Parallel Multi-objective Optimization on CPU Using Information Framework for Constructing Global Optimization Algorithms

Vladislav V. Sovrasov

State University of Nizhny Novgorod, Nizhny Novgorod, Russia
sovrasov.vlad@gmail.com

Abstract. In the present work, a parallel algorithm for the multi-objective optimization is considered. The considered approach is based on the application of the information-statistical algorithm to some reduced scalar problem, the set of the global optima in which coincides to the set of the weakly efficient solutions of the initial multi-objective problem. The sequential version of this method has been considered earlier. In the present work, the parallelization scheme based on the characteristics, which is common for all information global search algorithms, was applied to the sequential algorithm of the multi-objective optimization. Also, in the present work one of the techniques of accounting for the local properties of the optimized function allowing accelerating the convergence essentially has been considered for the multi-objective method for the first time.

Keywords: deterministic global optimization, multi-objective optimization, parallel numerical methods, derivative-free algorithms

1 Introduction

The multi-objective optimization problems attract an increasing interest in recent years since these ones reflect the controversy of the requirements arising to the modeled objects of real world the most completely. In such problems, the solutions are a compromised decision of a set of these ones. As a rule, these compromised solutions are chosen either from a set, in which all elements cannot be improved with respect to any criterion without the worsening of the other criteria (Pareto-optimal or efficient solutions) or from a set, in which no one element could be improved with respect to all criteria at once (Slater-optimal or weakly efficient solutions). The latter requirement is weaker, and all Pareto-optimal solutions are the Slater-optimal ones as well.

To find the complete set of efficient or weakly efficient points is the most computationally complex problem, however it provides a wide choice of possible compromised solutions. A plenty of approaches to the finding of these sets has been developed: the application of the genetic algorithms to the multi-objective problem directly [1], the use of the parametric convolutions reducing the problem

to a series of the scalar global optimization problems [2], various approaches to the solving the problems with the convex [3] or linear [4] objectives. The main problem in the using of majority of these methods is the absence of the theoretical guarantee of the uniform convergence to the whole sought set of the Pareto-optimal solutions on a wide class of problems including the ones with the non-convex multiextremal criteria. A method having this property in the case, when all criteria satisfy the Lipschitz condition has been developed on the basis of the information global search algorithm [5]. In the present paper, the parallel and accelerated by the local refinement technique [6](chapter 3) modification of this method will be considered.

2 Problem Statement and Dimensionality Reduction

The multi-objective optimization problems are stated in the following way:

$$\min\{f(y) : y \in D\}, D = \{y \in \mathbb{R}^n : a_i \leqslant y_i \leqslant b_i, 1 \leqslant i \leqslant n\} \quad (1)$$

Let us assume the components of the vector function (the partial criteria) $f_i(y), 1 \leqslant i \leqslant m$ to satisfy the Lipschitz condition with the Lipschitz constant L_i in D:

$$|f_i(y_1) - f_i(y_2)| \leqslant L_i \|y_1 - y_2\|, y_1, y_2 \in D, 0 < L_i < \infty, 1 \leqslant i \leqslant m$$

The set $S(D) \in D$ of strictly non-dominated points from the search domain is accepted as the solution of the problem (1) usually i. e.

$$S(D) = \{y \in D : \nexists z \in D, f_i(z) < f_i(y), 1 \leqslant i \leqslant m\} \quad (2)$$

which is usually referred as the set of semi-efficient (or weakly efficient) solutions. The conditions in the right-hand side of the definition (2) are known as the principle of weak Pareto-optimality (or Slater's optimality principle).

The use of the evolvents $y(x)$ i. e. the space-filling curves is a classic dimensionality reduction scheme for global optimization algorithms [7].

$$\{y \in R^N : -2^{-1} \leqslant y_i \leqslant 2^{-1}, 1 \leqslant i \leqslant N\} = \{y(x) : 0 \leqslant x \leqslant 1\}$$

Such a mapping allows the reduction of the problem (1) stated in a multidimensional space to solving a one-dimensional problem at the expense of worsening its properties. In particular, the one-dimensional functions $f_i(y(x))$ are not Lipschitzian but the Hölderian functions:

$$|f_i(y(x_1)) - f_i(y(x_2))| \leqslant H_i |x_1 - x_2|^{\frac{1}{N}}, x_1, x_2 \in [0, 1] \quad (3)$$

where the Hölder constants H_i are related to the Lipschitz constant L_i by the relation

$$H_i = 4L_i d\sqrt{N}, d = \max\{b_i - a_i : 1 \leqslant i \leqslant n\}$$

Therefore, not limiting the generality, one can consider the solving of the one-dimensional problem $\min\{f(y(x)) : x \in [0; 1]\}$, satisfying the Hölder condition.

The issues of the numerical building the mapping like a Peano curve and the corresponding theory have been considered in detail in [7]. Here we would note that an evolvent built numerically is an approximation of the theoretical Peano curve with a precision of the order 2^{-m} where m is the building parameter of the evolvent.

3 Description of the Parallel Algorithm With Local Refinement

Let us consider a scalarization scheme for the reduced problem (1), presented in [5]. Let

$$\varphi(x) = \max\{h(x,y) : y \in [0;1]\}, x \in [0;1]. \tag{4}$$

Let us consider a scalar problem

$$\varphi^* = \min\{\varphi(x) : x \in [0;1]\}. \tag{5}$$

As it has been shown in [5], the set of weakly efficient solutions of the reduced problem (1) coincides to the set of the globally optimal solutions of the problem (5) i. e.

$$S([0;1]) = \{x \in [0;1] : \varphi(x) = \varphi^*\} \tag{6}$$

Also, it has been shown in [5] that the function $\varphi(x)$ satisfies the Hölder condition when the requirements (3) are satisfied. Thus, the information-statistical global search algorithm can be applied to the function $\varphi(x)$ in order to solve the problem (5). However, $\varphi(x)$ is defined through the operator $\max\{...\}$, therefore, it is difficult to compute it directly. In [5], a modification of the classic information global search algorithm [8] is given, in which the values of $\varphi(x)$ are computed approximately. Let us consider the modified version of this algorithm. The modification consists in the use of the local refinement technique described in [6](chapter 3) as well as in the parallelization based on the characteristics [6](chapter 5).

The first two iterations are performed at the boundary points $x^0 = 0$ and $x^1 = 1$ of the interval $[0;1]$. The choice of the points $x^{k+j}, 1 \leqslant j \leqslant p$ is performed according to the following rules:

Step 1. Renumber the points in the set $X_k = \{x^1, \ldots, x^k\} \cup \{0\} \cup \{1\}$, which includes the boundary points of the interval $[0,1]$ as well as the points of preceding trials, by the lower indices in order of increasing coordinate values i. e.

$$0 = x_0 < x_1 < \ldots < x_{k+1} = 1$$

Step 2. Compute the lover bound μ_ν for the Lipzhitz constant for each objective function $f_\nu(x), 1 \leqslant \nu \leqslant m$:

$$\mu_\nu = \max_{1 \leqslant i \leqslant k} \frac{|f_\nu(x_i) - f_\nu(x_{i-1})|}{\Delta_i}, 1 \leqslant \nu \leqslant m \tag{7}$$

$$\Delta_i = (x_i - x_{i-1})^{\frac{1}{N}} \tag{8}$$

Step 3. To each point x_i, $0 \leqslant i \leqslant k$, juxtapose the value

$$z_i = \max\{h(x_i, x_j) : 0 \leqslant j \leqslant k\}, \tag{9}$$

where

$$h(x_i, x_j) = \min\{\frac{f_\nu(x_i) - f_\nu(x_j)}{\mu_\nu} : 1 \leqslant \nu \leqslant m\}, 0 \leqslant i, j \leqslant k \tag{10}$$

Step 4. For each interval $(x_i, x_{i-1}), 1 \leqslant i \leqslant k$ compute two characteristics:

$$R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{r^2 \Delta_i} - \frac{z_i + z_{i-1}}{2r} \tag{11}$$

$$R^*(i) = \frac{R(i)}{\sqrt{(z_i - z^*)(z_{i-1} - z^*)} + 1.5^{-\alpha}}, \tag{12}$$

where Δ_i is from (8), $z^* = \min\{z_i : 1 \leqslant i \leqslant k\}$, $r > 1$, and $\alpha \in [10; 30]$ are the input parameters of the algorithm.

Step 5. If $q \neq 0$ and $s \mod q \neq 0$, then arrange the characteristics $R(i)$, $1 \leqslant i \leqslant k+1$ in the decreasing order

$$R(t_1) \geqslant R(t_2) \geqslant \cdots \geqslant R(t_k) \geqslant R(t_{k+1})$$

and select p maximum characteristics with the indices of the intervals t_j, $1 \leqslant j \leqslant p$. Otherwise do the same with the characteristics $R^*(i), 1 \leqslant i \leqslant k+1$. Here s is the index of current iteration and q is the parameter of the method responsible for the degree of intensity of the local refinement. The less q, the more frequent the characteristics R^* are used making the method to choose the next points near the current minimum found.

Step 6. Perform the new trials at the points x^{k+j}, $1 \leqslant j \leqslant p$:

$$x^{k+j} = \frac{x_{t_j} + x_{t_j - 1}}{2} - \operatorname{sign}(z_{t_j} - z_{t_j - 1}) \frac{|z_{t_j} - z_{t_j - 1}|^n}{2r} \tag{13}$$

All p trials within this step can be performed in parallel on p computing devices.

The algorithm is terminated if the condition $\Delta_{t_j} \leqslant \varepsilon$ is fulfilled at least for one of the indices t_j, $1 \leqslant j \leqslant p$; here $\varepsilon > 0$ is the predefined accuracy. After the search is terminated, the set $S(\{x^0, \ldots, x^k\})$ of all non-dominated points of the truncated sequence $\{x^0, \ldots, x^k\}$ is accepted as an estimation for S from (6).

The theoretical substantiation of this method when $p = 1$ and $q = 0$ is presented in [6](chapter 3). The siffitient condition of convergence is: there exists an iteration such that $r\mu_\nu \geqslant 4H_\nu, 1 \leqslant \nu \leqslant m$.

4 Test Problems

In order to evaluate the degree of speedup of the convergence of the modified algorithm from Section 3, the following problems were used:

1. Markin-Strongin problem from [5]:

$$\begin{cases} f_1(y) = \min\{\sqrt{y_1^2 + y_2^2}, \sqrt{(y_1 - 1.5)^2 + (y_2 + 1.5)^2}\} \\ f_2(y) = \sqrt{(y_1 + 0.5)^2 + (y_2 - 0.5)^2} \end{cases} y_1 \in [-1; 2], y_2 \in [-2; 1] \tag{14}$$

2. Fonseca and Fleming problem [9]:

$$\begin{cases} f_1(y) = 1 - \exp\left(-\sum_{i=1}^{n}\left(y_i - \frac{1}{\sqrt{n}}\right)^2\right) \\ f_2(y) = 1 - \exp\left(-\sum_{i=1}^{n}\left(y_i + \frac{1}{\sqrt{n}}\right)^2\right) \end{cases} \quad y \in [-4; 4]^n \tag{15}$$

3. Viennet function [9]:

$$\begin{cases} f_1(y) = 0.5(y_1^2 + y_2^2) + \sin(y_1^2 + y_2^2) \\ f_2(y) = \frac{(3y_1 - 2y_2 + 4)^2}{8} + \frac{(y_1 - y_2 + 1)^2}{27} + 15 \\ f_3(y) = \frac{1}{y_1^2 + y_2^2 + 1} - 1.1\exp\{-(y_1^2 + y_2^2)\} \end{cases} \quad y \in [-3; 3]^2 \tag{16}$$

4. Poloni's two objective function [9]:

$$\begin{cases} f_1(y) = \left[1 + (A_1 - B_1(y))^2 + (A_2 - B_2(y))^2\right] \\ f_2(y) = (y_1 + 3)^2 + (y_2 + 1)^2 \end{cases} \quad y \in [-\pi; \pi]^2 \tag{17}$$

where

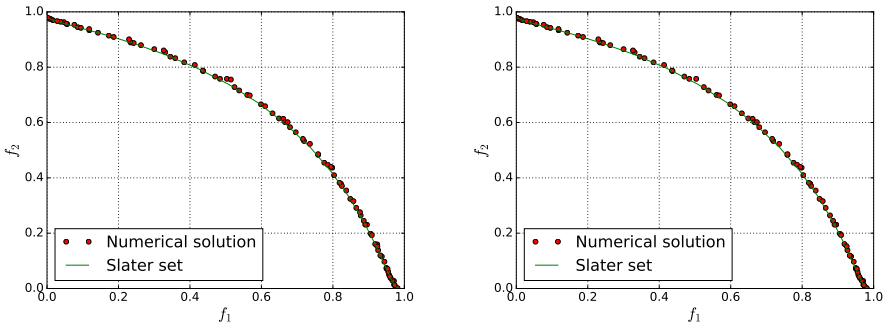
$$\begin{cases} A_1 = 0.5\sin(1) - 2\cos(1) + \sin(2) - 1.5\cos(2) \\ A_2 = 1.5\sin(1) - \cos(1) + 2\sin(2) - 0.5\cos(2) \\ B_1(y) = 0.5\sin(y_1) - 2\cos(y_1) + \sin(y_2) - 1.5\cos(y_2) \\ B_2(y) = 1.5\sin(y_1) - \cos(y_1) + 2\sin(y_2) - 0.5\cos(y_2) \end{cases}$$

5 Experimental Results

The computational experiments have been carried out on the Lobachevsky supercomputer at State University of Nizhny Novgorod. A computational node included 2 Intel Sandy Bridge E5-2660 2.2 GHz processors, 64 Gb RAM. The CPUs had 8 cores (i. e. total 16 cores were available per a node). In the present section, we will understand the speedup of the method as the speedup in the number of executed iterations, not in the time of execution. If the computation of the problem criteria (1) takes enough much time, the overhead costs for the execution of the decision rules of the optimization method are low as compared to the time of computing the criteria $f_i(y), 1 \leqslant i \leqslant m$.

Local refinement advantages. In [10] the numerical experiments demonstrating the speedup of convergence when using the local refinement technique in the algorithm solving the scalar global optimization problems are presented. One

could expect similar results from the multi-objective algorithm as well. The multi-objective algorithm with local refinement (MOALR) has been applied to the Fonseca and Fleming problem (15) at $n = 2$. The parameters of the method were as follows: $\varepsilon = 0.01,\ r = 4,\ q = 4,\ \alpha = 15,\ p = 1$. Before the method stops, 1176 iterations have been executed, the number of found weakly optimal points was 90. At $q = 0$ (without the local refinement) the method had performed 1484 iterations, the number of found weakly optimal points was 93. In Fig. 1 the set S in the problem (15) and the numerical estimates of this one obtained at $q = 0$ (fig. 1.a) and $q = 4$ (fig. 1.b) are presented. It is evident also from fig. 1 that the approximate solution covers the whole set of the weak-optimal solutions.



(a) MOALR with $q = 0$ (b) MOALR with $q = 4$

Fig. 1: Numericas estimation of S obtained by MOALR

Below, MOALR with $q = 4$ was used for all experiments.

Parallel method results. In order to demonstrate the speedup in iterations, which MOALR provides when $p > 1$, all problems from Section 4 have been solved for the values $p = 1, 2, 4, 8, 16$. The parameres of the method were the following: $r = 4.5$, $\varepsilon = 0.01$, $\alpha = 15$. The number of iterations is presented in Table 1 (in the braces, the cardinality of the set S is given) and the speedup in iterations is presented in in Table 2. As it is evident form Table 1, the cardinality of the weak-optimal solutions set changed insufficiently when varying p i. e. the quality of the obtained estimates remained the same. At the same time, the number of iterations decreased linearly with increasing value of p (Table 2). In Fig. 2, the examples of the numerical solutions of the considered problems are given.

Parallel method time speedup. The objective functions of all problems listed in this section are featured by low computational complexity. In order to demonstrate the possibility to obtain a speedup in time for the problems with the hard-to-compute criteria, additinal intensive froating point computations not

Table 1: Results of numerical experiments: number of iterations

Problem	p				
	1	2	4	8	16
Markin-Strongin	1041(198)	516(198)	256(185)	131(197)	68(191)
Fonseca and Fleming 2d	1181(93)	636(99)	386(111)	176(95)	106(97)
Fonseca and Fleming 3d	5346(160)	3551(183)	1186(143)	606(153)	351(142)
Viennet problem	4896(276)	2156(273)	1226(270)	631(287)	286(274)
Poloni's function	3351(102)	1706(90)	856(88)	426(96)	201(99)

Table 2: Results of numerical experiments: speedup in iterations

Problem	p			
	2	4	8	16
Markin-Strongin	2.02	4.07	7.95	15.31
Fonseca and Fleming 2d	1.86	3.06	6.71	11.14
Fonseca and Fleming 3d	1.51	4.51	8.82	15.23
Viennet problem	2.27	3.99	7.76	17.12
Poloni's problem	1.96	3.91	7.87	16.67

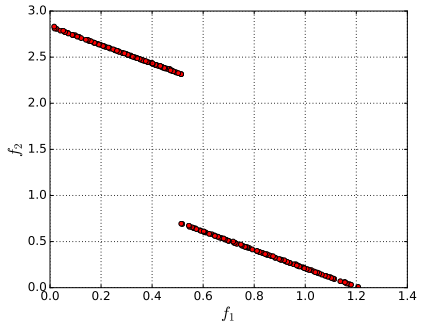
affecting the resulting values were introduced into the criteria. The speedups obtained are given in Table 3. In all cases, the speedup in time was less than the speedup in iterations because the operation of the optimizaion method itself takes a part of the execution time. However, in most cases the parallelizaion was efficient enough even in the case of 16 CPU threads. For more computation costly criteria, one can expect the increasing of the speedup in time for a large number of threads.

Table 3: Results of numerical experiments: speedup in time

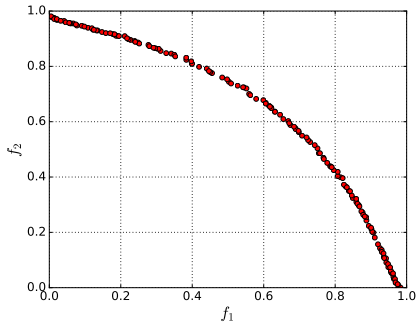
Problem	p				
	1(time, s)	2	4	8	16
Markin-Strongin	104.47	1.97	3.65	6.79	9.90
Fonseca and Fleming 2d	118.95	1.85	2.81	5.79	6.40
Fonseca and Fleming 3d	554.45	1.51	4.14	8.05	10.69
Viennet problem	1488.6	2.22	3.64	6.98	13.49
Poloni's problem	336.74	1.82	3.64	6.98	10.60

6 Conclusion

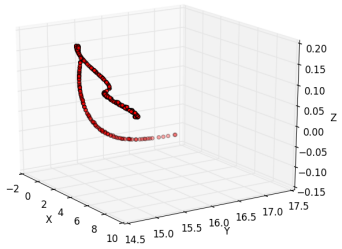
In the present work, a method for solving the multi-objective optimizaion problems allowing finding a uniform estimate of the set of the weakly-optimal points has been considered. The techniques of parallelization and of the acceleration of



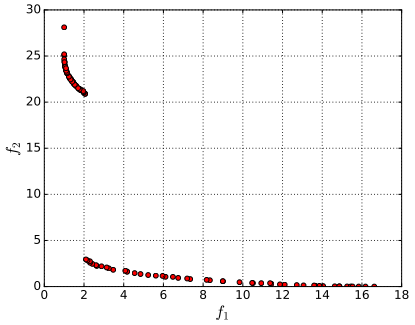
(a) Markin-Strongin problem



(b) Fonseca and Fleming 3d problem



(c) Viennet problem



(d) Poloni's problem

Fig. 2: Numericas estimations of S obtained by MOALR

convergence common for all algorithms of similar structure have been applied. The numerical experiments performed have demonstrated a speedup in convergence when using the local refinement. Also, the parallelization based on the characteristics was found to be efficient for this method. The speedup in iterations at the parallelization based on the characteristics appeared to be linear in most cases as for the information methods of scalar optimization (see the results of the parlallelization of the scalar optimization method, for example, in [11]). From the results of the experiments, one can conclude that it is expedient to apply the parallel MOALR in the problems of low dimensionality (1 to 3) with the computation-costly criteria. Besides, the property of uniform convergence of the considered algorithm to the whole set of the weakly-optimal solutions has been demonstrated on the numerical examples.

References

1. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.
2. V.P. Gergel and E.A. Kozinov. Accelerating parallel multicriterial optimization methods based on intensive using of search information. *Procedia Computer Science*, 108:1463 – 1472, 2017. International Conference on Computational Science, {ICCS} 2017, 12-14 June 2017, Zurich, Switzerland.
3. Guo-Li Zhang and Hua Zuo. Solution analysis of multi-objective programming problem. In *2013 International Conference on Machine Learning and Cybernetics*, volume 03, pages 1039–1044, 2013.
4. Harold P. Benson. An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem. *Journal of Global Optimization*, 13(1):1–24, 1998.
5. Markin, D.L., Strongin, R.G. Uniform estimates for the set of weakly effective points in multi-extremum multicriterion optimization problems. *Computational Mathematics and Mathematical Physics*, 33(2):171–179, 1993.
6. Strongin R.G., Sergeyev Ya.D. *Global optimization with non-convex constraints. Sequential and parallel algorithms.* Kluwer Academic Publishers, Dordrecht, 2000.
7. Sergeyev, Y.D., Strongin, R.G., Lera, D. *Introduction to Global Optimization Exploiting Space-Filling Curves.* Springer Briefs in Optimization, Springer, New York, 2013.
8. Markin, D.L., Strongin, R.G. A method for solving multi-extremal problems with non-convex constraints, that uses a priori information about estimates of the optimum. *Computational Mathematics and Mathematical Physics*, 27(1):33–39, 1987.
9. S. Huband, P. Hingston, L. Barone, and L. While. A review of multiobjective test problems and a scalable test problem toolkit. *Trans. Evol. Comp*, 10(5):477–506, October 2006.
10. Konstantin Barkalov and Ilya Lebedev. Local tuning in multilevel scheme of parallel global optimization. *AIP Conference Proceedings*, 1776(1):060006, 2016.
11. Konstantin Barkalov, Ilya Lebedev, Vladislav Sovrasov, and Aleksandr Sysoyev. Implementation of a parallel algorithm for searching the global extremum of a function on intel xeon phi. In *Proceedings of the 10th Annual International Scientific Conference on Parallel Computing Technologies*, pages 68–80, 2016.