

# Knowledge Extraction in Multi-objective Optimization Problem based on Visualization of Pareto Solutions

Fumiya Kudo

Graduated School of Engineering  
Nagoya University

Email: kudo@cmplx.cse.nagoya-u.ac.jp

Tomohiro Yoshikawa

Graduated School of Engineering  
Nagoya University

Email: yoshikawa@cse.nagoya-u.ac.jp

**Abstract**—Genetic Algorithm (GA) is one of the effective methods in the application to optimization problems. Recently, Multi-Objective Genetic Algorithm (MOGA), which is the application of Genetic Algorithm to Multi-objective Optimization Problems, is focused on in the engineering design field. In this field, the analysis of design variables in the acquired Pareto solutions, which gives the designers useful knowledge in the applied problem, is important as well as the acquisition of advanced solutions. This paper proposes a visualization method using an idea of Isomap, that visualizes manifold embedded in the high dimensional space, which was originally proposed in the field of multiple classification analysis. The proposed method visualizes the geometric distance of solutions in the design variable space considering their distance in the objective space. This method enables a user to analyze the design variables of the acquired solutions considering their relationship in the objective space. This paper applies the proposed method to the conceptual design optimization problem of hybrid rocket engine and studies the effectiveness of the proposed method. We found interesting structure in the distribution of Pareto solutions by applying the proposed method to this problem. This paper shows that the visualized result gives some knowledge on the features between design variables and fitness values in the acquired Pareto solutions.

## I. INTRODUCTION

Recently, Multi-objective Optimization Problems (MOPs) have been focused on in the engineering design field. Generally, it is difficult or impossible to acquire the optimized solution satisfying all objective functions because of their trade-offs. Then in MOPs, it is necessary to acquire Pareto solutions which are superior to other solutions by at least one fitness value. Multi-Objective Genetic Algorithm (MOGA), which is the application of GA[1] to MOPs, could be effective to solve MOPs because GA is multi-point search and it can search various Pareto solutions in one trial[2].

It has been reported in recent years that MOGAs are applied to engineering design problems due to the improvement of computing[3][4][5]. Obayashi[3] worked on design optimization of aircraft configuration problem using MOGA. [3] acquired Pareto solutions by MOGA method and analyzed the design variables of Pareto solutions to grasp the physical features in the problem through visualization using Self Organizing Map (SOM). Deb[4] studied the method to discover

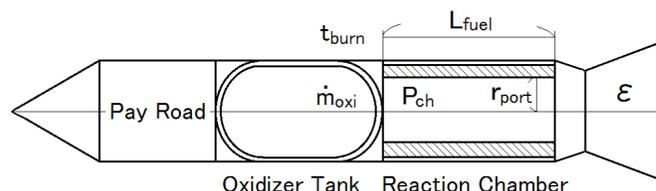


Fig. 1. Hybrid Rocket

useful information for designers from the Pareto solutions in engineering design problems. In engineering design field, it is important not only to search advanced Pareto solutions using MOGA, but also to grasp useful knowledge for designers and analyze physical relationships between fitness values and design variables[4][5].

This paper employs the conceptual design optimization problem of hybrid rocket engine (HRE) as an application of MOGA to engineering design problem[7]. This problem has been provided and published on the Web (in Japanese)[8] by Japan Aerospace Exploration Agency (JAXA).

This paper applies one of the representative multi-objective optimization method NSGA-II (Non-dominated Sorting Genetic Algorithm-II) [9] to the conceptual design optimization problem of HRE and analyzes the relationship between design variables and fitness values in the acquired Pareto solutions[10][11][12]. To consider their relationship, this paper proposes a visualization method which considers relative distance of data in the design variable space considering relative distance of data in the objective space using an idea of Isomap[13][14][15][16]. Isomap is the visualization method which preserves the distance between data and distributes them in a low-dimensional space based on "geodesic distance" instead of Euclidean distance in the original space. The experimental result shows that we can see the relationship of the design variables considering that of the fitness values as well, and we can grasp some knowledge on the features between them.

TABLE I  
DESIGN VARIABLES AND THEIR RANGES

V1	$\dot{m}_{oxi}(0)$ : Initial flow rate of oxidant 1.0 – 30.0[kg/s]
V2	$L_{fuel}$ : Length of fuel 1.0 – 10.0[m]
V3	$r_{port}(0)$ : Initial radius of port 0.01 – 0.2[m]
V4	$t_{burn}$ : Burning time 15.0 – 35.0[s]
V5	$P_{ch}$ : Pressure in reaction chamber 30.0 – 40.0[Bar]
V6	$\epsilon$ : Open area ratio 5.0 – 7.0

TABLE II  
FITNESS VALUES

Obj1	$M_{tot}$ : Weight of rocket [kg]	Min
Obj2	$H_{max}$ : Highest reachable altitude [km]	Max
Obj3	$M_{pay}$ : Weight of pay load [kg]	Max
Obj4	$L_{tot}$ : Length of rocket [m]	Min
Obj5	$a_{max}$ : Maximum acceleration [ $m/s^2$ ]	Min

## II. PROBLEM ESTABLISHMENT

This section explains the employed conceptual design optimization problem of HRE. HRE is the rocket engine that holds propellant in two different conditions, which has a smaller risk of explosion and a higher safeness. It is also environment-friendly, and it can adjust the thrust by using a slot ring. However, for a design of HRE, it is difficult to acquire general design knowledge because the thrust is generated by burning in the turbulent boundary layer in the reaction chamber, which is different from the conventional rocket engines. In the turbulent boundary layer, the O/F mixture ratio which dictates the thrust is calculated by the length of fuel, the radius of the port and the flow rate of oxidant. Thus, the geometric design of a HRE can be considered as an optimization problem which optimizes the weight of the rocket and the highest reachable altitude. As for the design variables of this problem, there are the flow rate of oxidant[kg/s], length of fuel[m], initial radius of the port[m], burning time[s], pressure in the reaction chamber[Bar] and open area ratio. As for the main objective functions, there are the minimization of weight of the rocket[kg] and the maximization of highest reachable altitude[km].

Fig.1 shows the conceptual figure of a hybrid rocket, and TABLE I and TABLE II show the design variables and the fitness values, respectively. This paper deals with three objective functions  $Obj1$ ,  $Obj2$  and  $Obj5$ , which can be expanded to 5 objective optimization problem in this problem as shown in TABLE II.

## III. VISUALIZATION METHOD

The proposed method uses an idea of geodetic distance to reflect the similarity of the solutions in the objective space into the geodetic distance in the design variable space. Isomap visualizes relationship between data defined by geodetic distance using Multidimensional Scaling (MDS).

### A. MDS

When the similarity between data in a multidimensional space were defined, MDS distributes the data in a visualiza-

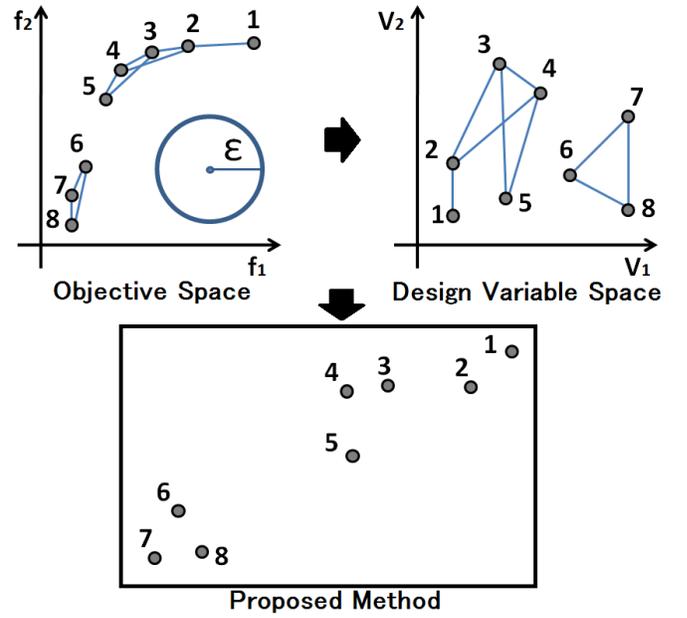


Fig. 2. Image of Proposed Method

tion space (low-dimensional space) preserving the similarities among data in the original space (high-dimensional space). This paper employs the Euclidean distance as the similarity in MDS and two-dimensional space to visualize the result. The fitness values are normalized based on the maximum and the minimum fitness values in each objective function acquired in genetic search and design variables are normalized by the feasible area shown in TABLE I into 0-1, respectively. The distance between data  $i$  and  $j$  in the original space  $d_{ij}$  and  $d'_{ij}$  in the visualization space are calculated by eq. (1) and (2). Then, it distributes the data to minimize the sum of the squares of error between two distances in the different spaces calculated by eq. (3).  $X_i$  and  $X'_i$  denote the coordinate in the original space and in the visualization space, respectively. In the error function eq. (3) in MDS, the coordinate itself for each data does not have meaning, but the distance between data does. Thus, in the visualization result by MDS, the axes and the coordinate of data does not have explicit meaning.

$$d_{ij} = \|X_i - X_j\| \quad (1)$$

$$d'_{ij} = \|X'_i - X'_j\| \quad (2)$$

$$\min \left\{ \sum_i \sum_j (d_{ij} - d'_{ij})^2 \right\} \quad (3)$$

### B. Isomap

Isomap is similar to MDS in terms of preserving the distance between data and distributes them in a lower-dimensional space. The feature of Isomap is using “geodetic distance” as the similarity of data. Geodetic distance is to measure the distance along the form of manifold. In this method, the distance in neighborhood is approximated by the Euclidean distance and others are calculated by the shortest path using the neighborhood distance. For example, in the distance between Hokkaido and Taiwan, the distance measured by straight line

that goes under the ground is the Euclidean distance, while the distance measured as (Hokkaido-Tokyo) + (Tokyo-Okinawa) + (Okinawa-Taiwan) is the geodetic distance. Isomap approximates the geodetic distance by the sum of the neighborhood distance.

Concretely, it defines “neighborhood” of each data in the original space and links them. There are two ways to define the neighborhood, one is to use the Euclidean distance in the original space and the other is to define the number of neighborhoods around each data. This paper employs the former. Next, the geodetic distance is calculated. In the geodetic distance, the distance between linked data is simply measured by its Euclidean distance and that between not linked data is calculated by the sum of the Euclidean distances of the data which give the shortest path to reach the target data along the linked ones. The distance between data which are not able to reach is given as  $D_{ue}$ . Usually,  $D_{ue}$  is set to relatively large value[16].

### C. Proposed Method

The Isomap calculates the geodetic distance along the manifold in the original space and visualizes the relationship among them. The feature of the proposed method is to define the neighborhood in the objective space and then to calculate the geodetic distance in the design variable space.

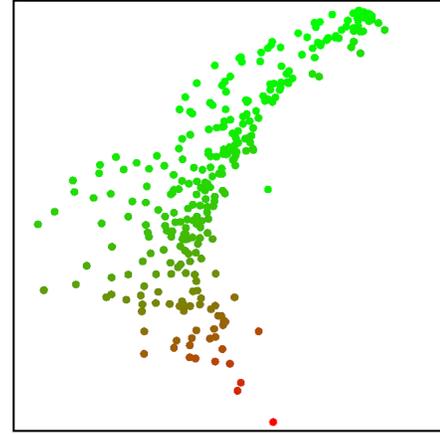
Figure.2 shows the image of the proposed method. In this paper, the radius  $\epsilon$  shown in the objective space in Fig.2 is calculated by eq. (4).  $\eta$  denotes the parameter that defines the distance of “neighborhood” and  $f_{imax}$  and  $f_{imin}$  denote the maximum and the minimum fitness value in the solutions in each objective function.

$$\epsilon = \frac{\sqrt{\sum_{i=1}^n (f_{imax} - f_{imin})^2}}{\eta} \quad (4)$$

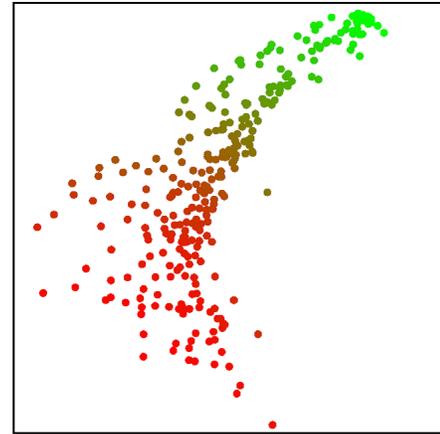
The proposed method defines the neighborhood in the objective space based on the distance in the sphere with the radius  $\epsilon$ , then it links the data in the design variable space using the neighborhood information in the objective space and calculates the geodetic distance in the design variable space.

For example in Fig.2, data 6 and data 8 become the neighborhood because the distance between data 6 and 8 is smaller than  $\epsilon$ , then they are linked each other. As for data 1 and data 4, they are not linked directly and the distance is (distance between data 1 and data 2 + distance between data 2 and data 4). Data 1-2-3-4-5 and data 6-7-8 are not linked together so that they are distributed far in the visualization space. In Fig.2, data that are near in both the objective space and the design variable space are distributed near in the visualization space (e.g. data 6,7,8), while data that are near in the design variable space but far in the objective space are distributed far in the visualization space (e.g. data 1 and data 5).

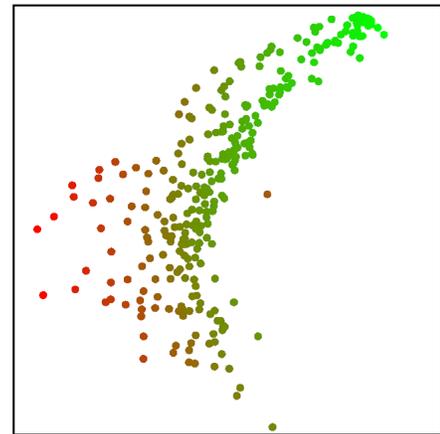
MDS and the original Isomap enable us to analyze the solutions acquired in optimization problems independently in the objective space or the design variable space. However, it is difficult to analyze both the objective space and the design variable space together and the relationship between them.



(a) Weight of Rocket



(b) Highest Reachable Altitude



(c) Maximum Acceleration

Fig. 3. Visualization of Objective Space by MDS

The proposed method uses the neighborhood information in

the objective space and calculate the geodetic distance in the design variable space to visualize the similarity. It is expected that the solutions can be analyzed considering both the objective space and the design variable space.

#### IV. EXPERIMENT AND CONSIDERATION

First, NSGA-II was applied to the conceptual design optimization problem of HRE explained in II. and 286 Pareto solutions were acquired. This paper used the real value coded GA, in which each gene represented each design variable directly shown in TABLE I. This paper employed PNX[17][18] as the crossover and Uniform mutation in which each variable was uniformly mutated in the range of area shown in TABLE I. Crossover rate was 1.0, mutation rate was 0.03, population size was 64 and the search was done for 32 generations.

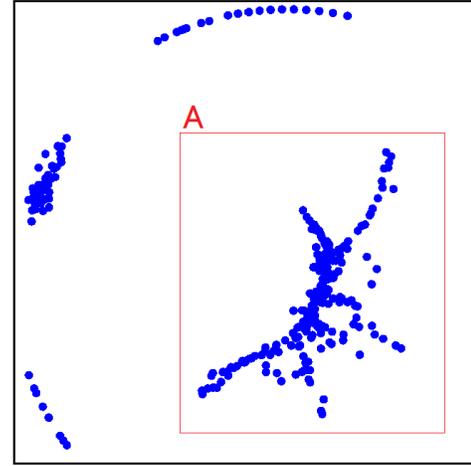
The 286 Pareto solutions visualized by MDS are shown in Fig.3(a)(b)(c). The gradation in each figure presents the weight of rocket ( $Obj1$ ), the highest reachable altitude ( $Obj2$ ) and the maximum acceleration ( $Obj5$ ). The red data have large value and green data have small value in each variable. As described in III.A., each fitness value is normalized by the maximum and the minimum fitness value acquired in genetic search into the range of  $[0, 1]$ . From Fig.3, the weight of rocket ( $Obj1$ : minimization) and the highest reachable altitude ( $Obj2$ : maximization) have trade-offs in the objective space, while the maximum acceleration ( $Obj5$ : minimization) is orthogonal to these objective functions.

Fig.4(a) shows the result of the proposed method to the acquired Pareto solutions shown above. This paper set  $\eta = 25$  in eq. (4). As for the distance  $D_{ue}$  for not linked data explained in III.B. was set to 10. In the proposed method, the solutions having similar fitness values are basically placed near one another because the neighborhood and the link structure are defined in the objective space. Fig.4(b) shows the solutions in the space  $A$  in Fig.4(a) in which the positions have been recalculated based on their geodetic distances. It can be seen that there are some branch structure. We picked up some branch structures shown in Fig.4(b) to see the characteristics of them. TABLE III shows the average of fitness values and design variables in each group labeled in Fig.4(b).

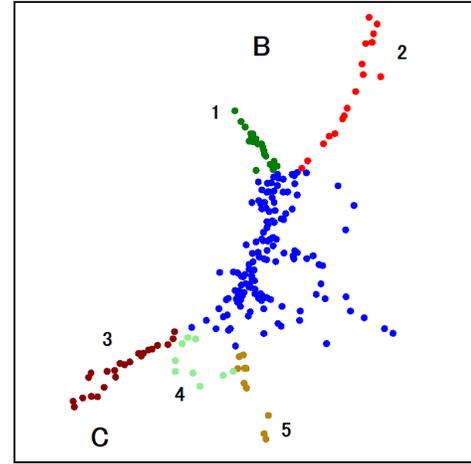
In TABLE III, group 1 and 2 have small fitness value while group 3, 4 and 5 have large value in comparison. Thus, the overall trend shows that the solutions are distributed from large fitness value in the low left part to small one in the upper right part in Fig.4(b). TABLE III also tells us that the design variables having large influence on this overall trend are the initial flow rate of oxidant ( $V1$ ) and the length of fuel ( $V2$ ).

In comparing group 1 and 2, group 1 have larger fitness value of the weight of rocket ( $Obj1$ ) and lower value of the maximum acceleration ( $Obj5$ ) than group 2. The same feature can be found in group 3 and 5 while the fitness values themselves are much different from group 1 and 2. And group 4 which is distributed between group 3 and 5 has the middle fitness value of them. The design variable of the influence on this feature will be the burning time ( $V4$ ), because group 1 and 3 have large value and group 2 and 5 have small one.

Figure.5 shows the gradation of the burning time ( $V4$ ). It can be seen that the branch between group 1 and 2, group 3 and 5 were caused by the burning time ( $V4$ ).



(a) Overall View



(b) Grouping in each Branch

Fig. 4. Visualization Result of Proposed Method

TABLE III  
VARIABLES IN EACH GROUP

	Group	1	2	3	4	5
$Obj1$	$M_{tot}$	221	174	1705	1377	1201
$Obj2$	$H_{max}$	82	73	193	192	195
$Obj5$	$a_{max}$	57	80	104	111	140
$V1$	$\dot{m}_{oxi}(0)$	2.1	2.5	20.3	17.6	19.5
$V2$	$L_{fuel}$	1.63	1.59	3.93	3.83	3.68
$V3$	$r_{port}(0)$	0.07	0.06	0.061	0.062	0.050
$V4$	$t_{burn}$	31	20.1	32.2	29.8	23.4
$V5$	$P_{ch}$	36.8	33.7	35.1	35.4	36.1
$V6$	$\epsilon$	5.7	5.9	6.3	6.1	6.1

Next, we focused on each branch. Fig.6 shows the result

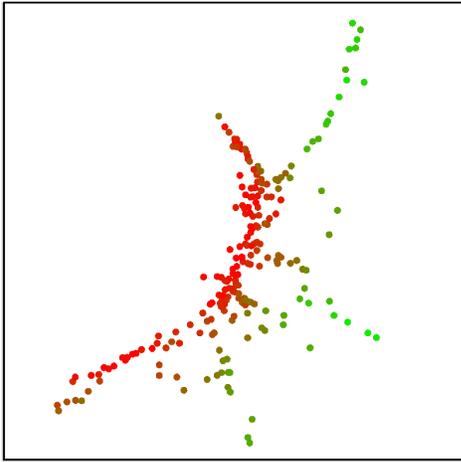


Fig. 5. Gradation by Burning Time ( $V4$ )

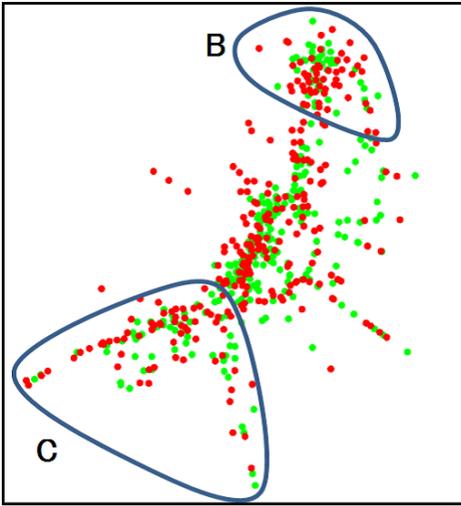


Fig. 6. Pareto Solutions (green) and Rank-2 Solutions (red)

when the proposed method was applied to the Pareto solutions (green) with the solutions of rank-2 (red). When some new solutions are added, the distribution of all solutions has changed due to the change of the information of the neighborhood and their links. In comparing Fig.6 with Fig.4(b), the space between group 1 and 2 (“ $B$ ” part in Fig.4(b) and Fig.6) was filled by the rank-2 solutions while the space between group 3 and 5 (“ $C$ ” part in Fig.4(b) and Fig.6) was not. It can be said that the  $B$  part was the space where dominated solutions were and the branch was generated on the progress of the search. On the other hand, there were no solutions in the  $C$  part even considering the dominated solutions.

Next, we generated some solutions to see if the  $C$  part could be filled in. First, we focused feasible solutions which can be generated from the solutions in the branches. The average fitness values of the solutions in the edge of group 3 and group 5 circled in Fig.7 (group 6 and group 7, respectively) were used as the bases. Only one design variable, the initial flow rate of oxidant ( $V1$ ) in first and the burning time ( $V4$ )

in second, was changed ( $i$ ) from the average value of group 6 to that of group 7 in 10 steps starting at the average values of design variables of group 6, and ( $ii$ ) that of 7 to that of 6 starting at group 7. Fig.7(a) shows the result of the proposed method applied to the Pareto solutions with these generated solutions for the initial flow rate of oxidant ( $V1$ ), and Fig.7(b) does that for the burning time ( $V4$ ). In Fig.7, the distribution of the Pareto solutions were fixed and the generated solutions were placed using Isomap. The solutions in the case ( $i$ ) described above are blue and ( $ii$ ) are red in Fig.7. The generated solutions were distributed around each group or along the distribution of the Pareto solutions. We could not fill in the space  $C$  by generating solutions which had the design variables between group 3 and 5.

We generated unfeasible solutions to find out if the space  $C$  could be filled in. Figure.8 shows the result of the proposed method to the Pareto solutions with the unfeasible solutions generated here. The unfeasible solutions had the same fitness values of solution  $b$  in Fig.8 and the design variables of the average of solution  $a$ ,  $b$  and  $c$  in Fig.8. However, they had  $35.0-70.0[kg/s]$  counted by  $5.0[kg/s]$  for the initial flow rate of oxidant ( $V1$ ) which were out of feasible area in this design variable. We can see in Fig.8 that these unfeasible solutions are distributed between group 3 and 5.

Thus, it can be said that the space between group 1 and 2 is the dominated solution space and the space between group 3 and 5 is the unfeasible solution space.

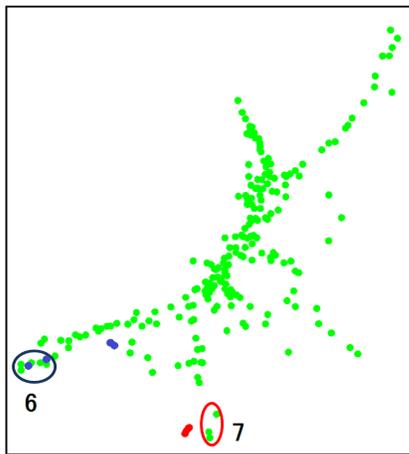
## V. CONCLUSION

This paper proposed a visualization method which visualized the geometric distances of solutions in the design variable space considering their distances in the objective space using the idea of Isomap. This paper applied NSGA-II to the conceptual design optimization problem of HRE to analyze the design variables in the acquired Pareto solutions. By using the proposed method, it was able to grasp the feature of the Pareto solutions and the correspondence of objective space and design variable space. The experimental results showed that the proposed method worked well to analyze the Pareto solutions and to extract knowledge of them. For the future work, we will apply the proposed method to other problems and feed back the extracted knowledge to genetic operations.

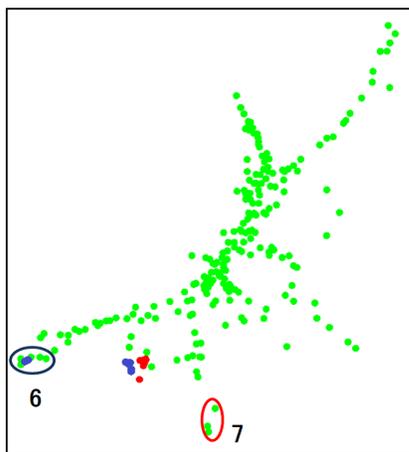
This research is partially supported by Strategic Programs for Innovative Research.

## REFERENCES

- [1] J. H.Holland, *Adaptation in Natural and Artificial Systems*, Univ. Michigan Press,1975
- [2] K .Deb, *Multi-Objective Optimization using Evolutionary Algorithms*, Chichester, UK, Wiley, 2001
- [3] Shigeru Obayashi, *Multiobjective Design Optimization of Aircraft Configuration (in Japanese)*, The Japanese Society for Artificial Intelligence, Vol.18, No.5, pp.495-501, 2003
- [4] K .Deb, *Unveiling Innovative Design Principles By Means of Multiple Conflicting Objectives*, Engineering Optimization, Vol35, Report Number 2002007, pp.1-6, 2003
- [5] Akira Oyama, Yasuhiro Kawakatsu, Kazuko Hagiwara, *Application of Multiobjective Design Exploration to Trajectory Design of the Next-Generation Solar Physics Satellite*, Japanese Society for Evolutionary Computation, 2010
- [6] Sunith Bandaru, Kalyanmoy Deb,*Automating Discovery of Innovative Design Principles through Optimization*,Indian Institute of Technology KanpurPIN 208016,KanGAL Report Number 2010001
- [7] Yukihiro Kosugi, Akira Oyama, Kozo Fuji, Masahiro Kanazaki, *Conceptual Design Optimization of Hybrid Rocket Engine (in Japanese)*, Utyuyusousinpojiumu, STCP-2009-75, 2010
- [8] [http://flab.eng.isas.ac.jp/member/oyama/realproblems\\_j.html](http://flab.eng.isas.ac.jp/member/oyama/realproblems_j.html) (in Japanese)
- [9] K .Deb, S.Agrawal, A .Pratab and T .Meyarivan, *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization*, NSGA-II, IEEE transactions on evolutionary computation, Vol.6, pp.182-197, 2002
- [10] Moubayed, N. and Petrovski, A. and McCall, J., *Clustering-Based Leaders' Selection in Multi-Objective Particle Swarm Optimisation*, Intelligent Data Engineering and Automated Learning-IDEAL 2011, p.100-107
- [11] Kudo, F. and Yoshikawa, T. and Furuhashi, T., *Study of Relationship in Design Space of Conceptual Design Optimization of Hybrid Rocket Engine*, The Japanese society for Evolutionary Computation, 2010(in Japanese)
- [12] Kudo, F. and Yoshikawa, T. and Furuhashi, T., *A study on analysis of design variables in Pareto solutions for conceptual design optimization problem of hybrid rocket engine*, Evolutionary Computation (CEC), 2011 IEEE Congress on, p.2558-2562, 2011
- [13] Ali Ghodsi, *Dimensionality Reduction A Short Tutorial*, Department of Statistics and Actuarial Science, University of Waterloo, Vol.41, pp.15-16, 2006
- [14] M .Bernstein, Vin de Silva, John C . Langford and J .B. Tenenbaum, *Graph approximations to geodesics on embedded manifolds*, Dept. Psychol.,Stanford Univ., Stanford, CA, 2000
- [15] Hisashi Handa, Hiroshi Kawakami, *Dimension Reduction by Manifold Learning for Evolutionary Learning with Redundant Sensory Inputs*, IEEE World Congress on Computational Intelligence, pp.1923-1928, 2010
- [16] Tenenbaum, J.B., Silva, V., Langford, J.C., *A global geometric framework for nonlinear dimensionality reduction*, Science, Vol.290, pp.2319-2323, 2000
- [17] Pedro J. Ballester, Jonathan N. Carter, *An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation*, Genetic and Evolutionary Computation-GECCO 2004, pp.901-913, 2004
- [18] Shigeyoshi Tsutsui, Masayuki Yamamura, Takahide Higuchi, *Multi-parent Recombination with Simplex Crossover in Real Coded Genetic Algorithms*,Proceedings of the Genetic and Evolutionary Computation Conference, volume 1, pp.657-664, 1999



(a) Shift of Initial Flow Rate of Oxidant (V1)



(b) Shift of Burning Time (V4)

Fig. 7. Generated Solutions

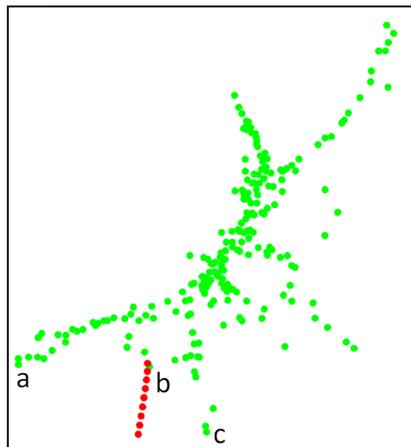


Fig. 8. Unfeasible Solutions (V1)