Directed Search Method for Indicator-Based Multi-objective Evolutionary Algorithms

Víctor Adrián Sosa H.*
CINVESTAV-IPN
Dpto. de Computacion
Mexico City, MEXICO
msosa@computacion.cs.cinvestav.mx

Oliver Schütze†
CINVESTAV-IPN
Dpto. de Computacion
Mexico City, MEXICO
schuetze@cs.cinvestav.mx

Günter Rudoph‡
TU-Dortmund
Fakultät für Informatik
Dortmund, GERMANY
Guenter.Rudolph@tu-dortmund.de

Heike Trautmann§
University of Münster
Statistics and Information Systems
Münster,GERMANY
trautmann@wi.uni-muenster.de

ABSTRACT

Indicator based evolutionary algorithms have caught the interest of many researchers for the treatment of multi-objective optimization problems in the recent past since they deliver the desired approximation of the solution set and due to a usually better performance compared to dominance based algorithms. Nevertheless, these methods still suffer the drawback that many function evaluations are required to obtain a suitable representation of the solution set. The aim of this study is to present the Directed Search (DS) Method as local searcher within global indicator based optimization algorithms. For this, we will present the DS in the context of hypervolume maximization leading to both a new local search algorithm and a new memetic algorithm. Further, we will present first attempts to adapt the DS to a class of parameter dependent problems.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization

General Terms

Algorithms, Design, Performance, Theory

*Student
†Advisor
‡Advisor
§Advisor

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.

Keywords

multi-objective optimization, evolutionary computing, directed search, memetic algorithms

1. INTRODUCTION

In many applications one is faced with the problem that several objectives have to be optimized leading to a *multi-objective optimization problem* (MOP). Since the solution set of a MOP does not consist of a single solution but forms a $(k-1)$-dimensional manifold, where k is the number of objectives, the numerical treatment of such problems is a challenging task. Among those algorithms, specialized evolutionary algorithms (evolutionary multi-objective algorithms, EMOAs) have caught the interest of many researchers in the recent past [4]. Reasons for this (among others) are that EMOAs are applicable to a wide range of problems, are of global nature, and allow to compute a finite size representation of the Pareto set in one single run of the algorithm. Among EMOAs there is a recent trend in the design of algorithms that are based on a particular performance indicator. Reasons for that include the improvement of the numerical treatment of the problem (e.g., the speed up of the convergence rate) and the fact that such optimal archives (i.e., optimal w.r.t. the given indicator) are in certain cases most appropriate for the related decision making problem.

The goal of this study is to use and adapt the recently developed *Directed Search Method* (DS) as local searcher within indicator based EMOAs. The DS allows to steer the search into any direction given in objective space which is well-suited for the problem at hand: Given an indicator, the MOP is (implicitly) transformed into a single-objective optimization problem. Hence, for every point out of the population that is selected for local search there exists an 'optimal' (i.e., greedy) search direction to locally improve the indicator value. Note that most performance indicators are defined in objective space (e.g., the hypervolume indicator [10] or the averaged Hausdorff distance [7]), and hence, the

use of the DS comes as a natural choice.
In this paper, we first recall some background (Sec. 2) and discuss further on how the DS can be used in the context of hypervolume maximization (Sec. 3). Further, we present some first results to adapt the DS to the context of parameter dependent MOPs (Sec. 4). Finally, we draw our conclusions and give paths for future research in Sec. 5.

2. BACKGROUND

A MOP can be stated as

$$\min_{x \in Q}\{F(x)\}, \tag{1}$$

where F is given by the vector of the objective functions $F : Q \to \mathbb{R}^k$, $F(x) = (f_1(x), \ldots, f_k(x))$, and where each objective is given by $f_i : Q \to \mathbb{R}$. We say that a vector $x \in Q$ dominates a vector $y \in Q$ ($x \prec y$) if $f_i(x) \leq f_i(y)$ $\forall i = 1, \ldots, k$ and $f_j(x) < f_j(y)$ for a $j \in \{1, \ldots, k\}$. A vector x is called (Pareto) optimal if there exists no $z \in Q$ such that $z \prec x$. The set of Pareto optimal points is called the Pareto set P_Q, and its image $F(P_Q)$ the Pareto front.
Recently, the DS has been proposed to allow to steer the search from a given point into a desired direction $d \in \mathbb{R}^k$ in objective space [8]. To be more precise, given a point $x_0 \in \mathbb{R}^n$, a search direction $\nu \in \mathbb{R}^n$ is sought such that

$$\lim_{t \searrow 0} \frac{f_i(x_0 + t\nu) - f_i(x_0)}{t} = d_i, \quad i = 1, \ldots, k. \tag{2}$$

Such a direction vector ν solves the system $J(x_0)\nu = d$ of linear equations, where $J(x)$ denotes the Jacobian of F at x. Since typically $k < n$, the equation is underdetermined. Among the solutions of $J(x_0)\nu = d$, the one with the least 2-norm can be viewed as the greedy direction for the given context. It is given by $\nu_+ := J(x)^+ d$, where $J(x)^+$ denotes the pseudo inverse of $J(x)$. Since there is no restriction on d the search can be steered in any direction, e.g., toward and along the Pareto set. See [8] for a Pareto descent method and a continuation method based on DS. In [5] a gradient free version of the DS is presented.

3. DS TO MAXIMIZE THE HYPERVOLUME

First we address the problem to maximize the hypervolume using DS (see [9] for a more detailed study). For this, we first divide the objective space into three regions:

1. **Region I** The objective vector $F(x)$ is 'far away' from $F(P_Q)$. A greedy search toward the rough location of $F(P_Q)$ is desired.

2. **Region II** $F(x)$ is 'in between', i.e., neither far away nor near to $F(P_Q)$. A descent direction has to be selected such that a movement in that direction maximizes the hypervolume.

3. **Region III** $F(x)$ is 'near' to $F(P_Q)$. A movement toward $F(P_Q)$ will lead to 'non-significant' improvements of the hypervolume. Instead, a search along $F(P_Q)$ will be performed.

The region assignment for a given candidate solution x can be done by considering the size of the descent cone which is in turn related to the angle between the 'objectives' gradients. Next we describe the use of DS for bi-objective problems in each region.

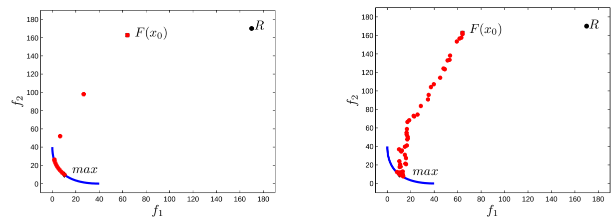

Figure 1: Result of the HVDS (left) and a hypervolume hill climber (right) on a convex MOP.

	SMS-EMOA		SMS-EMOA-HVDS	
	Average	Deviation	Average	Deviation
Convex*	2003.867	68.956	2161.668	18.039
Dent	17.234	0.031	17.245	0.023
ZDT1*	105.015	0.948	109.532	0.005
ZDT2*	97.592	2.965	109.230	0.040
ZDT3*	113.771	1.857	116.097	1.948
ZDT4	76.536	13.485	76.030	15.955

Table 1: Numerical results of SMS-EMOA with and without HVDS as local searcher (using a budget of 2500 function evaluations). Average over 20 runs.

Using DS it can be shown that for points x far away from P_Q large improvements in image space using one iteration step can only be obtained for $d_I = (-1, -|\mu|)^T$, where $||\nabla f_2(x)||_2 = |\mu| ||\nabla f_1(x)||_2$, which represents a movement toward the Pareto front. Hence, we suggest to use DS using direction d_I.
For x in Region II, the task is to find a search direction $d_{II} <_p 0$ such that a movement in that direction maximizes the hypervolume. It can be shown for one element archives the greedy solution is given by $d_{II} = F(x) - R$. For general archives, R is a given reference point for the extreme points and the nadir point of the two neighboring solutions in case $F(x)$ is located between two archive objective vectors.
In case x is already near to the Pareto front, a replacement of x by a dominating solution will increase the hypervolume since this indicator is Pareto compliant, however, only by a non-significant value. Instead, we propose to perform a search along the Pareto front. Using DS, this can be done by linearizing the Pareto front at $F(x)$ as done for the DS continuation [8]. Then, assuming (locally) a linear front, a one-dimensional optimization problem has to be solved in order to maximize the hypervolume. If the resulting step size is smaller than a given threshold, the algorithm can be stopped since no more improvements can be expected.
The resulting algorithm, HVDS (Hypervolume based Directed Search), can then constructed based on the above guidelines and then either be used as a standalone algorithm or as a local searcher within an EMOA. Figure 1 shows one result of the HVDS as standalone algorithm on a MOP with a convex Pareto front. For comparison, a simple hypervolume based hill climber has been used. Table 1 shows experimental results of SMS-EMOA [2] with and without HVDS on six benchmark models. Here, the hybrid wins significantly in 4 out of 6 models and loses in 1. Statistically significant differences due to the Wilcoxon-Rank-Sum Test with $\alpha = 0.05$ are marked with (*).

4. DS FOR PARAMETER DEPENDENT MOPS

In the sequel we consider the following parameter dependent multi-objective problems (PMOPs):

$$\min_{x\in Q}\{F_\lambda(x)\}, \tag{3}$$

where F is as in (1) and λ is an external parameter[1] within a given set $\Lambda \subset \mathbb{R}^l$. Note that for a fixed value λ problem (3) reads as the original problem (1). We denote by $P_{Q,\Lambda}$ the family of Pareto sets and by $F(P_{Q,\Lambda})$ the respective family of Pareto fronts.

To apply DS to the new problem, we have to use a trick: We will formally treat λ as 'normal' parameter leading to the mapping $F: \mathbb{R}^{n+l} \to \mathbb{R}^{k+l}$ with

$$F(x,\lambda) = \begin{pmatrix} f_1(x,\lambda) \\ \vdots \\ f_k(x,\lambda) \\ \lambda \end{pmatrix} =: \begin{pmatrix} g_1(x,\lambda) \\ \vdots \\ g_{k+l}(x,\lambda) \end{pmatrix}. \tag{4}$$

Now we can adapt DS to the current context: Given a point $(x,\lambda) \in \mathbb{R}^{n+l}$ in parameter space and a vector $d = (d_f, d_\lambda)^T \in \mathbb{R}^{k+l}$ in objective space, the task is to find a direction $\nu = (\nu_f, \nu_\lambda) \in \mathbb{R}^{n+l}$ such that $\forall i \in \{1,\dots,k+l\}$:

$$\lim_{t\searrow 0} \frac{g_i((x,\lambda)+t\nu) - g_i(x,\lambda)}{t} = \langle \nabla g_i(x,\lambda), \nu\rangle = d_i. \tag{5}$$

In matrix vector notation, Equation (5) can be written as

$$\begin{pmatrix} J_x & J_\lambda \\ 0 & I_l \end{pmatrix} \begin{pmatrix} \nu_f \\ \nu_\lambda \end{pmatrix} = \begin{pmatrix} d_f \\ d_\lambda \end{pmatrix}, \tag{6}$$

where J_x (J_λ) denotes the derivative of F with respect to $x(\lambda)$ at (x,λ) and I_l the $l \times l$ identity matrix. Using (6), we are now in the position to steer the search in any direction given in objective space where we can separate between 'f-space' (i.e., the objective space for a particular value of λ given by d_f) and 'λ-space'. For instance, the greedy solution ν_+ to perform a search in d-direction is given by

$$\nu_+(x,d) = J^+ d = \begin{pmatrix} J_x^+ & -J_x^+ J_\lambda \\ 0 & I_l \end{pmatrix} \begin{pmatrix} d_f \\ d_\lambda \end{pmatrix} = \begin{pmatrix} J_x^+ d_f - J_x^+ J_\lambda d_\lambda \\ d_\lambda \end{pmatrix} \tag{7}$$

DS Descent Method.

Assume we are given a point (x_0,λ_0) and the task is to steer the search into direction $d = (d_f, d_\lambda)$, where all entries of d_f are negative (i.e., a 'descent direction' in f-space). Then a movement in that direction is related to the numerical solution of the following initial value problem:

$$\begin{aligned} z(0) &= (x_0,\lambda_0)^T \in \mathbb{R}^{n+l} \\ \dot z(t) &= \nu_+(x(t),d) \end{aligned} \tag{λ-DS(x_0,λ_0,d)}$$

We define the critical point of the solution γ of λ-DS(x_0,λ_0,d) as the first point where no movement in d-direction can be performed (which does not have to be the end point of γ). Such points are always the boundary points of problem (3) but do not have to be KKT points of problem (1) for the critical value λ^*. The following discussion shows the relation to the normal boundary intersection (NBI, [3]) which is a well-known scalarization method for MOPs.

The NBI subproblem for problem (3) can be written as:

$$\begin{aligned} \max_{x,\lambda,t} \quad & t \\ \text{s.t.} \quad & F(x,\lambda) = F(x_0,\lambda_0) + td \\ & x \in Q, \quad \lambda \in \Lambda \end{aligned} \tag{NBI(x_0,λ_0,d)}$$

Using (NBI(x_0,λ_0,d)), we can state the following result (we omit here the all the proofs due to space limitations).

PROPOSITION 1. *Let $z^* = (x^*,\lambda^*)$ be the critical point of (λ-DS(x_0,λ_0,d)), then it is a local solution of ($NBI(x_0,\lambda_0,d)$).*

Hence, following γ up to the critical point leads by the above result to the maximal movement in d-direction. It remains to detect z^*. Since it is a boundary point, it follows that the matrix $(J_x\ J_\lambda) \in \mathbb{R}^{k\times(n+l)}$ has to have rank $k-1$ which can easily be checked numerically by looking at the condition number of the matrix during the numerical integration of (λ-DS(x_0,λ_0,d)).

DS Continuation Method.

Next, it is desirable to use DS to move *along* the set of interest $P_{Q,\Lambda}$. In [8], this is realized by performing a move in direction d that points along the linearized Pareto front. This is also possible for PMOPs due to the following result. Hereby, $I(A)$ denotes the interior of a set A.

PROPOSITION 2. *Let $(x,\lambda) \in I(Q)\times\Lambda$ and x be a KKT point of F_λ with weight vector α (i.e., $\alpha_i \geq 0$, $i = 1,\dots,k$, $\sum_{i=i}^k \alpha_i = 1$ and $\sum_{i=1}^k \nabla f_i(x,\lambda) = 0$). Then*

$$\eta = (\alpha, -{J_\lambda}^T \alpha)^T, \tag{8}$$

is orthogonal to the linearization of $F(P_{Q,\Lambda})$ at $F_\lambda(x)$.

Hence, using (8), one can perform a movement along $P_{Q,\Lambda}$ as for the original DS. Since a movement in f-space is analog to the one for static MOPs, we concentrate here on the case that we are given a KKT point for a fixed value of λ and aim for a movement along λ-space (i.e., orthogonal to the Pareto set of F_λ). In particular, we suggest to use the following predictor-corrector (PC) method: In the predictor step, compute the predictor direction d_{pred} and perform a step in that direction using DS. This direction is given by $d_{pred} = (\alpha, d_\lambda)^T$ where d_λ solves the equation $\alpha^T J_\lambda d_\lambda = \mu||\alpha||_2^2$. In a next step, this point can be projected back to $P_{Q,\Lambda}$ via DS using the corrector direction $d_{corr} = (-\alpha, 0)$, that is, $d_f = -\alpha$ (orthogonal projection to the Pareto front of $F_{\tilde\lambda}$, where $\tilde\lambda$ is the new value in Λ) and $d_\lambda = 0$ (no changes in λ-space allowed).

As an example, we consider the PMOP S5 proposed in [6] where Q is 2-dimensional and Λ is one-dimensional. Figure 2 shows the sets $P_{Q,\Lambda}$ and its image together with numerical result of the novel PC method where the process has been started with a Pareto optimal solution x_0 for the value $\lambda_0 = -1$. Note that the above method does not require any 2nd gradient information which is the case for other PC methods. To solve the above problem, one can e.g. apply classical continuation methods (e.g., [1]) on the zero finding problem

$$\tilde F(x,\lambda) = \sum_{i=1}^k \alpha_i \nabla_x f_i(x,\lambda) = 0. \tag{9}$$

Using the continuation method on (9) a result very similar to the one in Figure 2 can be obtained, however, with a

[1]As an example, consider bi-objective problem 'F=(speed, safety)' of a vehicle where the side wind has to be considered within a certain range.

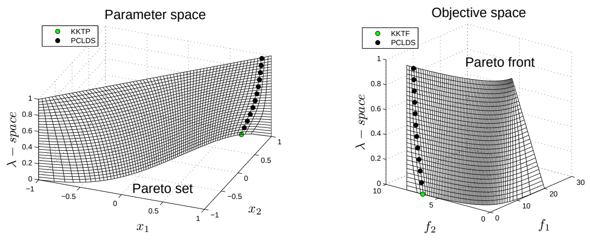

Figure 2: Continuation method through λ-space using λ-DS.

Table 2: Comparison of the cost of classical continuation and the DS approach to solve the problem shown in Figure 2.

Approach	Function	Jacobian	Hessian
Continuation	87	216	87
DS	255	11	0

much higher effort. Table 2 shows the cost of the methods apparently, the DS approach has a much lower overall cost.

Gradient Free DS for PMOPs.

One possible drawback of the DS is that it requires gradient information. The following result, however, shows a possible remedy. If neighboring solutions are at hand (which is typically the case for EMOAs), ν can be computed for free in terms of additional function evaluations.
Define the matrix $\mathcal{F}(x, \lambda) \in \mathbb{R}^{(k+l)\times r}$ as follows

$$\mathcal{F}(x,\lambda) := (\langle \nabla g_i(x,\lambda), \nu_j \rangle)_{\substack{i=1,\dots,k+l \\ j=1,\dots,r}} \tag{10}$$

Hence, every entry m_{ij} of $\mathcal{F}$ is defined by the directional derivative of objective g_i in direction ν_j. The following result is central for the gradient free computation of ν.

Proposition 3. *Let $(x, \lambda) \in \mathbb{R}^{k+l}$, $\nu_1, \dots, \nu_r \in \mathbb{R}^{k+l}$ be linear independent and $w \in \mathbb{R}^r$ such that $\nu := \sum_{i=1}^r w_i \nu_i$. Then $\mathcal{F}(x, \lambda)w = J(x, \lambda)v$.*

Hence, the gradient free DS can be realized as follows: Given a point (x_0, λ_0) where the local search has to be performed as well as r further test points (x_i, λ_i), $i = 1, \dots, r$, one can first approximate the entries of $\mathcal{F}$ via

$$m_{ij} = \langle \nabla g_i(x,\lambda), \nu_j \rangle = \frac{g_i(x_i,\lambda_j) - g_i(x_0,\lambda_0)}{\|(x_j,\lambda_j) - (x_0,\lambda_0)\|_2} \tag{11}$$

Then, ν can be computed by solving

$$\mathcal{F}(x,\lambda)w = d, \tag{12}$$

and setting

$$\nu := \sum_{i=1}^{r} w_i \nu_i. \tag{13}$$

It is important to note that only $r = k + l$ test points are needed in order to find a direction ν that solves (12). In contrast, a total of $(n + l) * k$ function evaluations are needed when approximating the Jacobian of F using finite differences. Thus, we can say that if $r > k + l$ test points are available in the vicinity of (x_0, λ_0), the search direction comes for free.

5. CONCLUSIONS AND FUTURE WORK

Here we have considered the DS as local searcher within EMOAs. We conjecture that DS is in particular beneficial for indicator based algorithms since for such problems greedy directions exist that are defined in objective space which makes DS a natural choice. We have applied the DS in the context of hypervolume maximization and have made a first attempt to adjust it to the context of PMOPs. There are many interesting aspects for future work. Though the first results are promising, further tests have to be performed and design parameters to be optimized. Further, other indicators than hypervolume have to be considered.

6. ACKNOWLEDGMENTS

The first author acknowledges support from CONACyT through a scholarship to pursue graduate studies at the Computer Science Department of CINVESTAV-IPN. The second author acknowledges support from CONACyT project no. 128554. The third and fourth author acknowledge support from DFG, project no. TR 891/5-1.

7. REFERENCES

[1] E. L. Allgower and K. Georg. *Numerical Continuation Methods*. Springer, 1990.

[2] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 2006.

[3] I. Das and J. Dennis. *Normal-boundary intersection*: A new method for generating the Pareto surface in nonlinear MOP. *SIAM Journal of Optimization*, 8:631–657, 1998.

[4] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001. ISBN 0-471-87339-X.

[5] A. Lara, S. Alvarado, S. Salomon, G. Avigad, C. A. C. Coello, and O. Schütze. The gradient free directed search method as local search within MOEAs. In *EVOLVE*, pages 153–168, 2013.

[6] O. Schütze. *Set Oriented Methods for Global Optimization*. PhD thesis, Universität Paderborn, 2004.

[7] O. Schütze, X. Esquivel, A. Lara, and C. A. C. Coello. Using the averaged Hausdorff distance as a performance measure in evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012.

[8] O. Schütze, A. Lara, and C. A. C. Coello. The directed search method for unconstrained MOPs. In *Proceedings of the EVOLVE*, 2011.

[9] V. A. Sosa, O. Schütze, G. Rudolph, and H. Trautmann. The Directed Search Method for Pareto Front Approximations with Maximum Dominated Hypervolume. *Accepted in EVOLVE, Leiden, The Netherlands*, 2013.

[10] E. Zitzler and L. Thiele. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.