Uniform Sampling of Local Pareto-Optimal Solution Curves by Pareto Path Following and its Applications in Multi-objective GA

Ken Harada
Tokyo Institute of Technology
4259 Nagatsuta-cho,
Midori-ku
Yokohama-shi, Kanagawa,
Japan
ken@fe.dis.titech.ac.jp

Jun Sakuma
Tokyo Institute of Technology
4259 Nagatsuta-cho,
Midori-ku
Yokohama-shi, Kanagawa,
Japan
jun@fe.dis.titech.ac.jp

Shigenobu Kobayashi
Tokyo Institute of Technology
4259 Nagatsuta-cho,
Midori-ku
Yokohama-shi, Kanagawa,
Japan
kobayasi@dis.titech.ac.jp

ABSTRACT

Although multi-objective GA (MOGA) is an efficient multi-objective optimization (MOO) method, it has some limitations that need to be tackled, which include unguaranteed uniformity of solutions and uncertain finding of periphery of Pareto-optimal solutions. It has been shown that, on bi-objective problems, which are the subject of this paper, local Pareto-optimal solutions form curves. In this case, some of the limitations of MOGA can be resolved by sampling the curves uniformly in the variable space and in the objective space. This paper proposes Pareto Path Following (PPF) which does the sampling by extending the framework of Numerical Path Following, verifies that PPF exhibits the desired behaviors, and addresses the extension of PPF for problems with more than two objective functions.

Application of PPF is not limited to refinement of solutions obtained with MOGA. PPF makes it natural to have a local Pareto-optimal solution curve as the unit of search, which leads to curve-based MOGA. PPF also enables examination of which Pareto-optimal solution curves are found by MOO methods, and performance metrics based on it can be defined. This paper proposes these applications of PPF in MOGA and compares standard MOGA and curve-based MOGA using the metrics to reveal their characteristics.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*Gradient methods*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Algorithms, Performance, Experimentation, Theory

Keywords

Multi-objective optimization, local search, constraint handling

1. INTRODUCTION

Multi-objective function optimization (MOO) is required in many real-world problems. The solutions to which no other feasible solutions are superior in all objective functions are called Pareto-optimal, and those to which no other solutions in the feasible ε-vicinity are superior are called locally Pareto-optimal. The dimension N of the variable space is generally bigger than the dimension M of the objective space, and it has been shown that local Pareto-optimal solutions locally form $(M-1)$-dimensional manifold [7, 3]. Considering the abundance of bi-objective problems, $M = 2$ is assumed in this paper. Under this assumption, [7, 3] imply that local Pareto-optimal solutions form curves.

MOO methods attempt to find solutions that approximate all Pareto-optimal solutions. Among many MOO methods, Genetic Algorithms (GA) have been known to be effective and studied extensively [2]. As described in Subsection 2.3, multi-objective GA (MOGA) has some limitations that remain to be addressed, which include unguaranteed uniformity of solutions and uncertain finding of end-points of Pareto-optimal solution curves. These limitations can be lifted if local Pareto-optimal solution curves can be sampled uniformly in the variable space and in the objective space, as explained in Subsection 2.4. This paper proposes Pareto Path Following (PPF) which does the sampling by extending the framework of Numerical Path Following [1].

Use of PPF is not limited to refinement of MOGA's final solutions. When PPF can be used, it is natural to have a local Pareto-optimal solution curve, instead of a point solution, as the unit of search in MOGA. Such MOGA is proposed as curve-based MOGA in this paper. PPF also allows for the calculation of the ratio of the Pareto-optimal solution curves reached by the solutions of MOO methods. This paper proposes performance metrics based on it as well.

2. MOO AND MOGA

2.1 Multi-objective Optimization

Denote the vector of variables by $\boldsymbol{x} = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$

and the vector of objective functions by $\boldsymbol{f} = (f_1, \ldots, f_M)^T$. *Multi-objective function optimization (MOO)* is to minimize $\boldsymbol{f}(\boldsymbol{x})$ subject to $\boldsymbol{x} \in S \subset \mathbb{R}^N$. The *feasible region* S satisfies inequality constraints $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$, where $\boldsymbol{g} = (g_1, \ldots, g_L)$ is the vector of constraint functions. A solution is *feasible* if it is in S, and *infeasible* otherwise. Since there are a considerable number of real-world problems for which analytical or approximate gradients of f_i and g_j are available, f_i and g_j are assumed to be continuously differentiable. If $g_j(\boldsymbol{x}) = 0$ for some $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{x}$ is on the *boundary* of the constraint, and the constraint is *active*. A direction $\boldsymbol{d} \in \mathbb{R}^N$ at a solution $\boldsymbol{x}$ on the boundary is *feasible* if $\boldsymbol{d} \cdot \nabla g_j(\boldsymbol{x}) \leq 0$, and *infeasible* otherwise.

If $(\forall i\ f_i(\boldsymbol{x}_1) \leq f_i(\boldsymbol{x}_2)) \wedge (\exists i\ f_i(\boldsymbol{x}_1) < f_i(\boldsymbol{x}_2))$ holds for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$, $\boldsymbol{x}_1$ is *superior* to $\boldsymbol{x}_2$, which is denoted by $\boldsymbol{x}_1 \succ \boldsymbol{x}_2$. If there is no feasible solution $\boldsymbol{x}'$ such that $\boldsymbol{x}' \succ \boldsymbol{x}$, $\boldsymbol{x}$ is *Pareto-optimal*. If there is no solution $\boldsymbol{x}'$ such that $\boldsymbol{x}' \succ \boldsymbol{x}$ in the feasible ε-vicinity of $\boldsymbol{x}$, $\boldsymbol{x}$ is *locally Pareto-optimal*. In general, there are a number of Pareto-optimal and locally Pareto-optimal solutions. It has been shown that local Pareto-optimal solutions locally form a $(M-1)$-dimensional manifold [7, 3], assuming $N > M$. This implies that they form curves in the variable space and in the objective space when $M = 2$. These curves are called *local Pareto-optimal solution curves*. In general, there are multiple such curves. Since Pareto-optimal solutions are also locally Pareto-optimal, *Pareto-optimal solution curves* can be defined similarly.

2.2 Evaluation of MOO Methods

MOO methods attempt to find a set of solutions that approximate Pareto-optimal solutions, and many evaluation schemes for MOO methods have been suggested [2]. This paper notes that, while existing evaluation schemes are suitable for problems with only one or a few Pareto-optimal solution curves, they do not explicitly take into consideration the case of *multiple* curves. Hence, this paper considers different perspectives on evaluation of MOO methods: inter-curve coverage and intra-curve coverage.[1]

Inter-curve coverage means that all Pareto-optimal solution curves are reached by at least one solution given by MOO methods, where a curve is considered to be reached by a solution if applying multi-objective local search (MOLS) to the solution brings it to the curve. *Intra-curve coverage* means that each Pareto-optimal solution curve is uniformly covered with high precision from an end-point to the other. It is intra-curve coverage that has been studies extensively in previous studies. It is known to be composed of *proximity* and *diversity* of the solutions obtained, and diversity in turn consists of *uniformity*, a.k.a. *distribution*, and *extent* [2]. These two coverages can be evaluated either in the variable space or in the objective space.

2.3 Multi-objective GA

GA is known as an effective MOO method and has been studied extensively [2]. Multi-objective GA (MOGA) maintains a set of solutions and efficiently brings them closer to Pareto-optimal solutions by repeatedly applying a *crossover* operator and a *selection* operator to them. *Survival selection* is considered to be the most vital constituent of selection operators for MOO. Survival selection generally consists of some form of *ranking*, which selects solutions closer

---

[1]The ideas can be easily extended to the cases of $M \geq 3$.

to Pareto-optimal solutions, and *sharing*, which enhances the diversity of solutions by removing solutions in crowded areas in the objective space.

Regarding inter-curve coverage, it is generally believed that MOGA is good at reaching many Pareto-optimal solution curves since maintenance of a set of solutions allow MOGA to overcome locally but not globally Pareto-optimal solution curves. However, MOGA does not necessarily achieve good inter-curve coverage since some Pareto-optimal solution curves found during the search are lost in the end, as demonstrated in Subsection 4.4.

Regarding intra-curve coverage, one limitation of MOGA that has been addressed in existing studies is low precision [5]. It has been shown that applying MOLS to the final solutions of MOGA, rather than to those of each generation, resolves the precision limitation [5]. This paper notes that there are two other limitations that remain to be addressed. The first is that MOGA does not guarantee uniformity of solutions since it attempts to achieve uniformity *as a result* of removing solutions in crowded areas in the objective space. The second is that the solutions obtained with MOGA are not necessarily extended to the periphery of each Pareto-optimal solution curve since a crossover operator does not necessarily generate offsprings near the end-points of each curve, as demonstrated in Subsection 4.4. This paper primarily tackles the two limitations regarding intra-curve coverage, and the one regarding inter-curve coverage will be discussed in Subsection 4.4.3.

2.4 The Approach of This Paper

In order to resolve the limitations of MOGA regarding intra-curve coverage, this paper considers the problem of sampling local Pareto-optimal solution curves uniformly in the variable space and in the objective space. Apparently, resultant samples guarantee uniformity. Additionally, the end-points of the curves are reached while the curves are being sampled. Section 3 proposes Pareto Path Following (PPF) which does the sampling, and applications and implications of PPF in MOGA will be detailed in Section 4.

3. PARETO PATH FOLLOWING

This section first describes Numerical Path Following (NPF) and proposes PPF which incorporates the concepts of NPF. It then verifies through experiments that PPF exhibits the desired behaviors.

3.1 Numerical Path Following

3.1.1 Framework

Given variables $\boldsymbol{x} \in \mathbb{R}^{N+1}$ and a smooth map $H : \mathbb{R}^{N+1} \to \mathbb{R}^N$ whose Jacobian has maximal rank, the solutions of the underdetermined system of equations $H(\boldsymbol{x}) = \boldsymbol{0}$ form a curve. The solution curve $\boldsymbol{x}(s)$, where s is a one-dimensional parameter, can be regarded as solutions to the initial-value problem $\dot{\boldsymbol{x}}(s) = t(s)$ with $\boldsymbol{x}(0) = \boldsymbol{x}_0$, where $t(s')$ is the tangent vector of $\boldsymbol{x}(s)$ at $s = s'$ (cf. Fig. 1), and $H(\boldsymbol{x}_0) \approx 0$. Suppose that an initial solution $\boldsymbol{x}_0 \in \mathbb{R}^{N+1}$ such that $H(\boldsymbol{x}_0) \approx 0$ and a sufficiently small stepsize h are given. Numerical Path Following (NPF) [1] is a general framework of calculating the solutions $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ on the curve, one after another as depicted in Fig. 1, by alternately applying a predictor step and a corrector step with the following functions:
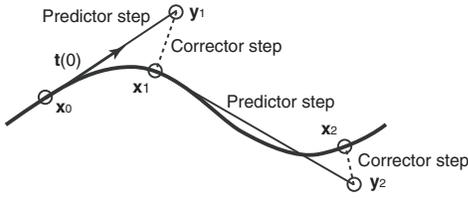
Figure 1: The initial-value problem corresponding to the solution curve of an underdetermined system of equations, and path following by a predictor step and corrector step [1]

Predictor step: Calculate $\boldsymbol{y}_i$ that is at the distance of h from $\boldsymbol{x}_{i-1}$ in the direction of the (unit) tangent vector $\bar{\boldsymbol{t}}_{k-1}$ of the curve at $\boldsymbol{x}_{i-1}$, i.e. $\boldsymbol{y}_i = \boldsymbol{x}_{i-1} + h\bar{\boldsymbol{t}}_{i-1}$.

Corrector step: Since $\boldsymbol{y}_i$ is generally off the curve, push it back onto the curve, and set the resultant solution as the next search point, i.e. $\boldsymbol{x}_i = \arg\min_{\boldsymbol{x}}\{||\boldsymbol{y}_i - \boldsymbol{x}|| : H(\boldsymbol{x}) = \boldsymbol{0}\}$.

3.1.2 Use of NPF in MOO

If constraints are all equality constraints, the necessary condition of local Pareto-optimality can be expressed as an underdetermined system of equations, and a method that follows local Pareto-optimal solution curves using NPF has been proposed [7]. When there are inequality constraints, they may be converted to equality constraints by some means, e.g. introduction of slack variables. However, these conversions make the method vulnerable to numerical errors [7]. Therefore, inequality constraints must be dealt with as they are. In the presence of inequality constraints, however, the necessary condition can *not* be expressed as an underdetermined system of equations, so NPF is not applicable.

3.2 Proposal of Pareto Path Following

In order to extend the path following framework of NPF for MOO, this subsection presents a predictor step and corrector step that follow local Pareto-optimal solution curves.

Predictor Step: Consider the case of no active constraint at a current search point $\boldsymbol{x}_{i-1}$. In this case, the gradients of objective functions are tangent to local Pareto-optimal solution curves. Therefore, the predictor step for MOO needs only to calculate a point $\boldsymbol{y}_i$ at a distance of h in the direction of the gradient of one of the two objective functions.[2]

Now consider the case of some active constraints at $\boldsymbol{x}_{i-1}$. If the gradient is feasible, active constraints can be ignored. If not, the search direction must be the projection of the gradient onto the tangent planes of the active constraints, as in *gradient projection method* [8], in order to avoid searching toward the outside of the feasible region. Since the active constraints may be nonlinear, $\boldsymbol{y}_i$ in that direction may be infeasible. If so, it can be made feasible by applying an appropriate repair operator.[3] In this paper, *Pareto Descent Repair operator (PDR)* [6] is used, which is briefly described in Appendix A.

Corrector Step: Since applying MOLS to $\boldsymbol{y}_i$ pushes it

---

[2]Using the other objective function gives the opposite tangent direction.

[3]When stepsize h is sufficiently small, the resultant solution of repairing $\boldsymbol{y}_i$ is expected to be approximately at the distance of h from $\boldsymbol{x}_{i-1}$.

back onto the local Pareto-optimal solution curve, the corrector step for MOO needs only to apply MOLS to $\boldsymbol{y}_i$ and set the resultant solution to be $\boldsymbol{x}_i$. In this paper, *Pareto Descent Method (PDM)* [4], which efficiently decreases all objective functions simultaneously with relatively small computational cost, is used as the MOSL.

Given an initial solution $\boldsymbol{x}'_0$, applying PDR, if $\boldsymbol{x}'_0$ is infeasible, and PDM to $\boldsymbol{x}'_0$ gives a local Pareto-optimal solution $\boldsymbol{x}_0$. The local Pareto-optimal solution curve to which $\boldsymbol{x}_0$ belongs can be followed using the predictor step and corrector step. This paper proposes this method as *Pareto Path Following (PPF)*. The primary constituents of PPF are PDM and PDR, whose main computation occurs in solving linear programming problems. Since computationally efficient solvers such as simplex method are available, it can be said that the computational complexity of PPF is accordingly small. Some practical uses of PPF are described below.

Uniform Sampling: With fixed stepsize h, PPF can sample local Pareto-optimal solution curves uniformly in the variable space. On the other hand, choosing $\boldsymbol{y}_i$ such that $||\boldsymbol{f}(\boldsymbol{y}_i) - \boldsymbol{f}(\boldsymbol{x}_{i-1})|| \approx h$ by binary search in the predictor step allows for uniform sampling in the objective space. Hence, PPF can sample curves uniformly in either space.

Finding the End-Points of Curves: If local Pareto-optimal solutions are bounded, PPF reaches the end-points of local Pareto-optimal solution curves while sampling the curves. When PPF reaches an end-point, the corrector step either pushes the search point $\boldsymbol{y}_i$ back to the end-point, or pushes it away to another curve. In either case, the distance $||\boldsymbol{x}_i - \boldsymbol{y}_i||$ by which the corrector step moves the search point becomes big. Therefore, arrival at an end-point can be detected, for example, by examining whether $||\boldsymbol{x}_i - \boldsymbol{y}_i|| \geq h\eta$, where $\eta \in (0, 1)$ is a user-specified parameter.

Determining Stepsize from Sample Size: It is convenient to be able to determine stepsize h so that a specified number n of solutions are obtained. Suppose that there are n_P Pareto-optimal solution curves and that, when the stepsize is h', n' solutions are obtained. The sum of the lengths of all curves is approximately $h'(n' - n_P)$, and PPF with $h = h'\frac{n' - n_P}{n - n_P}$ gives approximately n solutions.

3.3 Verification of the Behaviors of PPF

This subsection verifies through experiments that PPF samples local Pareto-optimal solution curves uniformly either in the variable space or in the objective space.

3.3.1 Experiment Setup

Benchmark Problems: Since PPF does not depend on the dimension N of a variable space, low dimensional benchmark problems Skewed QUAD and TNK are used, in order to examine in detail the behaviors of PPF *inside* and *on boundaries of* feasible regions, respectively.

Skewed QUAD is a 3-variable-2-objective problem with no constraint. Its objective functions are

$$f_1(\boldsymbol{x}) = \left((x_1/2)^2 + (x_2 - 1)^2 + x_3^2\right)^2,$$
$$f_2(\boldsymbol{x}) = \left((x_1 - 1)^2 + (x_2/2)^2 + x_3^2\right)^2.$$

Its Pareto-optimal solutions form a curve connecting $(0, 1, 0)^T$ and $(1, 0, 0)^T$ in the variable space. The initial solution for this problem is $(0.2, 0.5, 0.8)^T$, which was chosen arbitrarily.

TNK [2] is a 2-variable-2-objective problem with non-linear constraints. Its objective functions are $f_1(\boldsymbol{x}) = x_1$ and $f_2(\boldsymbol{x}) = x_2$. It has three local Pareto-optimal solution curves on a boundary. The initial solutions $(0.9, 1.0)^T$, $(1.1, 0.2)^T$, and $(0.2, 1.1)^T$ for this problem were chosen arbitrarily while ensuring that all three curves will be sampled. **PPF:** Given an initial point $\boldsymbol{x}_0'$, PDR is applied, if $\boldsymbol{x}_0'$ is infeasible, and PDM (a pair of direction calculation and linear search) is applied 12 times to obtain a local Pareto-optimal solution $\boldsymbol{x}_0$. Since solutions $\boldsymbol{y}_i$ that the predictor step gives are in the proximity of the local Pareto-optimal solution curve to which $\boldsymbol{x}_0$ belongs, PDM is applied 4 times in the corrector step. Stepsize h is determined so that approximately 35 solutions are obtained, as explained in Subsection 3.2. The settings of PDM and PDR are detailed below:

(Common Settings) Approximate gradients of functions are calculated using the forward difference of 10^{-4}. For linear search, golden section method is used, with basic search interval length of 10^{-2}, the maximum number of interval extension of 20, and the basic number of iteration of 20. When a distance between a solution and a constraint boundary is less than $10^{-2} \times \tau^{20}$, where τ is the golden ratio, the solution is assumed to be on the boundary.

(PDM) In order to sufficiently approximate the complete convex cone of feasible descent directions, 40 combination weights are randomly drawn for direction calculation.

(PDR) When active constraints are $\hat{C}^{\mathrm{u}} = \{\hat{c}_1^{\mathrm{u}}, \ldots, \hat{c}_{\hat{L}^{\mathrm{u}}}^{\mathrm{u}}\}$, $\hat{C}^{\mathrm{u}}$ and $\{\hat{c}_i^{\mathrm{u}}\}$ for each $i = 1, 2, \ldots, \hat{L}^{\mathrm{u}}$ are considered for inactivation. Search direction calculation and linear search are applied at most 30 times.

3.3.2 Experimental Results

PPF was applied to Skewed QUAD and TNK to sample their local Pareto-optimal solution curves uniformly in the variable space and in the objective space. The distributions of the resultant solutions are shown in Fig. 2. For Skewed QUAD, solutions in the variable space are projected onto the x_1-x_2 plane since the x_3-component of the Pareto-optimal solutions of Skewed QUAD is 0. For TNK, only the distribution in the variable space is shown, since each objective function of TNK equals the corresponding variable.

Examining the results on Skewed QUAD, one can see that solutions uniformly spaced in the variable space are obtained when the Pareto-optimal solution curve is sampled uniformly in the variable space, and those uniformly spaced in the objective space when the curve is sampled uniformly in the objective space. In addition, end-points of the curve are identified. Next, examining the results on TNK, one can observe that solutions uniformly spaced in the variable/objective space are obtained, and the end-points of the three curves are identified. These results confirm that PPF samples local Pareto-optimal solution curves uniformly either in the variable space or in the objective space, regardless of whether the curves are inside or on boundaries of feasible regions, and thus *ensures intra-curve coverage.*

3.4 Extension of PPF for Problems with More than Two Objective Functions
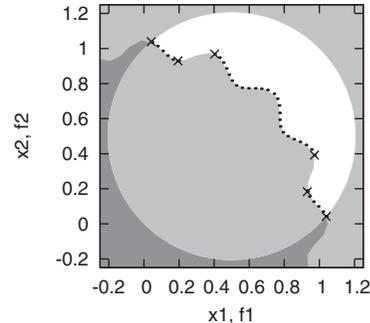
PPF, in principle, can be extended for problems with more than two objective functions, for which local Pareto-optimal solutions form surfaces or solids. The extension of PPF is explained below assuming $M = 3$. Similar extension is possible for the cases of $M > 3$.
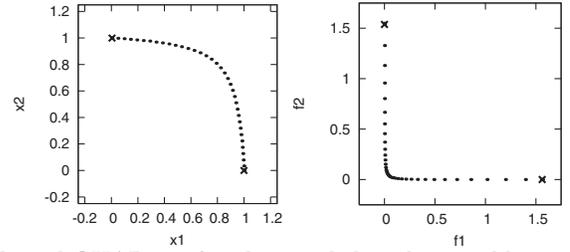
Skewed QUAD, uniformly sampled in the <u>variable space</u>
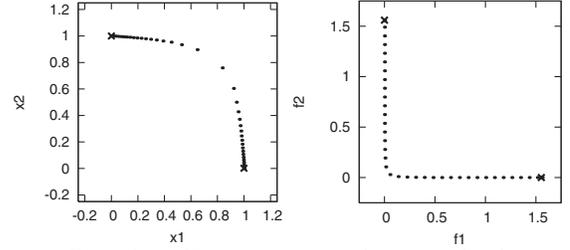
Skewed QUAD, uniformly sampled in the <u>objective space</u>

TNK, uniformly sampled in the <u>variable/objective space</u>

Figure 2: The distribution of solutions obtained with PPF. The end-points of local Pareto-optimal solution curves that PPF identified are denoted by $\times$. The shaded areas are infeasible, and, the more constraints they violate, the darker they are shaded.

When $M = 3$, local Pareto-optimal solutions form 2-dimensional manifolds, i.e. surfaces. Note that the gradients of objective functions are tangent to the surfaces. Choose f_1, for example, as the primary objective function, although the choice is arbitrary. Given an initial local Pareto-optimal solution $\boldsymbol{x}_{00}$, the surface to which $\boldsymbol{x}_{00}$ belongs can be sampled along the direction of ∇f_1 by taking ∇f_1 as the search direction in the predictor step. Now choose f_2 as the secondary objective function, although the choice is arbitrary as long as it has not already been chosen. Assuming that ∇f_2 and ∇f_1 are linearly independent, projecting ∇f_2 onto the null-space of ∇f_1 gives a direction that is tangent to the surface *and* orthogonal to ∇f_1. Using this direction in predictor step samples the surface in the direction orthogonal to ∇f_1. The extended PPF consists of the predictor step that uses these search directions and corrector step which is the same as before. Note that $(M - 1)$-dimensional integer coordinates (i, j) can be assigned to each solution $\boldsymbol{x}_{ij}$ given by the extended PPF, with $\boldsymbol{x}_{00}$ being assigned $(0, 0)$.

If a local Pareto-optimal solution surface is flat and ∇f_1 is constant on the surface, the extended PPF samples the surface in the square grid pattern in the variable space. How-
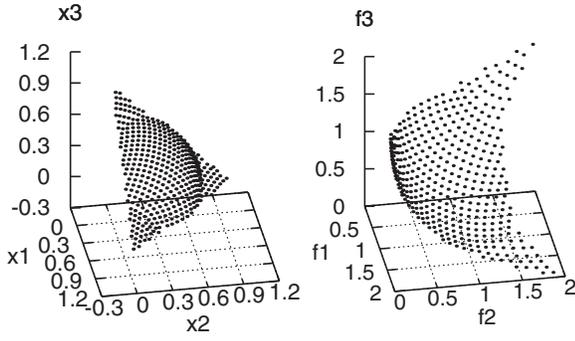
Figure 3: The distribution of solutions obtained by applying the extended PPF to BRT

Figure 4: UNDX-based crossover for curve-based MOGA

Figure 5: Survival selection for curve-based MOGA

ever, since that is generally not the case, the samples in the square grid pattern may be folded or stretched, which undermine the uniformity. Such fold or stretch may be detected by examining the distances to neighboring solutions implied by the integer coordinates. This, however, makes the entire algorithm cumbersome. It is also difficult for the extended PPF to sample in the square grid pattern in the objective space and to identify the edges of the surfaces.

In order to demonstrate the principles of the extended PPF which attempts to sample the surface in the square grid pattern in the variable space, it was applied to a 3-variable-3-objective problem with a non-linear constraint, Bumped Regular Triangle (BRT), whose objective functions are

$$f_1(\boldsymbol{x}) = ||\boldsymbol{x} - \boldsymbol{e}_1||^2, \; f_2(\boldsymbol{x}) = ||\boldsymbol{x} - \boldsymbol{e}_2||^2, \; f_3(\boldsymbol{x}) = ||\boldsymbol{x} - \boldsymbol{e}_3||^2,$$

where $\boldsymbol{e}_1 = (1,0,0)^T$, $\boldsymbol{e}_2 = (0,1,0)^T$, $\boldsymbol{e}_3 = (0,0,1)^T$, and the constraint is $||\boldsymbol{x}|| \geq 0.75$. Its Pareto-optimal solutions form a regular triangle in the variable space whose vertices are at $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, but part of the triangle is pushed up by the constraint. Fig. 3 shows the distribution of the resultant solutions. Although the samples are relatively uniform, one can observe that the uniformity on some parts of the surface is disturbed by the curvature of the surface.

4. APPLICATIONS OF PPF IN MOGA

This section gives three applications of PPF in MOGA. The first is refinement of MOGA's final solutions, which includes clustering of the solutions w.r.t local Pareto-optimal solution curves. The others are curve-based MOGA and performance metrics that measure the two coverages.

4.1 Refinement of Final Solutions of MOGA

Consider applying PPF to the solutions obtained with MOGA. Now the precision of solutions is ensured by PDM [5], and uniformity and extent within each curve are guaranteed by PPF, as demonstrated in Subsection 3.3. Therefore, applying PPF to the final solutions of MOGA ensures high intra-curve coverage. The experiment in Subsection 4.4 will show that it is difficult for MOGA alone to achieve comparable intra-curve coverage.

From a practical standpoint, it is favorable to have MOGA's solutions clustered w.r.t. local Pareto-optimal solution curves. This enhances the readability of solutions in, for example, lens system optimization problems [10], for which solutions belonging to the same local Pareto-optimal solution curve represent lenses with similar shapes, and those belonging
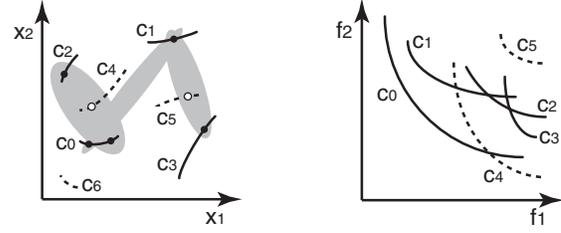
to different curves represent lenses with dissimilar shapes. PPF allows for such clustering as explained below. Suppose that curves are uniformly sampled in the variable space with a stepsize sufficiently small to ensure that the distances between curves are greater than the stepsize, and a local Pareto-optimal solution $\boldsymbol{x}$ is given by PDM to be examined for its belonging to a curve. One can conclude that $\boldsymbol{x}$ belongs to a curve if the minimum distance between $\boldsymbol{x}$ and the samples of the curve is smaller than the stepsize.

4.2 Curve-Based MOGA

When PPF can be used, a different approach of MOGA, i.e. curve-based MOGA, makes sense, as developed below.

Population: Standard MOGA has a point solution as the unit of search and attempts to approximate (local) Pareto-optimal solution curves with many points. However, since PPF generates an entire local Pareto-optimal solution curve from a single solution, it is natural to have a curve as the unit of search. This paper proposes MOGA whose population is a set of local Pareto-optimal solution curves as *curve-based MOGA*. Each curve, in practice, is represented by a set of solutions. One can use PPF to generate representative samples of each curve that are uniform in the variable space.

Crossover: When the unit of search is a local Pareto-optimal solution curve, the role of crossover is to find new curves. Assuming that a conventional crossover is used for generating offspring point solutions, it is appropriate, for that purpose, to take randomly-chosen solutions from different curves as parents. Applying PDM to the resultant offspring solutions gives local Pareto-optimal solutions. Those that do not belong to any of the curves in the population can be identified as described in Subsection 4.1. Since they belong to new curves, one needs to generate the curves, i.e. offspring curves, by PPF.

Fig. 4 shows an example. Solid curves $\boldsymbol{c}_0, \boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ are local Pareto-optimal solution curves in the population, and dashed ones $\boldsymbol{c}_4, \boldsymbol{c}_5, \boldsymbol{c}_6$ are those not in the population. Filled circles denote parent solutions, and offspring solutions are likely to be generated in the shaded areas when UNDX [9], for example, is used for crossover. Among the solutions obtained by applying PDM to the offspring solutions, those on curves not in the population are shown as white circles. Hence, the offspring curves are $\boldsymbol{c}_4$ and $\boldsymbol{c}_5$.

Survival Selection: Suppose that the population size is fixed. A survival selection must choose as many superior curves from the set of parent and offspring curves as the population size. In order to do this, one can order the curves according to their superiority by extending Non-Dominated

Sorting [2]: take away the curves not entirely dominated by other curves[4] and rank them 0, repeat the same with increasing rank until all curves are ranked, and select the curves in the ascending order of the rank. When not all curves of the same rank can fit into the population, the curves to be selected must be determined within the rank. Since the diversity of Pareto-optimal solutions must be maximized with the limited population size, it is appropriate to choose curves in the descending order of the length of the curve's non-dominated part(s) in the objective space.[5]

In the example given in Fig. 5, curves c_0, c_4 are ranked 0, c_1, c_2, c_3 are ranked 1, and c_5 is ranked 2. When the population size is 4, the curves to be selected are c_0, c_4, c_1, c_3.

Curve-based MOGA maintains high intra-curve coverage at all times, and never worsens inter-curve coverage since, once it finds a Pareto-optimal solution curve, the curve will never be discarded, as long as population size is greater than the number of Pareto-optimal solution curves. Standard MOGA, on the contrary, does worsen inter-curve coverage, as demonstrated in Subsection 4.4. However, curve-based MOGA requires many function evaluations for gradient calculation. Hence, it is necessary to investigate whether the effectiveness of curve-based MOGA measures up to the many function evaluations.

4.3 Performance Metrics for MOO Methods

Define *Inter-Curve Coverage Rate (Inter-CCR)* as the ratio of Pareto-optimal solution curves reached by at least one solution given by an MOO method. This directly measures inter-curve coverage. Additionally, for each Pareto-optimal solution curve, define *Reach Indicator (RI)* which is 1 if at least one solution given by the MOO method reaches the curve, and 0 otherwise. RI can be evaluated since PPF enables clustering of solutions w.r.t Pareto-optimal solution curves as explained in Subsection 4.1. Inter-CCR is simply the average RI *over all Pareto-optimal solution curves*. The average RI *over all simulation runs* gives the probability, hence ease or difficulty, of finding the particular curve.

Intra-curve coverage can also be evaluated using PPF. Given a benchmark problem, generating a large number of initial solutions uniformly at random in the feasible region and applying PDM to them give a picture of the local Pareto-optimal solutions. Applying PPF to them gives samples of the local Pareto-optimal solution curves. Among the samples, those not inferior to any others are good approximation of Pareto-optimal solutions and suitable reference points for evaluating performances. They conveniently allow for calculation of an approximate length of each Pareto-optimal solution curve in either space. Consider evaluating intra-curve coverage of an MOO method that gives n solutions when there are n_P Pareto-optimal solution curves and the sum of the curves' lengths is l. If the n solutions are precise, uniform, and extended to the periphery of each Pareto-optimal solution curve, each reference solution must be within the distance of $\frac{l}{2(n-n_P)}$ from at least one of the

---

[4]Since each curve is represented by a set of solutions in practice, the curves which has some constituent solutions not dominated by any solutions of any other curves are the ones not entirely dominated by other curves.

[5]Since each curve is represented by an ordered set of solutions generated by PPF in practice, an approximate length of the non-dominated parts of each curve can be calculated using the non-dominated constituent solutions.

n solutions. Define *Intra-Curve Coverage Rate (Intra-CCR)* for each Pareto-optimal solution curve as the ratio of the reference points on the curve that are within the distance of $\frac{(1+\epsilon)l}{2(n-n_P)}$ from at least one solution that reaches that curve, where $\epsilon > 0$ introduces tolerance. Intra-CCR directly measures intra-curve coverage. Note that, when PPF is used, Intra-CCR is either 0 or 1 since PPF guarantees intra-curve coverage of a curve that is reached.

4.4 Comparison of Standard MOGA and Curve-Based MOGA

This subsection compares standard MOGA and curve-based MOGA with two aims. The first one is to demonstrate that standard MOGA does not necessarily give solutions extended to the periphery of Pareto-optimal solution curves, which validates refinement of the solutions with PPF. The second is to demonstrate that standard MOGA worsens inter-curve coverage while curve-based MOGA does not.

4.4.1 Experiment Setup

Benchmark Problem: Multi-objective Rastrigin problem (MR), which is a 3-variable-2-objective problem with constraints, is used. The first objective function is the Rastrigin function, which has many local optima. The second is also the Rastrigin function but is rotated counterclockwise about the x_1-axis by 45 degrees, about x_2-axis by 70 degrees, about the x_3-axis by 20 degrees, and translated by $(0.3, 1.6, 2.8)^T$. The feasible region is $[-5.12, 5.12]^3$.

In order to visualize the local Pareto-optimal solutions around the Pareto-optimal solutions, 2×10^4 initial solutions were generated uniformly at random in $[-2, 1] \times [-1, 4]^2$, and PDM was applied to each of them 40 times. The resultant local Pareto-optimal solutions are shown in Fig. 6. They form numerous curves even in the limited region, making MR a difficult problem. PPF with stepsize $h = 0.005$ was applied to the solutions in Fig. 6, and, among the samples, the ones to which no others are superior were separated and are shown in Fig. 7. Single-point Pareto-optimal solutions, which are weakly but not globally Pareto-optimal and accrue to the structure of the Rastrigin function, have been removed. These are good approximation of the Pareto-optimal solutions and form 5 curves.

Performance Metrics: In addition to Inter-CCR and RI, extent ratio, defined below, is used. Given the final solutions of an MOO method, cluster them w.r.t. Pareto-optimal solution curves as explained in Subsection 4.1. For each curve reached by the final solutions, calculate the sum of the distances from each end-point to the final solution that reaches the curve and is the closest to the end-point in the normalized objective space. Define *extent ratio* as the ratio of the sum to the length of the Pareto-optimal solution curve in the normalized objective space. Extent ratio measures how spread the final solutions are on each curve, and the smaller the better. Note that extent ratio is defined only for curves reached by at least one final solution. Intra-CCR is not used since Subsection 3.3 has already shown that PPF ensures intra-curve coverage.

Standard MOGA: The standard population size 100 is used. Initial solutions are generated uniformly at random in the feasible region. UNDX [9] is used for crossover, and SPEA2 [11] for survival selection. For each generation, 50 parent pairs are formed, and each pair gives 2 or 20 off-
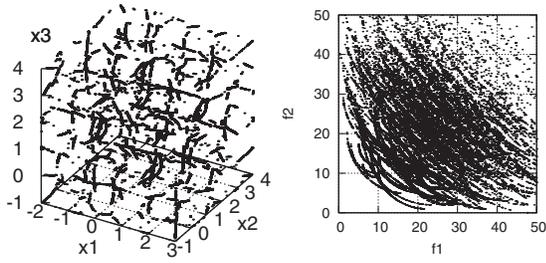
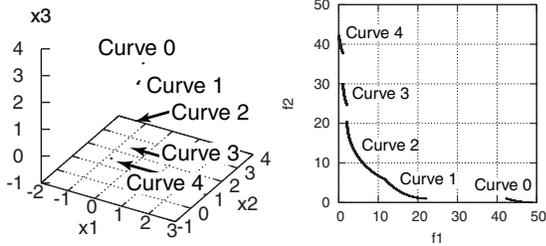**Figure 6: Local Pareto-optimal solutions of MR**

Figure 7: Pareto-optimal solutions of MR

Table 1: Average extent ratio

	C. 0	C. 1	C. 2	C. 3	C. 4
MOGA 02	0.809	0.181	0.013	0.665	0.904
MOGA 20	0.101	0.035	0.010	0.117	0.120
C.-b. MOGA	0.017	0.008	0.005	0.019	0.018

springs. These two versions of standard MOGA will be denoted by MOGA 02 and MOGA 20, respectively.

Curve-Based MOGA: Population size is 10. Initial curves are generated by applying PPF to 10 solutions distributed uniformly at random in the feasible region. UNDX is used for crossover. For each generation, 10 parent pairs are formed, and each pair gives 2 offsprings. The setting of PPF is the same as in Subsection 3.3, except $h = 0.005$.

4.4.2 Experimental Results

Table 1 shows the average extent ratios for each curve over 100 trials, *excluding* the trials for which the final solutions did *not* reach the corresponding curve. One can see that MOGA 20 performs better than MOGA 02 for all curves, which implies that simply increasing the number of offsprings for each parent pair gives better spread of solutions within each curve. However, curve-based MOGA achieved substantially better average extent ratios than MOGA 20 did for all curves because of PPF. These results validate application of PPF to the solutions obtained with MOGA in order to refine them, as explained in Subsection 4.1.

The transitions of the average Inter-CCR and RI over 100 trials are shown in Fig. 8. Comparing MOGA 02 and MOGA 20, one can see that MOGA 20 performed better in both Inter-CCR and RI, even when taking into consideration that MOGA 20 consumes 10 times as many function evaluations as MOGA 02 does. This implies that, although the number of offspring solutions generated from each parent pair is 2 in most existing studies, more Pareto-optimal solution curves are found when the number is bigger, outweighing the increase in function evaluations. One can also see that both MOGA *worsened* both Inter-CCR and RI in the course

of search. This implies that, although *more* Pareto-optimal solution curves are reached at some point during the search, *fewer* curves are reached in the end, which undermines inter-curve coverage. The average RI for curve 2 shows that the curve is found with high probability by all methods, implying that curve 2 is easy to find. On the contrary, curves 0, 3, and 4 are found with low probabilities by MOGA, implying that these curves are difficult for MOGA to find.

Comparing MOGA 20 and curve-based MOGA, one can see that curve-based MOGA performs better in both Inter-CCR and RI, and it monotonically improves both of them. However, the number of function evaluations must be taken into account when comparing them since curve-based MOGA requires more function evaluations than MOGA does. Preliminary experiments have shown that curve-based MOGA consumed about 50 times as many function evaluations as MOGA 20 did, where PPF is applied to the final solutions of MOGA 20 so that the outputs of the two methods are similar. Therefore, it can be inferred that, given a fixed number of function evaluations, multi-start MOGA 20 will probably find more Pareto-optimal solution curves with higher probability than single-run curve-based MOGA.

4.4.3 Discussion

On a large-scale problem with many local Pareto-optimal solution curves, e.g. lens system optimization problem [10], it becomes essential not to worsen inter-curve coverage so that, once any Pareto-optimal solution curves are found, they are never discarded. For this reason, curve-based MOGA is a valid approach although it requires more function evaluations. Therefore, it remains to compare standard MOGA and curve-based MOGA on such problems.

5. CONCLUSION

This paper tackled the limitations of MOGA regarding intra-curve coverage by proposing PPF that samples local Pareto-optimal solution curves uniformly in the variable space and in the objective space. When PPF can be used, the final solutions of MOGA can be refined, curve-based MOGA can be constructed, and performance metrics such as Inter-CCR and Intra-CCR can be evaluated. This paper verified through experiments that PPF exhibits the desired behaviors, and compared standard MOGA and curve-based MOGA w.r.t. extent ratio, Inter-CCR, and RI, giving some insight into the characteristics of these methods.

As explained in Subsection 3.4 and Subsection 4.4.3, it remains to investigate how to circumvent the difficulties that arise in extending PPF for problems with more than two objective functions, and to examine the approach of curve-based MOGA on more complex problems.

6. ADDITIONAL AUTHORS

Additional authors: Isao Ono (Tokyo Institute of Technology, email: `isao@dis.titech.ac.jp`)

7. REFERENCES

[1] E. L. Allgower and K. Georg. Numerical path following. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, volume 5, pages 3–207. North-Holland, 1997.
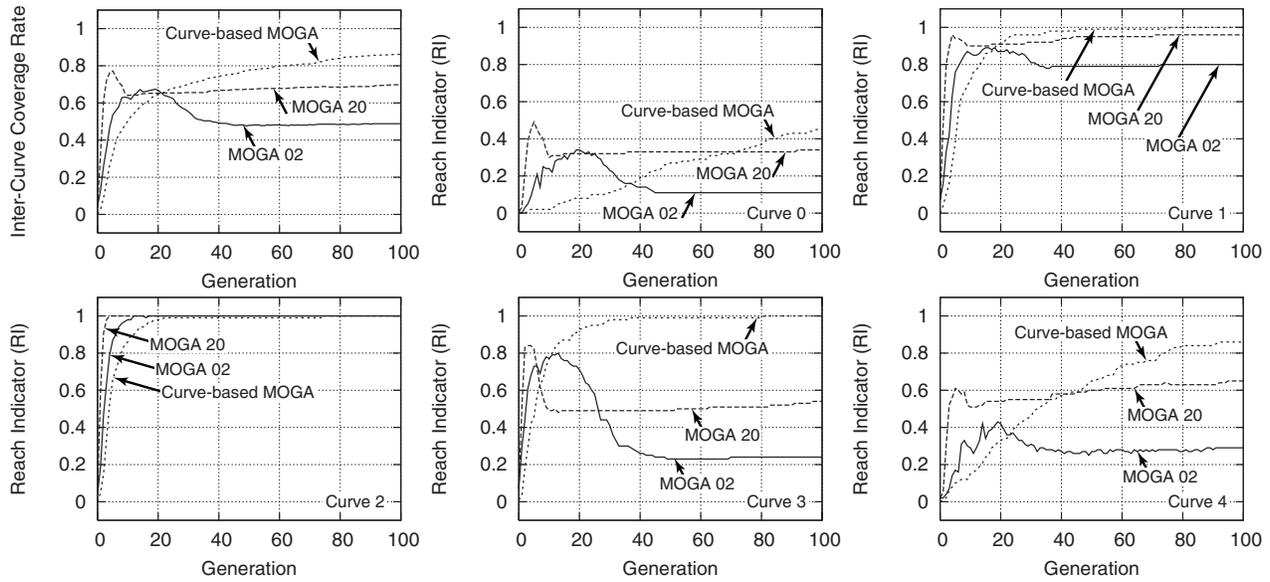
Figure 8: Transitions of average Inter-CCR and RI

[2] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms.* John Wiley & Sons, Chichester, 2001.

[3] M. Dellnitz, O. Schütze, and T. Hestermeyer. Covering pareto sets by multilevel subdivision techniques. *Journal of Optimization, Theory and Applications*, 124(1):113–136, 2005.

[4] K. Harada, J. Sakuma, K. Ikeda, I. Ono, and S. Kobayashi. Local search for multiobjective function optimization: Pareto descent method. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 659–666, New York, NY, 2006. ACM Press.

[5] K. Harada, J. Sakuma, S. Kobayashi, K. Ikeda, and I. Ono. Hybridization of genetic algorithm and local search in multiobjective function optimization: Recommendation of GA then LS. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 667–674, New York, NY, 2006. ACM Press.

[6] K. Harada, J. Sakuma, I. Ono, and S. Kobayashi. Constraint-handling method for multi-objective function optimization: Pareto descent repair operator. In *Evolutionary Multi-Criterion Optimization (EMO 2007)*, LNCS 4403, pages 156–170, Berlin Heidelberg, 2007. Springer-Verlag.

[7] C. Hillermeier. *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach*, volume 135 of *International Series of Numerical Mathematics.* Birkhäuser Verlag, 2001.

[8] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA, second edition, 1984.

[9] I. Ono and S. Kobayashi. A real-coded genetic algorithm for function iptimization using unimodal normal distribution crossover. In *Proceedings of the 7th International Conference on Genetic Algorithms (ICGA7)*, pages 246–253, 1997.
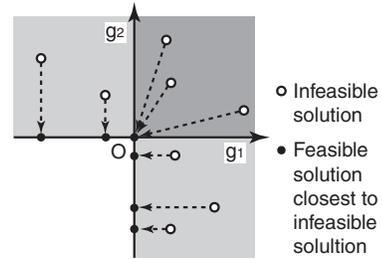
Figure 9: Repairing scheme of PDR in the presence of two constraints. The areas that violated constraints are shaded, and, the more constraints they violate, the darker they are shaded.

[10] I. Ono, S. Kobayashi, and K. Yoshida. Optimal lens design by real-coded genetic algorithms using UNDX. In *Computer methods in applied mechanics and engineering*, volume 186, pages 483–497, 2000.

[11] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In K. Giannakoglou et al., editors, *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN 2001)*, pages 12–21, 2001.

APPENDIX

A. PARETO DESCENT REPAIR OPERATOR

One approach for repairing an infeasible solution is to search for the feasible solution closest to the infeasible solution in the constraint space [6] (cf. Fig. 9). Pareto Descent Repair operator (PDR) [6] does this by incorporating the concepts of gradient projection method [8] and using PDM [4] to efficiently reduce all violated constraint functions simultaneously.