Diversifying Multi-Objective Gradient Techniques and their Role in Hybrid Multi-Objective Evolutionary Algorithms for Deformable Medical Image Registration

Kleopatra Pirpinia
Netherlands Cancer Institute
k.pirpinia@nki.nl

Tanja Alderliesten
Academic Medical Center
t.alderliesten@amc.uva.nl

Jan-Jakob Sonke
Netherlands Cancer Institute
j.sonke@nki.nl

Marcel van Herk
Netherlands Cancer Institute
m.v.herk@nki.nl

Peter A.N. Bosman
Centrum Wiskunde & Informatica
Peter.Bosman@cwi.nl

ABSTRACT

Gradient methods and their value in single-objective, real-valued optimization are well-established. As such, they play a key role in tackling real-world, hard optimization problems such as deformable image registration (DIR). A key question is to which extent gradient techniques can also play a role in a multi-objective approach to DIR. We therefore aim to exploit gradient information within an evolutionary-algorithm-based multi-objective optimization framework for DIR. Although an analytical description of the multi-objective gradient (the set of all Pareto-optimal improving directions) is available, it is nontrivial how to best choose the most appropriate direction per solution because these directions are not necessarily uniformly distributed in objective space. To address this, we employ a Monte-Carlo method to obtain a discrete, spatially-uniformly distributed approximation of the set of Pareto-optimal improving directions. We then apply a diversification technique in which each solution is associated with a unique direction from this set based on its multi- as well as single-objective rank. To assess its utility, we compare a state-of-the-art multi-objective evolutionary algorithm with three different hybrid versions thereof on several benchmark problems and two medical DIR problems. Results show that the diversification strategy successfully leads to unbiased improvement, helping an adaptive hybrid scheme solve all problems, but the evolutionary algorithm remains the most powerful optimization method, providing the best balance between proximity and diversity.

CCS Concepts

•**Mathematics of computing → Evolutionary algorithms;** •**Applied computing → Multi-criterion optimization and decision-making;**

Keywords

Evolutionary algorithms; Multi-objective Optimization; Gradients; Memetic Algorithms; Deformable Image Registration

1. INTRODUCTION

Local search algorithms that employ the gradient (i.e., the direction of greatest increase) of a function to iteratively arrive at a (locally) optimal solution have been studied for a long time in single-objective continuous optimization. In such approaches, gradient-based methods constitute a very powerful mathematical tool and are an efficient and practical way of finding (local) optima, especially when the function f to be optimized possesses desirable properties such as differentiability. As a consequence, their use is widespread in solving real-world optimization problems.

In contrast to local search algorithms, population-based algorithms such as evolutionary algorithms (EAs) seek to exploit the global properties of the search space. The synergy of EAs with local search has been investigated and has resulted in the so-called memetic algorithms or genetic local search [8], which are frequently successfully applied in single-objective- and multi-objective optimization.

Multi-objective optimization problems arise frequently in real-world situations, where it is often the case that optimal decisions need to be taken in the presence of two or more conflicting objectives. In a multi-objective approach, there are sets of optimal solutions to the problem, which represent the most efficient trade-offs that can be considered to be equally good [5].

One of the real-world problems that could benefit from such an approach is deformable image registration (DIR). DIR is a medical image processing task that can be of great value for healthcare. Its clinical implementation is still limited and presents many challenges [9]. Currently, a registration outcome is computed based on a single combination of different objectives, using predominantly single-objective gradient methods. However, there is currently no insightful way of finding the singular optimal combination of the objectives beforehand, thus making clinical implementation of such algorithms difficult. Different combinations lead to different outcomes, which can ultimately only be judged in quality by experts. For this reason, multi-objective optimization where multiple optimal outcomes with efficient

trade-offs are computed and presented simultaneously could be the key in obtaining wider clinical use. In this framework, studying the design and use of gradient-based methods for DIR becomes also of interest, as their efficient use could help achieve improved results.

However, as a consequence of the existence of these numerous (possibly infinite) trade-offs between objectives, the notion of improvement also becomes different. This makes the use of gradient information in multi-objective optimization less straightforward than in the single-objective case. For this reason, many different strategies have been proposed for its exploitation in such settings. In some cases, gradient information is not used explicitly [10]. Other multi-objective methods make explicit use of the gradient after having reduced the multi-objective problem to a single-objective one by constructing one function via weighted aggregation (see e.g., [11]) which is then optimized. Therefore, there is no guarantee that all objectives improve simultaneously. In another approach, the direction of greatest simultaneous improvement has been derived analytically [7], and applied in a hybrid algorithm [4]. Such approaches however can lead to a bias, if the objectives have different ranges, towards the objective with the largest range. Moreover, there are still infinitely many directions, that cannot be compared to each other; each one of them improves one objective more than the others. A complete analytic description of the best improving directions has been derived and has subsequently been utilized as a part of a multi-objective gradient search method [2]. Even then, a non-uniform spatial distribution of directions can lead to a bias due to differences in range or difficulty of the objectives. Therefore, to ensure gradient exploitation leads to a well-spread set of optimal solutions, it is essential to determine a strategy in which the multi-objective gradient is exploited such that it results in a balanced improvement of all objectives.

In this paper, we propose a diversification technique to exploit in such an unbiased way this multi-objective gradient information. We assess its utility by comparing a state-of-the-art multi-objective evolutionary algorithm to three different hybrid versions thereof on several benchmark problems, which gives us insight into the performance of the proposed approach when it comes to problems with great variation in difficulty and scale. Additionally, we test the best performing hybridization scheme on two DIR problems. Considering that the objectives of our DIR problem have a large difference in difficulty, we aim to assess whether the new technique results in a clearer advantage of the hybrid algorithm compared to the EA.

The remainder of the paper is organized as follows. In Section 2 we present some definitions for multi-objective optimization. In Section 3 we discuss the selected benchmark problems with an emphasis on those which exhibit the aforementioned behavior. In Section 4 we discuss in more detail the initial multi-objective gradient technique and the new diversification technique. Further, we present results of their comparison. In Section 5 we discuss the hybridization schemes, and the DIR problems we will test them on. In Section 6 we compare the performance of the schemes on all problems and in Section 7 we discuss our findings and conclusions.

2. DEFINITIONS

2.1 Multi-objective optimization

We assume to have m objective functions $f_i(\boldsymbol{x})$, $i \in \{0, 1, \ldots, m\}$ and, without loss of generality, we assume that the goal is to *minimize* all objectives. A solution $\boldsymbol{x}^1$ is said to (Pareto) dominate a solution $\boldsymbol{x}^2$ (denoted $\boldsymbol{x}^1 \succ \boldsymbol{x}^2$) if and only if $f_i(\boldsymbol{x}^1) \leq f_i(\boldsymbol{x}^2)$ holds for all $i \in \{0, 1, \ldots, m\}$ and $f_i(\boldsymbol{x}^1) < f_i(\boldsymbol{x}^2)$ holds for at least one $i \in \{0, 1, \ldots, m\}$. A *Pareto set* (denoted $\boldsymbol{\mathcal{P}_S}$) of size n is a set of solutions $\boldsymbol{x}^j$, $j \in \{0, 1, \ldots, n\}$ for which no solution dominates any other solution, i.e., there are no $j, k \in \{0, 1, \ldots, n\}$ such that $\boldsymbol{x}^j \succ \boldsymbol{x}^k$ holds. A *Pareto front* (denoted $\boldsymbol{\mathcal{P}_F}$) corresponding to a Pareto set is the set of all m-dimensional objective function values corresponding to the solutions, i.e., the set of all $\boldsymbol{f}(\boldsymbol{x}^j)$, $j \in \{0, 1, \ldots, n\}$. A solution $\boldsymbol{x}^1$ is said to be *Pareto optimal* if and only if there is no other $\boldsymbol{x}^2$ such that $\boldsymbol{x}^2 \succ \boldsymbol{x}^1$ holds. Further, the *optimal Pareto set* is the set of all Pareto-optimal solutions and the *optimal Pareto front* is the Pareto front that corresponds to the optimal Pareto set.

2.2 Multi-objective gradient

For any real-valued function $f(\boldsymbol{x})$, $\boldsymbol{x} = (x_0, x_1, \ldots, x_{l-1})$, the gradient of f is defined as:

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_0}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_{l-1}} \right).$$

The directional derivative then is a function that gives the rate of change of f in direction $\hat{\boldsymbol{u}}$ at any $\boldsymbol{y}$:

$$\nabla_u f(\boldsymbol{x})(\boldsymbol{y}) = (\nabla f(\boldsymbol{x})(\boldsymbol{y}))^T \hat{\boldsymbol{u}}. \qquad (1)$$

Therefore, to find the direction of greatest decrease the following single-objective optimization problem must be solved:

$$\min_{\hat{u}} \{ \nabla_u f(\boldsymbol{x})(\boldsymbol{y}) \}.$$

It is easy to see that the directional derivative at a point is minimal when the direction points to the direction of the negative gradient, i.e., the solution to this problem is naturally $-\nabla f(\boldsymbol{x})(\boldsymbol{y})$.

If we extend the notion of gradient and directional derivatives to a multi-objective setting, we must now solve the optimization problem

$$\min_{\hat{u}} \{ \nabla_u \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y}) \},$$

where $\boldsymbol{f} = (f_0, f_1, \ldots, f_{m-1})$. The directional derivative of $\boldsymbol{f}$ is now also a vector whose i-th component is the directional derivative in the i-th objective, i.e., $(\nabla_u \boldsymbol{f}(\boldsymbol{x}))_i = \nabla_u f_i(\boldsymbol{x})$ and therefore (1) for the multi-objective case is formulated as

$$\nabla_u \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{G}\hat{\boldsymbol{u}},$$

where $\boldsymbol{G} = (\nabla f_0(\boldsymbol{x}), \ldots, f_{m-1}(\boldsymbol{x}))^T$. Now, similar to the single-objective case we are ultimately interested in the direction that maximizes the improvement (i.e., the negative gradient). However, now this is a multi-objective optimization problem. Hence, no single direction of greatest increase of $\boldsymbol{f}$ starting from $\boldsymbol{y}$ exists, but a set of directions that

correspond to the optimal Pareto front of all improving directions. A parametric description of these Pareto-optimal directions is available [2], and a method called Combined-Objectives-Repeated-Line search (CORL), that will be explained in more detail in Section 4, exploits them by choosing a random direction from this parametrically described set. However, choosing a direction at random can lead to a bias, in particular when the objectives to be optimized vary a lot in difficulty. To investigate this, we have specifically chosen some problems in our multi-objective optimization test suite that exhibit this bias.

3. BENCHMARK PROBLEMS

We have selected a set of well-known benchmark problems in multi-objective evolutionary optimization, see Table 1. We first considered the well-known problems ZDT_i, $i \in \{1, 2, 3, 6\}$. The initialization ranges (IRs) of the ZDT_i problems are also constraints. For more details, see [12].

We also considered two problems with smooth properties. These are the most simple ones, and they are a generalization of the MED (Multiple Euclidean Distances) problems, which we will refer to as GenMED. There are two cases, one where the optimal Pareto front is concave and one where it is convex, but both cases are smooth functions with equally scaled objectives and without any local Pareto fronts.

Lastly, we have selected two more problems, which are characterized by a strong difference in difficulty between their objectives. Both of these problems, labeled BD_i, $i = 1, 2$, make use of the Rosenbrock function in one of their objectives. The real-world problems we are interested in share this difference in difficulty.

Name	Objectives	IR
ZDT_1	$f_0 = x_0, \quad f_1 = \gamma \left(1 - \sqrt{f_0/\gamma}\right)$ $\gamma = 1 + 9 \left(\sum_{i=1}^{l-1} x_i/(l-1)\right)$	$[0;1]^{30}$ $(l = 30)$
ZDT_2	$f_0 = x_0, \quad f_1 = \gamma \left(1 - (f_0/\gamma)^2\right)$ $\gamma = 1 + 9 \left(\sum_{i=1}^{l-1} x_i/(l-1)\right)$	$[0;1]^{30}$ $(l = 30)$
ZDT_3	$f_0 = x_0, \quad f_1 = \gamma \left(1 - \sqrt{f_0/\gamma} - (f_0/\gamma)\sin(10\pi f_0)\right)$ $\gamma = 1 + 9 \left(\sum_{i=1}^{l-1} x_i/(l-1)\right)$	$[0;1]^{30}$ $(l = 30)$
ZDT_6	$f_0 = 1 - e^{-4x_0}\sin^6(6\pi x_0), \quad f_1 = \gamma \left(1 - (f_0/\gamma)^2\right)$ $\gamma = 1 + 9 \left(\sum_{i=1}^{l-1} x_i/(l-1)\right)^{0.25}$	$[0;1]^{10}$ $(l = 10)$
BD_1	$f_0 = x_0, \quad f_1 = 1 - x_0 + \gamma$ $\gamma = \sum_{i=1}^{l-2} \left(100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right)$	$[0;1] \times$ $[-5.12; 5.12]^9$ $(l = 10)$
BD_2^s	$f_0 = \frac{1}{l} \sum_{i=0}^{l-1} x_i^2$ $f_1 = \frac{1}{l-1} \sum_{i=0}^{l-2} \left(100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right)$	$[-5.12; 5.12]^{10}$ $(l = 10)$
$GenMED_{1,2}$	$f_0 = \|\frac{1}{2}(\boldsymbol{x} - \boldsymbol{c^0})\|^d$ $f_1 = \|\frac{1}{2}(\boldsymbol{x} - \boldsymbol{c^1})\|^d$ $\boldsymbol{c^0} = (1, 0, 0, \dots) \quad \boldsymbol{c^1} = (0, 1, 0, 0, \dots) \quad d = 1, \frac{1}{2}$	$[-1;1]^{10}$ $(l = 10)$

Table 1: The MO problem test suite.

4. EXPLOITING MULTI-OBJECTIVE GRADIENT INFORMATION

4.1 Combined Objectives Repeated Line-search (CORL)

In this multi-objective version of gradient descent, the set of all Pareto-optimal improving directions for a vector function $\boldsymbol{f}$ is initially calculated [2]. To do so, every direction

$\hat{\boldsymbol{u}}$ is mapped to its directional derivative $\nabla_{\hat{\boldsymbol{u}}} \boldsymbol{f}$ from a unit l-dimensional hypersphere (where l is the problem dimensionality) to the surface and interior of an m-dimensional hyperellipsoid, where m is the number of objectives. The directions that are of interest, i.e., that lead to maximum improvement of all objectives, correspond to the negative non-dominated directional derivatives which lie in the intersection of the surface of the hyperellipsoid with $(-\infty, 0]^m$ (Figure 1). This set of directions can be described and sampled.
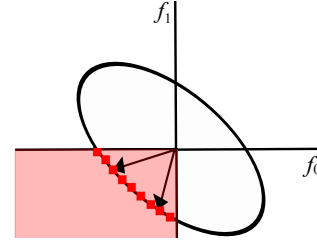

Figure 1: Part of the ellipsoid (in red) that includes the non-dominated improving directions.

One of these promising directions is chosen randomly and a multi-objective line search is performed along that direction. Typically, in a single-objective line search scheme, it is decided how far the algorithm should move in the chosen search direction to reach a local minimum along that direction, which maximizes the objective value difference between the starting point and the point where the line search ends. The multi-objective version of line-search is a natural generalization of that scheme, where the negative scaled Euclidean distance in objective space between the starting point $\boldsymbol{x}$ and the local minimum is used instead. The further $\boldsymbol{x} + \alpha\hat{\boldsymbol{u}}$ travels in objective space along the chosen direction $\hat{\boldsymbol{u}}$ while improving upon the starting point, the better. When the line search terminates, a new promising direction is sampled and a line search is repeated. This technique is called Combined Objectives Repeated Line-search (CORL).

To understand why a uniform distribution of directions is not necessarily preserved, one needs to look at the mapping of a direction $\hat{\boldsymbol{u}}$ from the parameter space to its directional derivative in the objective space. When one chooses a direction $\hat{\boldsymbol{u}}$, this direction is taken from the unit l-dimensional hypersphere. This direction is then collapsed onto the surface of the m-dimensional hyperellipsoid via a sequence of rotations, translations, and a projection. Therefore, the uniform distribution of the directions in the hypersphere is not necessarily preserved in the distribution of the directional derivatives on the hyperellipsoid, which depends on the orientation of the hyperellipsoid as well as its scale.

The lack of a spatially uniform distribution is possibly one of the reasons why in problems where the objectives have a large difference in difficulty, powerful single-objective methods such as conjugate gradients applied randomly or alternatingly to the different objectives perform better than CORL.

4.2 Rank-Based Combined Objectives Repeated Line-search (RB-CORL)

In this technique, we change the way directions are chosen for each member of the population. The multi-objective line search remains the same as in CORL.

4.2.1 A rank-based scheme

Aiming to exploit the analytically described set of Pareto-optimal improving directions in a truly uniform manner, we employ a diversification scheme, where as a first step the solutions are classified according to their multi-objective rank, which is calculated using the well-known domination-ranking scheme [6]. Consequently, the non-dominated solutions are of rank 0, the second best ones rank 1, and so on.
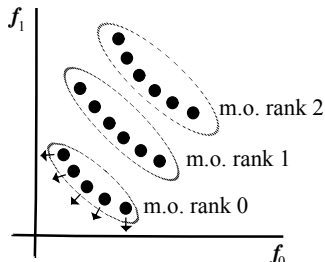

Figure 2: Directions assigned according to their multi-objective (m.o.) as well as their single-objective rank (within each multi-objective rank separately).

Subsequently, in each set of solutions that belong to the same rank, the solutions get a direction that depends on their ranks with respect to the individual objectives. In other words, the better a solution ranks in one objective compared to the other solutions that belong in the same multi-objective rank, the more favorable the direction for that objective (Figure 2). Each direction is associated with a solution in objective space that expresses this relation. Therefore, if we have k grouped solutions for one multi-objective rank, and r_0 is the rank of a solution s according to f_0, r_1 the rank of s according to f_1, r_{m-1} the rank of s according to f_{m-1}, the corresponding vector

$$\left(\frac{k-1-r_0}{k}, \frac{k-1-r_1}{k}, \ldots, \frac{k-1-r_{m-1}}{k} \right)$$

will be used to find a direction via sampling.

4.2.2 Monte-Carlo sampling

A uniform spatial distribution of the directional derivatives is what we need in order to assign the directions according to our rank-based scheme. We aim to partition uniformly the set of directional derivatives in objective space. The set of directional derivatives is parametrically described (see [2]) and therefore it can be sampled. This set can be easily visualized for $m = 2$ objectives, as in Figure 1. We initially normalize the directions to the observed ranges of solutions in the population, to ensure that the outcome is independent of the scaling of the objectives. To obtain a discrete uniformly distributed approximation, we employ Monte-Carlo sampling of this set of directions. Subsequently, each direction (which is an m-dimensional vector) is mapped via spherical coordinates to $m-1$ angles. For instance, if $m = 2$, each direction gets associated with one angle. We partition our objective space into equisized bins, each represented by an angle, and consequently by the direction associated with that angle.

Finally, we can map the direction we want from the rank-based scheme (by also converting it to spherical coordinates) to its closest available one from our obtained discrete set.

4.3 CORL versus RB-CORL

We first tested the new technique alone in a random-restart fashion, i.e., we applied it iteratively to 10000 randomly generated starting points. We are mostly interested in its behavior on problems with great difference in difficulty between objectives, such as BD_1 and BD_2.

We see that for BD_1, applying the multi-objective gradient method and choosing a random improving direction does not lead to a uniformly distributed improvement (Figure 3). CORL pushes all the members of the population to the vertical axis, as a result of the difficult landscape of the objective on the vertical axis, that uses the Rosenbrock function. The different level of difficulty as well as scale between the objectives translates into uneven spatial distribution of the directional derivatives in objective space. It demonstrates furthermore the need for a spatial uniformization of the choice of directions. The new technique is able to arrive much closer to the Pareto front of the problem, while achieving a much better spread of the population, as is the case also for BD_2. Moreover, it performs equally well on the GenMED problems (Figure 3). For BD_1 neither of the two techniques alone are able to actually find points on the optimal Pareto front, and therefore the hybridization of the new technique with an EA is considered necessary.
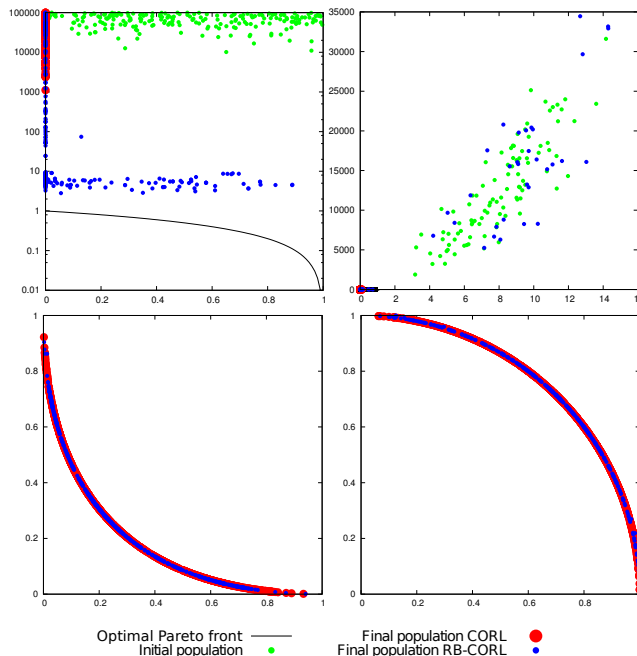

Figure 3: Distributions of optimized solutions using CORL and RB-CORL on BD_1 (upper left), BD_2 (upper right), convex (lower left) and concave (lower right) GenMED.

5. EXPERIMENT SETUP

5.1 Hybridization

5.1.1 iMAMaLGaM

Here we describe briefly the multi-objective EA we use for our hybridization schemes. We use an Estimation-of-Distribution Algorithm (EDA). EDAs are model-building EAs that aim to exploit structural features of the problem landscape. EDAs do so by estimating a probability distribution from selected solutions and sampling the estimated distribution to generate offspring.

The EDA we use is called iMAMaLGaM - incremental Multi-objective Adapted Maximum-Likelihood Gaussian Model. iMAMaLGaM, in contrast to well-known algorithms such as NSGA-II [6], is specifically designed for real-valued, multi-objective optimization. It should be noted that the variation operators of iMAMaLGaM could be easily used within NSGA-II as well and would approximately give the same results. In iMAMaLGaM, the selected solutions are grouped into clusters in objective space. For each cluster a multivariate Gaussian distribution is estimated using maximum-likelihood estimates, which are subsequently potentially adapted to prevent premature convergence. iMAMaLGaM samples then the Gaussian distributions to generate new solutions. An elitist archive of non-dominated solutions is also maintained. iMAMaLGaM has been shown to perform excellently in various benchmark as well as real-world problems [3].

5.1.2 Three hybridization schemes

In a multi-objective setting, it is not clear what is the most efficient way of integrating local search. For this reason, we consider different integration schemes. We compare these hybrid variants of iMAMaLGaM to iMAMaLGaM alone as well as to an adaptive hybridization scheme that was considered before for CORL [2]. The first two hybridization schemes are common in evolutionary optimization and local search [8].

Genetic Local Search 1 (GLS1): Given the large improvement obtained in initial stages of optimization using gradient techniques, in this hybridization scheme RB-CORL is applied only during initialization to all randomly generated solutions.

Genetic Local Search 2 (GLS2): In this scheme, RB-CORL is additionally called at the end of every evolutionary cycle to all members of the population.

Adaptive Genetic Local Search (AGLS): In this method RB-CORL is integrated in an adaptive hybridization scheme taken from literature [2], which exploits gradient information in three different ways. In this scheme, for maximum efficiency, the scheme decides during the run which exploitation method to use at the end of each evolutionary cycle, based on the number of improved solutions found by that method compared to the total number of evaluations used. The first exploitation method is called Random Objective Conjugate Gradients (ROCG); here, every time the method is called, one objective is chosen randomly and the conjugate gradients technique is performed for that objective only. The second exploitation method is called Alternating Objectives Repeated Line search (AORL). AORL performs a line search in the direction of the negative gradient of an alternatingly chosen objective. The third method of this hybrid scheme is CORL. We call this scheme *old AGLS*.

We made some changes to obtain a *new AGLS* version. Considering that both ROCG and AORL play approximately the same role during optimization - that is, they are able to find extreme points on the Pareto front, in this paper for the adaptive scheme we choose to only consider a variant of ROCG. Moreover, instead of applying conjugate gradients to a randomly chosen objective for a solution, we apply the method to only members of the population following again a rank-based scheme. More specifically, conjugate gradients is applied to (maximum) two solutions of each multi-objective rank that rank best at each objective individually. Therefore, conjugate gradients is applied to the solution that ranks best in one objective to further improve it. Finally, we replaced the old CORL with RB-CORL.

5.2 Benchmark problems

The problems we use to test the performance of iMAMaL-GaM and our hybridization schemes are described in Table 1.

5.2.1 Measuring performance

To measure performance we consider the set of the non-dominated solutions of the final population of the EDA, which we denote by S. To measure how well S approximates the optimal Pareto set $\mathcal{P_S}$ of each problem, we use a well-known indicator in multi-objective optimization which takes into account proximity to $\mathcal{P_S}$ as well as the spread along $\mathcal{P_S}$. This indicator, called Inverted Generational Distance (IGD) does so by computing the average distance to the closest solution in an approximation set S over all solutions in $\mathcal{P_S}$. The lower the value of the IGD, the better the overall approximation is, and IGD is 0 if and only if the approximation set and $\mathcal{P_S}$ are identical. IGD is denoted by $D_{P_f \to S}$, and formulated as follows:

$$D_{P_f \to S}(S) = \frac{1}{|\mathcal{P_S}|} \sum_{\boldsymbol{x}^1 \in \mathcal{P_S}} \min_{\boldsymbol{x}^0 \in S} \left\{ d(\boldsymbol{x}^0, \boldsymbol{x}^1) \right\},$$

where $d(\boldsymbol{x}^0, \boldsymbol{x}^1)$ is the Euclidean distance between the objective values $f(\boldsymbol{x}^0)$, $f(\boldsymbol{x}^1)$. For the problems in our test suite, given the ranges of the objectives for the optimal Pareto front configurations, a value of 0.01 for the $D_{P_f \to S}$ indicator corresponds to fronts that are quite close to the optimal Pareto front.

5.2.2 Algorithmic setup

We experimentally investigate the performance of RB-CORL as part of hybridization with iMAMaLGaM on the benchmark problems taken from literature shown in Table 1. We gave each algorithm a budget of 1,000,000 evaluations (per objective). The number of clusters was set to 20, since it was shown previously to yield good results. The cluster size for problems of dimension 10 was 11 and for problems of dimension 30 was 19. The selection percentile was set to $\tau = 0.35$, and the elitist archive target size was set to 1000. The evaluations for the computation of the gradient were not counted, considering that for our real-world application where we would like to use our hybrid version the gradient can be computed analytically.

5.3 Deformable image registration

5.3.1 Definition

Deformable image registration, i.e., finding the optimal transformation to align two images, is of great value in medical imaging. Healthcare specialists use it to combine images of different imaging modalities, as well as images taken at different time points. This allows them to get a clearer and more complete picture of the situation at hand as well as see how the anatomy of a patient changes over time. This is very important for example for radiation treatment planning, for diagnostic as well as for follow-up purposes.

In image registration, one image is referred to as the source image, whereas the other image is referred to as the target image. In the case of affine transformations (e.g. rotations or translations), the registration is called rigid. If non-affine transformations are needed to transform one image to the other (in the case of large anatomical differences, e.g., due to different positioning of the patient in image acquisitions) the registration is called deformable.

In this work, DIR is approached from a multi-objective optimization perspective [1]. This approach removes the need for pre-determining a singular combination of objectives, which is difficult to determine but necessary in existing registration methods. Moreover, since multiple outcomes are produced, this approach gives more options to the expert to choose the best outcome for the case at hand.

An image registration algorithm can be decomposed in three parts: the transformation model (the representation of all possible transformations from one image to the other image), the objective(s) to be optimized, and the optimization method.

5.3.2 Transformation model

We use a dual-dynamic transformation model based on B-splines that allows us to formulate differentiable objectives and calculate the gradient analytically.

B-splines (cubic polynomial functions used to model deformable objects) are widely used in image registration because they provide flexibility, transformation smoothness, and computational efficiency. In such transformation models, the source image is deformed by manipulating an underlying grid of control points that govern, via interpolation, the deformation and thereby the transformation of the image. In our dual-dynamic transformation model we have two grids associated with source- and target image, respectively, allowing us to handle large deformations. The correspondence between these two grids is established via a mapping from a 'virtual' fixed grid to the other two (Figure 4). More specifically, we loop over the virtual grid and we map the points to a location in the source- and a location in the target image. These locations are derived via the B-spline transformations T_s, T_t (which depend on the control points). Finally, the grey value at the obtained location is derived via pixel grey value interpolation. The sets of control points ϕ_s, ϕ_t that correspond to the source- and target- image are the parameters to be optimized.

5.3.3 Objectives

There are at least two objectives that are of interest in DIR and have to be minimized simultaneously. The first objective is quality of fit, i.e., the degree of similarity between the two images. The second objective is related to the smoothness of the transformation and aims to prevent implausible deformations.

The similarity measure we use is the sum of squared differences in pixel grey values between the source- and the target image:

$$Similarity = \sum_{x,y} \left[I_s(T_s(x,y;\phi)) - I_t(T_t(x,y;\phi)) \right]^2,$$

where I_s, I_t are the grey values at a point (x,y) on the source- and target image, respectively, and T_s, T_t are the B-spline transformation functions that determine the deformation of the source- and target- image. I_s, I_t are interpolated functions constructed again via B-splines which make them differentiable.

The second objective is introduced to measure smoothness, using the sum of squared second order derivatives of each B-spline transformation $\left(\frac{\partial^2 T}{\partial x^2} \right)^2$, $\left(\frac{\partial^2 T}{\partial y^2} \right)^2$. We formulate the smoothness objective as the sum of the squared differences of the smoothing terms evaluated at pairs of points in the source- and target image. This objective is easier to optimize than the similarity objective, as no displacement of control points is already (close to) an optimum.
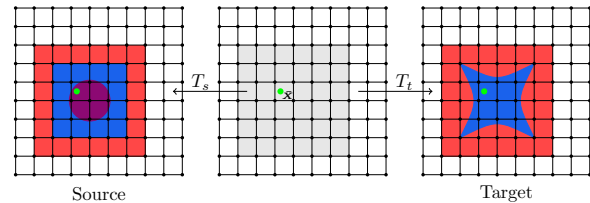

Figure 4: Dual-dynamic transformation model.

5.3.4 Optimization method

We apply three optimization methods to the DIR problems: iMAMaLGaM alone, the old AGLS scheme and and the new AGLS scheme. We use settings for the EDA which were shown to yield good results [3].

5.3.5 Algorithmic setup

We considered two test cases; the first one is a large-deformation registration case: prone-supine breast MRI registration. All MRI scans were acquired from a healthy volunteer. First, the two MRI scans were rigidly registered on the bony anatomy. Subsequently, one central pair of 2D slices was selected. The second case concerns MRI bladder registration, where there is also a large deformation of the bladder between image acquisitions, with an empty bladder in one image and a full bladder in the other. For both cases, we performed registration using two 8×8 grids of control points, which resulted in 256 parameters to be optimized. These settings are not necessarily sufficient to obtain the best possible registration result, but certainly sufficient to demonstrate the capabilities of the approach, and the impact of choosing different multi-objective optimizers.

6. RESULTS

6.1 Results on the benchmark problems

Results show that the use of gradient-based local search every generation, i.e., in the traditional GLS sense, is very
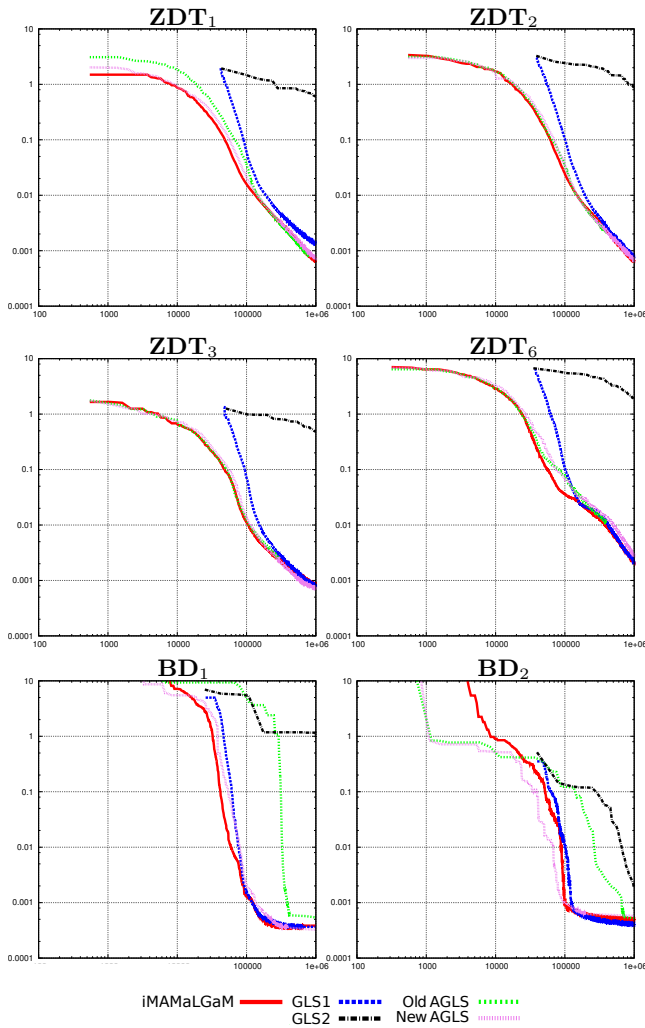
Figure 5: Convergence of all algorithms on 6 benchmark problems, averaged over 10 runs. Horizontal axis: number of evaluations (both objectives per evaluation). Vertical axis: $D_{\mathcal{P}_F \to \mathcal{S}}$.

computationally expensive (Figure 5). Note that evaluations required to approximate gradients are not even counted here as we are interested in the real-world application case where we have analytical expressions for the gradients of the objectives. iMAMaLGaM is capable of making many improvements all along the Pareto front at a much smaller budget of evaluations, which results in overall superior performance. This is in line with the common conception that EAs are well-suited and among the state-of-the-art for multi-objective optimization. Only by making use of the adaptive scheme for gradient integration that only uses gradient exploitation methods more often if they are really capable of making improvements more frequently than the EA, results can be obtained that come close to the performance of iMAMaLGaM. However, for problems BD_1 and BD_2, which exhibit most strongly the difference in objective complexity that we are interested in and that we redesigned AGLS for, the performance of the new version of AGLS is equal or even better than iMAMaLGaM. Moreover, the new AGLS clearly outperforms the old version of AGLS on these problems, which was one of our main objectives.

6.2 Results on the registration problems

For the breast MRI problem, iMAMaLGaM finds the most interesting solutions (i.e., solutions with a low value for the similarity measure, resulting in highly similar images). Compared to the old AGLS, we see that the new AGLS finds solutions which are slightly better in the interesting region, and it is closer to the final Pareto front of iMAMaLGaM (Figure 6). For the bladder MRI case, we can see a large difference between the old and the new AGLS; the new AGLS is able to obtain much better solutions regarding both objectives compared to the old AGLS. Also here, the EA is the most successful at finding solutions in the interesting region (Figure 6).

Although iMAMaLGaM finds the best solutions in a strict sense, the registration results of all algorithms on the breast MRI case are visually comparably good (Figure 7). For the bladder MRI case (Figure 8), the registration outcomes are overall not as good, however iMAMaLGaM and the new AGLS are able to obtain much better registration results compared to the old AGLS. It is worth noting that both the bladder and the breast MRI cases are difficult tasks, since there can be structures between these image acquisitions that move in the non-visible third plane, making it impossible to find a perfect match. Furthermore, many small structures do remain visible but are difficult to match using a low complexity deformation model of 8×8 control points. Therefore, it is possible that a larger number of control points is necessary to improve the registration results, especially in the bladder case. Additional guidance information such as contours could also be of great help.

Nonetheless, the above test cases demonstrate the capability of producing promising registration results using a multi-objective optimization approach to DIR. Moreover, the ability of RB-CORL to improve in an unbiased way all objectives, as shown in problems such as BD_1 and BD_2, can have a potentially large effect when it is part of a hybridization scheme: on hard, DIR problems like the bladder MRI case, the new AGLS is capable of obtaining results as good as of iMAMaLGaM. Lastly, its use could become more favourable in a hybridization scheme where not only the number of improvements are taken into account, but also the length of each improvement.
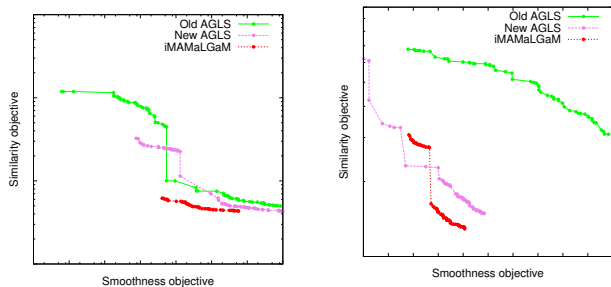

Figure 6: Final Pareto fronts for the breast MRI case (left) and the bladder MRI case (right).

7. DISCUSSION AND CONCLUSIONS

In this work, we developed and studied a new diversification technique for the exploitation of multi-objective gradient information, based on an analytically described set of Pareto-optimal improving directions. We assessed whether
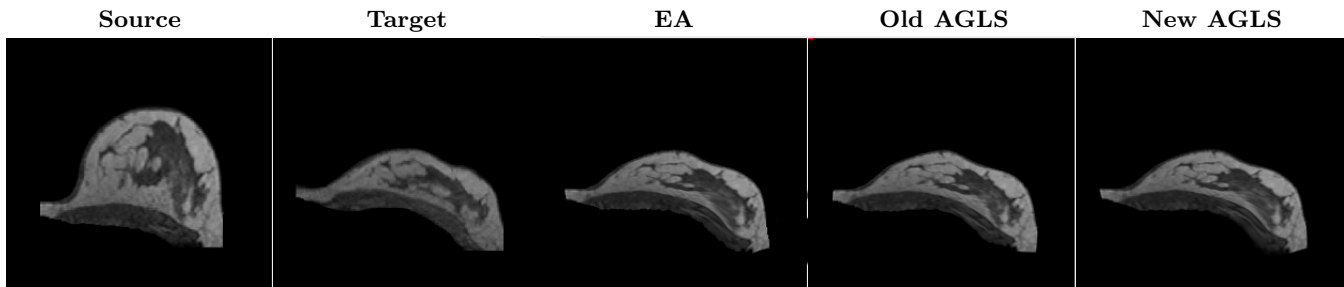
Source	Target	EA	Old AGLS	New AGLS

Figure 7: Selected DIR results on the breast MRI problem instance.

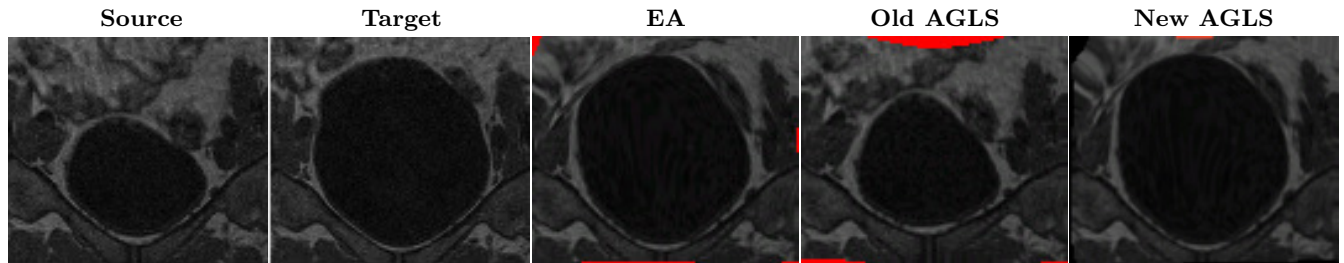| Source	Target	EA	Old AGLS	New AGLS

Figure 8: Selected DIR results on the bladder MRI problem instance.

this new technique can be combined efficiently with a multi-objective evolutionary algorithm, by considering three different hybridization schemes. Results showed improvement over existing adaptive hybridization schemes, especially for the problems we are particularly interested in: when there is a large difference in difficulty and/or scale between the objectives to be optimized, as is the case for the deformable image registration problem. The unbiased improvement of all objectives in a multi-objective gradient technique as part of a hybridization scheme was indeed successful at solving the tested hard DIR cases, as opposed to a previously introduced adaptive hybridization scheme. Nonetheless, the EA is still the most robust and powerful optimization algorithm, having the best balance between proximity and diversity. This is testimony again to the power of state-of-the-art EAs for multi-objective optimization, even for medical image processing applications, that have predominantly been solved using other types of algorithms, albeit from a single-objective optimization perspective.

For problems such as DIR, where a particular region of the Pareto front is of interest, it would also be interesting to investigate methods that are able to discover individual solutions on only part of the Pareto front, using linear/Chebyshev combinations of objectives.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] T. Alderliesten, J.-J. Sonke, and P. A. N. Bosman. Multi-objective optimization for deformable image registration: proof of concept. In *Proceedings of SPIE Medical Imaging*, pages 831420–831420, 2012.

[2] P. A. N. Bosman. On gradients and hybrid evolutionary algorithms for real-valued multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 16(1):51–69, 2012.

[3] P. A. N. Bosman and T. Alderliesten. Incremental Gaussian model-building in multi-objective edas with an application to deformable image registration. In *Proceedings of Genetic and Evolutionary Computation - GECCO-2012*, pages 241–248. ACM, 2012.

[4] M. Brown and R. E. Smith. Effective use of directional information in multi-objective evolutionary computation. In *Proceedings of Genetic and Evolutionary Computation - GECCO 2003*, pages 778–789. Springer, 2003.

[5] C. A. C. Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary algorithms for solving multi-objective problems*, volume 242. Springer, 2002.

[6] 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.

[7] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

[8] A. Jaszkiewicz. On the performance of multiple-objective genetic local search on the 0/1 knapsack problem-a comparative experiment. *IEEE Transactions on Evolutionary Computation*, 6(4):402–412, 2002.

[9] M. R. Kaus and K. K. Brock. Deformable image registration for radiation therapy planning: algorithms and applications. In C. T. Leondes, editor, *Biomechanical systems technology: computational methods*, pages 1–28. World Scientific Publishing Company, Incorporated, 2007.

[10] J. D. Knowles and D. W. Corne. M-PAES: A memetic algorithm for multiobjective optimization. In *Proceedings of Congress on Evolutionary Computation - CEC 2000*, volume 1, pages 325–332. IEEE, 2000.

[11] M. Lahanas, D. Baltas, and S. Giannouli. Global convergence analysis of fast multiobjective gradient-based dose optimization algorithms for high-dose-rate brachytherapy. *Physics in Medicine and Biology*, 48(5):599, 2003.

[12] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195, 2000.