)}{(|B|)} \quad (2)$$

This criteria could evaluate the quality of achieve solutions by two different methods. As C is larger, it means that more member of B is dominated by A. And show that A has better quality than B.

Normalizing the C, Q expressible as “(3)”:

$$Q(A, B) = C(A, B) / (C(A, B) + C(B, A)) \quad (3)$$

In “table I” the numerical criteria, Q is shown for three algorithms. The result reported as average and standard deviation of achieve solutions in 40 runs. The number that shows better performance is bold. As it is clear, the proposed method in 11 out of 18 test functions has better performance than RM-MEDA and outperforms of MO-PSO-EDA in 11 of 16. Since MO-PSO-EDA did not show a sufficiently good quality on F9 and F15, these functions are not used here.

Table I. Comparing proposed (A), RM-MEDA (B) and MO-PSO-EDA (C) with Q criteria in 40 runs

	Q(A,B)	Q(A,C)
F1	0.7981 ± 0.3186	0.5922 ± 0.3438
F2	0.967 ± 0.1528	0.9565 ± 0.0837
F3	0.1122 ± 0.1285	0.1694 ± 0.2720
F4	0.2164 ± 0.1793	0.6032 ± 0.2253
F5	0.9999 ± 3.0847E-4	0.9999 ± 2.7732E-4
F6	0.9947 ± 0.032	0.999 ± 0.0044
F7	0.8765 ± 0.1016	0.8599 ± 0.846
F8	0.9429 ± 0.1685	0.9217 ± 0.1413
F9	1 ± 0
F10	0.9575 ± 0.029	0.2170 ± 0.0831
F11	0 ± 0	0.0429 ± 0.0222
F12	0.0441 ± 0.0332	0.1033 ± 0.0601
F13	0.5935 ± 0.1219	0.5947 ± 0.1493
F14	1 ± 0	0.8253 ± 0.23
F15	0.3147 ± 0.1318
F16	0.0028 ± 0.0045	0.6963 ± 0.135
F17	0.5222 ± 0.2427	0.2541 ± 0.1594
F18	0.1055 ± 0.2599	0.9218 ± 0.0981

B.2 Convergence and diversity criteria

The inverse generational distance (IGD) criteria could show not only the amount of diversity but also the solutions convergence. To compute this criterion, the test functions true Pareto front is needed, and because this optima front is available for 10 test functions, this criteria is used for these functions. If P^* is a set of points with uniform distribution in fitness space along

the true pareto front, and P is an approximate of the Pareto front, the inverse generational distance define as “(4)”:

$$IGD(P^*, P) = \frac{\sum_{v \in P^*} d(v, P)}{|P^*|} \quad (4)$$

In “(4),” $d(v, P)$ is the minimum Euclidean distance between v and the points in P . If $|P^*|$ is large enough and could well explain the Pareto front, $IGD(P^*, P)$ could measure both the diversity and convergence. As $IGD(P^*, P)$ is smaller, the approximate front is closer to the true Pareto front.

In table II average and standard deviation of IGD reported in 40 runs for the proposed method and RM-MEDA. Also p-value of t-test is reported.

Table II. IGD for proposed (A), RM-MEDA (B) and MO-PSO-EDA (C) in 40 run

	Method	MEAN ± STD	p-value	p-value
F1	A	0.0018 ± 0.0018	0.0043	A & B H=0
	B	0.0037 ± 0.0015		
	C	0.0031 ± 0.0066		
F2	A	0.0020 ± 1.9868E-4	3.74E-5	H=0
	B	0.0071 ± 0.0033		
	C	0.0027 ± 6.1347E-4		
F3	A	0.0485 ± 0.086	0.004	0.0155
	B	0.0074 ± 0.0034		
	C	0.0138 ± 0.006		
F4	A	0.0419 ± 0.0134	1.2043E-11	1.1698E-05
	B	0.022 ± 0.0042		
	C	0.0547 ± 0.0142		
F10	A	0.0025 ± 1.1408E-4	7.7825E-33	9.8177E-08
	B	0.0045 ± 2.7221E-4		
	C	0.0023 ± 11.8156E-4		
F11	A	0.0575 ± 0.0097	6.1304E-4	3.6589E-58
	B	0.0518 ± 0.0016		
	C	0.0525 ± 0.0018		
F12	A	0.0197 ± 0.0019	5.3377E-17	2.8285E-11
	B	0.024 ± 7.7952E-4		
	C	0.0242 ± 8.895E-4		
F14	A	0.0021 ± 0.0039	1.5928E-35	1.1735E-43
	B	0.3773 ± 0.0512		
	C	0.0030 ± 3.5536E-4		
F16	A	0.0273 ± 0.0042	2.1856E-15	9.2907E-06
	B	0.2148 ± 0.0953		
	C	0.0313 ± 0.0026		
F17	A	0.0598 ± 0.0051	0.0251	3.4166E-70
	B	0.0578 ± 0.0019		
	C	0.0594 ± 0.0016		

As the results show the proposed method IGD criteria is better in 5 out of 10.

V. Conclusion

This paper introduce an algorithm that simultaneously use global search space information for better exploration and local information of the solutions for more accurate exploitation. Moreover, the use of Voronoi diagram in proposed algorithm tends to make more efficient probability model. As the simulation results show, the algorithm achieves appropriate approximation of continues\discontinues, convex\non-convex optimum fronts as it is helpful in problems with linear\non-linear relationship between variables.

References

- [1] M. Laumanns and J. O'cen'a'sek, "Bayesian optimization algorithms for multi-objective optimization," presented at 7th International Conference on Parallel Problem Solving from Nature (PPSN VII), 2002.
- [2] M. Pelikan, K. Sastry, and D. E. Goldberg, "Multiobjective hBOA, clustering, and scalability," presented at Conference on Genetic and Evolutionary Computation (GECCO '05), New York, NY, USA: ACM, 2005.
- [3] X. Zhong and W. L. I. C. Society, "A decision-tree-based multi-objective estimation of distribution algorithm," presented at International Conference on Computational Intelligence and Security (CIS'07), 2007.
- [4] C. W. Ahn and R. S. Ramakrishna, "Multiobjective real-coded Bayesian optimization algorithm revisited: Diversity preservation," presented at 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07), New York, NY, USA: ACM, 2007.
- [5] H. Karshenas, R. Santana, C. Bielza, and P. Larranaga, "Multi-objective Estimation of Distribution Algorithm Based on Joint Modeling of Objectives and Variables," *Evolutionary Computation*, IEEE Transactions on vol. 18, no. 4, pp. 519–542, 2014.
- [6] H. Tang, V. A. Shim, K. C. Tan, and J. Y. Chia, "Restricted Boltzmann machine based algorithm for multi-objective optimization," *IEEE Congress on Evolutionary Computation (CEC'10)*, pp. 1-8, 2010.
- [7] V. A. Shim, K. C. Tan, and J. Y. Chia, "Multi-objective optimization with estimation of distribution algorithm in a noisy environment," *Evolutionary Computation*, vol. 21, pp. 149-177, 2013.
- [8] T. Okabe, Y. Jin, B. Sendoff, and M. Olhofer, "Voronoi-based estimation of distribution algorithm for multi-objective optimization," *IEEE Congress on Evolutionary Computation (CEC'04)*, vol. 2, pp. 1594–1601, 2004.
- [9] Q. Zhang, A. Zhou, and Y. Jin, "RM-MEDA: A regularity model based multiobjective estimation of distribution algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 12, pp. 41-63, 2008.
- [10] C.W. Ahn, J. An, and J. Yoo, "Estimation of particle swarm distribution algorithms: Combining the benefits of PSO and EDAs", *Inf. Sci.*, vol.192, no. 1, pp. 109-119, 2012.
- [11] Y. Gao, X. Hu, H. Liu, and Y. Feng, "Multiobjective estimation of distribution algorithm combined with PSO for RFID network optimization," presented at International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2010.
- [12] J. Luo, Y. Qi, J. Xie, and X. Zhang, "A hybrid multi-objective PSO-EDA algorithm for reservoir flood control operation," *Applied Soft Computing*, vol. 34, pp. 526-538, 2015.
- [13] Q. Xu, Ch. Zhang, J. Sun and Li. Zhang, "Adaptive Learning Rate Elitism Estimation of Distribution Algorithm Combining Chaos Perturbation for Large Scale Optimization" *The Open Cybernetics & Systemics Journal* , , vol. 10, pp.20-40, 2016.
- [14] Q. Zhang and H. Li, "Moea/d: A multiobjective evolutionary algorithm based on decomposition," *Evolutionary Computation*, *IEEE Transactions on*, vol. 11, no. 6, pp. 712–731, 2007.
- [15] Bo Wang; Hua Xu; Yuan Yuan, "Scale adaptive reproduction operator for decomposition based estimation of distribution algorithm," in *Evolutionary Computation (CEC)*, 2015 IEEE Congress on, pp.2042-2049, 25-28 May 2015
- [16] Y. Li, A. Zhou, and G. Zhang, "A decomposition based estimation of distribution algorithm for multiobjective knapsack problems," in *Natural Computation (ICNC)*, 2012 Eighth International Conference on. IEEE, pp. 803–80 , 2012.
- [17] A. Zhou, Q. Zhang, and G. Zhang, "A multiobjective evolutionary algorithm based on decomposition and probability model," in *Evolutionary Computation (CEC)*, 2012 IEEE Congress on. IEEE, pp. 1–8, 2012.
- [18] V. A. Shim, K. Tan, and C. Cheong, "A hybrid estimation of distribution algorithm with decomposition for solving the multiobjective multiple traveling salesman problem," *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on*, vol. 42, no. 5, pp. 682–691, 2012.
- [19] A. Zhou, F. Gao, and G. Zhang, "A decomposition based estimation of distribution algorithm for multiobjective traveling salesman problems," *Computers & Mathematics with Applications*, vol. 66, no. 10, pp. 1857–1868, 2013.
- [20] V. A. Shim, K. C. Tan, and C. Y. Cheong, "A hybrid estimation of distribution algorithm with decomposition for solving the multiobjective multiple traveling salesman problem," *IEEE Transactions on Systems*, vol. 42, pp. 682-691, 2012.
- [21] W. Dong, and X. Yao, "Unified eigen analysis on multivariate Gaussian based estimation of distribution algorithms", *Inf. Sci.*, vol. 178, no.15, pp. 3000-3023, 2008.
- [22] P.A. Bosman, and D Thierens, "Numerical optimization with real-valued estimation of distribution algorithms", In: *Scalable Optimization via Probabilistic Modeling*. Springer: Berlin, Heidelberg, 2006, pp. 91-120.
- [23] S. Muelas, A. Mendiburu, A. LaTorre, and J. Peña, "Distributed estimation of distribution algorithms for continuous optimization: How does the exchanged information influence their behavior?", *Inf. Sci.*, vol. 268, pp. 231-254, 2014.
- [24] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *Evolutionary Computation*, *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 20