

Dominance-Based Multiobjective Simulated Annealing

Kevin I. Smith, Richard M. Everson, *Member, IEEE*, Jonathan E. Fieldsend, *Member, IEEE*, Chris Murphy, and Rashmi Misra

Abstract—Simulated annealing is a provably convergent optimizer for single-objective problems. Previously proposed multiobjective extensions have mostly taken the form of a single-objective simulated annealer optimizing a composite function of the objectives. We propose a multiobjective simulated annealer utilizing the relative dominance of a solution as the system energy for optimization, eliminating problems associated with composite objective functions. We also propose a method for choosing perturbation scalings promoting search both towards and across the Pareto front.

We illustrate the simulated annealer's performance on a suite of standard test problems and provide comparisons with another multiobjective simulated annealer and the NSGA-II genetic algorithm. The new simulated annealer is shown to promote rapid convergence to the true Pareto front with a good coverage of solutions across it comparing favorably with the other algorithms.

An application of the simulated annealer to an industrial problem, the optimization of a code-division-multiple access (CDMA) mobile telecommunications network's air interface, is presented and the simulated annealer is shown to generate nondominated solutions with an even and dense coverage that outperforms single objective genetic algorithm optimizers.

Index Terms—Code-division multiple-access (CDMA) networks, dominance, multiple objectives, simulated annealing.

I. INTRODUCTION

A POPULAR and robust algorithm for solving single-objective optimization problems (those in which the user cares only about a single dependant variable of the system) is simulated annealing (SA) [1], [2]. Geman and Geman [3] provided a proof that SA, if annealed sufficiently slowly, converges to the global optimum, and although the required cooling rate is infeasibly slow for most purposes, SA often gives well converged results when run with a faster cooling schedule. It is frequently the case in optimization problems, however, that there are several objectives of the system which the user is interested in optimizing simultaneously. Clearly, simultaneous optimization of several objectives is usually impossible and the curve (for two objectives) or surface (for three or more objectives)

that describes the tradeoff between objectives is known as the Pareto-front. Although there are several well developed genetic algorithms and evolutionary schemes to address such multiobjective problems (see [4] and [5] for recent reviews), SA does not, in its usual formulation, provide a method for optimizing more than a single objective. SA has been adapted to multiobjective problems by combining the objectives into a single-objective function [6]–[10]; however, these methods either damage the proof of convergence, or are limited (potentially severely) in their ability to fully explore the tradeoff surface.

We propose a modified SA algorithm which maps the optimization of multiple objectives to a single-objective optimization using the true tradeoff surface, maintaining the convergence properties of the single-objective annealer, while encouraging exploration of the full tradeoff surface. A method of practical implementation is also described, using the available nondominated data points from the current optimization to overcome the limitation that the true tradeoff surface is unavailable for most real-world problems.

In this paper, following some introductory material in Section II, we start by briefly discussing methods that combine objectives into a single composite objective. In Section III, we describe our dominance-based SA algorithm and, in Section IV, methods are described for improving the quality of the optimization energy measure when the available data points are few. Choosing an efficient scale for perturbations is an important component of scalar SA algorithms. The issue is further complicated in multiobjective algorithms because a perturbation may not only move the current state closer to or further from the Pareto front, but also transversally (i.e., across the front). In Section VI, we describe a method for setting the scale of perturbations and other runtime parameters. Results showing that the algorithm converges on a range of standard test problems are given in Section VII, and we show that the algorithm compares favorably with both the popular NSGA-II multiobjective genetic algorithm [11] and a multiobjective simulated annealer suggested by Nam and Park [8]. In Section VIII, we present results demonstrating the simulated annealer's performance on the optimization of the air interface of a code-division multiple-access (CDMA) network in the mobile telecommunications domain. We draw conclusions in Section IX. A preliminary report on this work appeared in [12]; here, we provide additional detail on the theoretical foundations of the algorithm and present extensive empirical results comparing the algorithm with the NSGA-II genetic algorithm and the Nam and Park simulated annealer, together with the application to CDMA networks.

Manuscript received April 3, 2006; revised October 17, 2006 and February 22, 2007. The work of K. Smith was supported in part by the School of Engineering, Computer Science and Mathematics, University of Exeter and Motorola. The work of J. Fieldsend was supported in part by the Engineering and Physical Sciences Research Council under Grant GR/R24357/01.

K. I. Smith, R. M. Everson, and J. E. Fieldsend are with the Department of Computer Science, University of Exeter, Exeter EX4 4QF, U.K. (e-mail: K.I.Smith@exeter.ac.uk; R.M.Everson@exeter.ac.uk; J.E.Fieldsend@exeter.ac.uk).

C. Murphy and R. Misra are with the Motorola, Swindon, SN25 4XY, U.K. (e-mail: Chris.Murphy@motorola.com; Rashmi.Misra@motorola.com).

Digital Object Identifier 10.1109/TEVC.2007.904345

II. BACKGROUND

A. Dominance and Pareto Optimality

In multiobjective optimization, we attempt to simultaneously maximize or minimize D objectives, y_i , which are functions of P variable parameters or decision variables, $\mathbf{x} = (x_1, x_2, \dots, x_P)$

$$y_i = f_i(\mathbf{x}), \quad i = 1, \dots, D. \quad (1)$$

Without loss of generality, we assume that the objectives are to be minimized, so that the multiobjective optimization problem may be expressed as

$$\text{minimize } \mathbf{y} = \mathbf{f}(\mathbf{x}) \equiv (f_1(\mathbf{x}), \dots, f_D(\mathbf{x})). \quad (2)$$

The idea of dominance is generally used to compare two solutions \mathbf{f} and \mathbf{g} . If \mathbf{f} is no worse for all objectives than \mathbf{g} and wholly better for at least one objective, it is said that \mathbf{f} *dominates* \mathbf{g} , written $\mathbf{f} \prec \mathbf{g}$. Thus, $\mathbf{f} \prec \mathbf{g}$ iff

$$\begin{aligned} f_i &\leq g_i \forall i = 1, \dots, D \text{ and} \\ f_i &< g_i \text{ for at least one } i. \end{aligned} \quad (3)$$

By a slight abuse of notation, dominance in *objective* space is extended to *parameter* space; thus it is said that $\mathbf{a} \prec \mathbf{b}$ iff $\mathbf{f}(\mathbf{a}) \prec \mathbf{f}(\mathbf{b})$.

The dominates relation is not a total order and two solutions are *mutually nondominating* if neither dominates the other. A set of F solutions is said to be a nondominating set if no element of the set dominates any other

$$\mathbf{a} \not\prec \mathbf{b} \forall \mathbf{a}, \mathbf{b} \in F. \quad (4)$$

A solution is said to be globally nondominated, or Pareto-optimal, if no other feasible solution dominates it. The set of all Pareto-optimal solutions is known as the Pareto-optimal front, or the Pareto set \mathcal{P} ; solutions in the Pareto set represent the possible optimal tradeoffs between competing objectives. A human operator can select a solution with a knowledge of the tradeoffs involved once this set has been revealed. Heuristic procedures, such as multiple objective evolutionary algorithms and the multiobjective simulated annealing (MOSA) algorithms discussed here, yield sets of mutually nondominating solutions which will be only an approximation to the true Pareto front. Some care with terminology is therefore required, and in this paper, the set produced by such an algorithm is referred to as the estimated Pareto front, which we denote by F .

B. Simulated Annealing (SA)

Simulated annealing (SA), introduced by Kirkpatrick *et al.* [1] may be thought of as the computational analogue of slowly cooling a metal so that it adopts a low-energy, crystalline state. At high temperatures, particles are free to move around, whereas as the temperature is lowered they are increasingly confined due to the high energy cost of movement. It is physically appealing to call the function to be minimized the *energy*, $E(\mathbf{x})$,

of the state \mathbf{x} , and to introduce a parameter T , the computational temperature, which is lowered throughout the simulation according to an annealing schedule. At each T , the SA algorithm aims to draw samples from the equilibrium distribution $\pi_T(\mathbf{x}) \propto \exp\{-E(\mathbf{x})/T\}$. If $T \rightarrow 0$ is sufficiently slow, an increasing proportion of the probability mass of π_T , is concentrated in the region of the global minimum of E , so eventually, assuming a sufficiently slow annealing schedule is used, any sample from π_T will almost surely lie at the minimum of E .

Algorithm 1: Simulated annealing

Inputs:

$\{L_k\}_{k=1}^K$ Sequence of epoch durations
 $\{T_k\}_{k=1}^K$ Sequence of temperatures, $T_{k+1} < T_k$
 \mathbf{x} Initial feasible solution

```

1:  for  $k := 1, \dots, K$ 
2:    for  $i := 1, \dots, L_k$ 
3:       $\mathbf{x}' := \text{perturb}(\mathbf{x})$ 
4:       $\delta E(\mathbf{x}', \mathbf{x}) := E(\mathbf{x}') - E(\mathbf{x})$ 
5:       $u := \text{rand}(0, 1)$ 
6:      if  $u < \min(1, \exp(-\delta E(\mathbf{x}', \mathbf{x})/T_k))$ 
7:         $\mathbf{x} := \mathbf{x}'$ 
8:      end
9:    end
10:  end
```

Sampling from the equilibrium distribution $\pi_T(\mathbf{x})$ at any particular temperature is usually achieved by Metropolis–Hastings sampling [2], which involves making proposals \mathbf{x}' that are accepted with probability

$$A = \min(1, \exp\{-\delta E(\mathbf{x}', \mathbf{x})/T\}) \quad (5)$$

where

$$\delta E(\mathbf{x}', \mathbf{x}) \equiv E(\mathbf{x}') - E(\mathbf{x}). \quad (6)$$

Intuitively, when T is high perturbations from \mathbf{x} to \mathbf{x}' which increase the energy are likely to be accepted (in addition to perturbations which decrease the energy, which are always accepted) and the samples can explore the state space. Subsequently, as T is reduced, only perturbations leading to small increases in E are accepted, so that only limited exploration is possible as the system settles on (hopefully) the global minimum. The algorithm is summarized in Algorithm 1: during each of K epochs, the computational temperature is fixed at T_k and L_k samples are drawn from π_{T_k} before the temperature is lowered in the next epoch. Each sample is a perturbation (“mutation” in the nomenclature of evolutionary algorithms) of the current state from a proposal density (line 3); the perturbed state \mathbf{x}' is accepted with probability given by (5), as shown in lines 4–8.

As already alluded to, convergence is guaranteed if and only if the cooling schedule is sufficiently gradual [3], but experience has shown SA to be a very effective optimization technique even with relatively rapid cooling schedules [13], [14].

C. Multiobjective Simulated Annealing (MOSA) With Composite Objective Functions

An attractive approach to multiobjective simulated annealing (MOSA), adopted by several investigators [7]–[10], [15]–[18], is to combine the objectives as a weighted sum

$$E(\mathbf{x}) = \sum_{i=1}^D w_i f_i(\mathbf{x}). \quad (7)$$

The composite objective is then used as the energy to be minimized in a scalar SA optimizer. An equivalent alternative [6] is to sum $\log f_i(\mathbf{x})$, and others (e.g., [16] and [8]) have investigated a number of nonlinear and stochastic composite energies.

It is clear that SA with a composite energy (7) will converge to points on the Pareto optimal front, where the objectives have ratios given by w_i^{-1} , if such points exist. However, it is unclear how to choose the weights in advance, indeed, one of the principal advantages of multiobjective optimization is that the relative importance of the objectives can be decided with the estimated Pareto front on hand. Perhaps more importantly, parts of the front are inaccessible with fixed weights [19]. Recognizing this, investigators have proposed a variety of schemes for adapting the w_i during the annealing process to encourage exploration along the front. See, for example, [20].

It is natural to keep an archive F of all the nondominated solutions found so far, and this archive may be utilized to further exploration by periodically restarting the annealer from a randomly chosen element of F [10].

A proposal \mathbf{x}' in scalar SA is either better or worse than the current state \mathbf{x} depending on the sign of $\delta E(\mathbf{x}', \mathbf{x})$; except for pathological problems the probability that $\delta E = 0$ is vanishingly small. In multiobjective SA, however, \mathbf{x}' may dominate \mathbf{x} or \mathbf{x}' may be dominated by \mathbf{x} or they may be mutually nondominating: in fact, the probability that a pair of randomly chosen points in D -dimensional space are mutually nondominating is $1 - 2(1/2)^D$, so the mutually nondominating case becomes increasingly common with more objectives. However, energies such as (7) may lead to \mathbf{x}' being accepted unconditionally ($\delta E < 0$) even though $\mathbf{x}' \not\prec \mathbf{x}$, because a large negative energy change from one objective may outweigh small positive changes on the other objectives. Each MOSA algorithm which utilizes a composite objective function must, therefore, deal with this behavior in some manner.

A good example of a composite objective function approach to MOSA is given by Suppaitnarm *et al.* [10]. Instead of weighting and summing the objectives to produce a composite energy difference for the acceptance criteria, this algorithm uses a multiplicative function with individual temperatures for each objective each of which is adjusted independently by the algorithm. This negates the need for *a priori* weighting of the objectives, and can be considered to function as a weighted composite sum approach with algorithmically controlled weightings. This is vulnerable to the concentrated

search properties of other composite objective techniques and Suppaitnarm *et al.* employ a return-to-base scheme, whereby the current solution is reseeded with another solution from the nondominated archive to promote a more even coverage. Suppaitnarm *et al.* promote exploration along the front by unconditionally accepting proposals that are not dominated by any member of F , otherwise using (5).

Of the MOSA techniques in the current literature, perhaps the most promising is that of Nam and Park [8] due to their use of dominance in state change probabilities. In this algorithm, the relative dominance of the current and proposed solutions is tested and when the proposed solution dominates the current solution the proposal is accepted; this is analogous to the automatic acceptance of proposals with a lower state energy in single-objective simulated annealing. In addition to the widespread practise of employing a state change probability which guarantees acceptance of strictly superior perturbations, Nam and Park modify the acceptance rule so that proposals are accepted with probability given by (5) and (7) if they are dominated by \mathbf{x} , but unconditionally accepted if $\mathbf{x}' \prec \mathbf{x}$ or if \mathbf{x}' and \mathbf{x} are mutually nondominating. This promotes exploration of the search space and escape from local fronts but as the dimensionality increases so does the proportion of all moves which are accepted unconditionally. This limits the behavior of the algorithm to that of a random walk through the search space when dealing with problems with high dimensionality. When the proposed solution is dominated by the current solution, Nam and Park define the energy difference controlling acceptance as the average difference in objective values. Nam and Park also employ 100 separate agents during optimization, where each agent is an independent copy of the algorithm; this serves a similar function to Suppaitnarm *et al.*'s return-to-base approach to promoting diversity of the solutions located by the algorithm.

Although it is clear that the assurance of a convergence proof can be provided for a multiobjective simulated annealer using a scalar objective function and fixed weights (7), such annealers are fundamentally limited in their coverage of the Pareto front. On the other hand, it is difficult to see how proofs of convergence might be obtained with the heuristic modifications designed to promote exploration transversal to the front. Given these difficulties, defining a multiobjective simulated annealer which utilizes a composite objective function is undesirable. With this in mind, we investigate the efficacy of an energy function based on the defining notion of dominance. The aim is the definition of a single energy function appropriate to all cases of relative dominance between \mathbf{x} and \mathbf{x}' without requiring special cases for where $\mathbf{x}' \prec \mathbf{x}$, or where \mathbf{x}' and \mathbf{x} are mutually nondominating, as has been the case in previous algorithms.

III. A DOMINANCE-BASED ENERGY FUNCTION

In single-objective optimization problems, the sign of the difference in energy $\delta E(\mathbf{x}', \mathbf{x})$ tells us whether the proposal \mathbf{x}' is a better, worse, or (very rarely) an equally good solution as the current solution \mathbf{x} . Likewise, the dominance relation can be used to compare the relative merit of \mathbf{x}' and \mathbf{x} in multiobjective problems, but note that it gives essentially only three values of quality—better, worse, equal—in contrast to the energy difference in unobjective problems which usually gives a continuum.

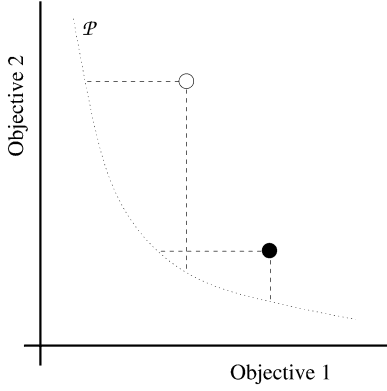


Fig. 1. Energy from area of the true Pareto front \mathcal{P} dominating a solution. Solutions are marked by circles and lines indicate the regions of \mathcal{P} dominating each one.

If the true Pareto front \mathcal{P} were available, we could define a simple energy of \mathbf{x} as the measure of the front that dominates $\mathbf{f}(\mathbf{x})$. Let $\mathcal{P}_{\mathbf{x}}$ be the portion of \mathcal{P} that dominates $\mathbf{f}(\mathbf{x})$

$$\mathcal{P}_{\mathbf{x}} = \{\mathbf{y} \in \mathcal{P} \mid \mathbf{y} \prec \mathbf{f}(\mathbf{x})\}. \quad (8)$$

Then, we define

$$E(\mathbf{x}) = \mu(\mathcal{P}_{\mathbf{x}}) \quad (9)$$

where μ is a measure defined on \mathcal{P} . We shall be principally interested in finite sets approximating \mathcal{P} and so shall take $\mu(\mathcal{P}_{\mathbf{x}})$ to be simply the cardinality of $\mathcal{P}_{\mathbf{x}}$. If \mathcal{P} is a continuous set, we can take μ to be the Lebesgue measure (informally, the length, area, or volume for 2, 3, or 4 objectives); we further discuss measures induced on \mathcal{P} in Section VII-E. As illustrated in Fig. 1, this energy has the properties we desire: if $\mathbf{x} \in \mathcal{P}$, then $E(\mathbf{x}) = 0$, and solutions more distant from the front are, in general, dominated by a greater proportion of \mathcal{P} and so have a higher energy; in Fig. 1 the solution marked by an open circle has a greater energy than the one marked by a filled circle.

Clearly this formulation of an energy does not rely on an *a priori* weighting of the objectives and the assurances of convergence [3] for uniojective SA continue to hold in this case. Since all solutions lying on the front have equal minimum energy, we would anticipate that a simulated annealer using this energy would, having reached the front, perform a random walk exploration of the front.

We note that Fleischer [21] has proposed an alternative measure of a nondominated set, which may be loosely characterized as being based on the volume dominated by the set rather than the area of the dominating set.

Unfortunately, the true Pareto front \mathcal{P} is unavailable during the course of an optimization. We therefore propose to use an energy defined in terms of the current estimate of the Pareto front, F , which is the set of mutually nondominating solutions found thus far in the annealing. Denote by \tilde{F} the union of F , the current solution \mathbf{x} , and the proposed solution \mathbf{x}' , that is

$$\tilde{F} = F \cup \{\mathbf{x}\} \cup \{\mathbf{x}'\}. \quad (10)$$

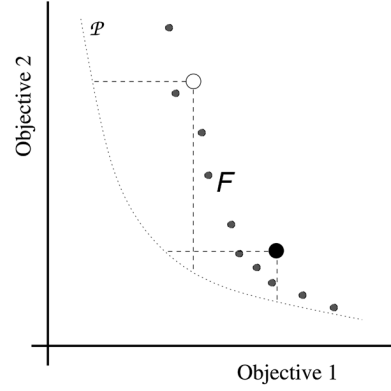


Fig. 2. Energy from proportion of the estimated Pareto front F dominating points dominating a solution. Elements of F are shown as small gray circles, solutions are shown as larger open or filled circles.

Then, in a similar manner to (8), let $\tilde{F}_{\mathbf{x}}$ be the elements of \tilde{F} that dominate \mathbf{x}

$$\tilde{F}_{\mathbf{x}} = \{\mathbf{y} \in \tilde{F} \mid \mathbf{y} \prec \mathbf{x}\}. \quad (11)$$

We note that $|\tilde{F}_{\mathbf{x}}|$ is a quantity similar to one used in the ranking method proposed by Fonseca and Fleming [22], namely, the number of solutions in a search population that dominate \mathbf{x} plus 1. Using $\tilde{F}_{\mathbf{x}}$, we obtain an energy difference between the current and proposed solutions of

$$\delta E(\mathbf{x}', \mathbf{x}) = \frac{1}{|\tilde{F}|} (|\tilde{F}_{\mathbf{x}'}| - |\tilde{F}_{\mathbf{x}}|). \quad (12)$$

Division by $|\tilde{F}|$ ensures that δE is always less than unity and provides some robustness against fluctuations in the number of solutions in F . If \tilde{F} is a nondominating set, the energy difference between any two of its elements is zero. Note also that $\delta E(\mathbf{x}', \mathbf{x}) = -\delta E(\mathbf{x}, \mathbf{x}')$. The inclusion of the current solution and the proposal in \tilde{F} means that $\delta E(\mathbf{x}', \mathbf{x}) < 0$ if $\mathbf{x}' \prec \mathbf{x}$, which ensures that proposals that move the estimated front towards the true front are always accepted. Proposals that are dominated by one or more members of the current archive are accepted with a probability depending upon the difference in the number of solutions in the archive that dominate \mathbf{x}' and \mathbf{x} . We emphasize that this probability does not depend upon metric information in objective space; we put no *a priori* weighting on the objectives and the acceptance probability is unaffected by rescalings of the objectives.

A further benefit of this energy measure is that it encourages exploration of sparsely populated regions of the front. Imagine two proposals, each dominated by some solutions in F ; for example, the solutions illustrated by the filled and unfilled circles in Fig. 2. The solution that is dominated by fewer elements (the unfilled circle) has the lower energy, and would therefore be more likely to be accepted as a proposal.

Defining the energy in this manner, unlike some proposed multiobjective enhancements to SA discussed in Section II-C, provides a single energy function encouraging both convergence to and coverage of the Pareto front without requiring other modifications to the single-objective simulated annealing algorithm (beyond the obvious storage of an archive of the

estimated Pareto front). In particular, no additional rules are required for cases in which the current and proposed solutions are mutually nondominating.

Convergence to the true Pareto front is no longer an immediate consequence of Geman and Geman's work [3], because the energy based on F is only an approximation to (9). However, Greening [23] offers proof of convergence, albeit more slowly, even when the energy contains errors. Current work is investigating the application of this work to MOSA, and in Section VII, we offer empirical evidence of the convergence.

An energy function based on (12) is straightforward to calculate; counting the number of elements of \hat{F} that dominate \mathbf{x} and \mathbf{x}' can be achieved in logarithmic time [24], [25]. Our proposed multiobjective algorithm closely follows the standard SA algorithm (Algorithm 1), the only addition that is necessary is to maintain an archive, F of the current estimate of the Pareto front, and to calculate the energy difference using (12). However, we postpone detailed description of the algorithm until methods of increasing the empirical energy resolution have been discussed.

IV. INCREASING ENERGY RESOLUTION

As mentioned earlier, the true Pareto-optimal front of solutions is, in general, unavailable to us. While using the archive of the estimated Pareto front F provides an estimate of solution energy, when F is small, the resolution in the energies can be very coarse. In fact, the difference in energy between two solutions is an integer multiple of $1/|\hat{F}|$ between 0 and 1. Since the acceptance criterion (5) for new solutions is determined by the difference in energy $\delta E(\mathbf{x}, \mathbf{x}')$ between the current solution and the proposed solution, low resolution of the energies leads to a low resolution in acceptance probabilities. At low computational temperatures and with small archives, it will become increasingly likely that this granularity will make it almost impossible for even slightly detrimental moves (i.e., moves that increase $E(\mathbf{x})$) to be made. This is undesirable as, at its most severe, this effect reduces the algorithm to behavior similar to a greedy search optimizer, and prevents the exploratory behavior provided by detrimental moves.

For this reason, and because a limited archive may inhibit convergence [24], [26], we do not constrain the size of the archive. In fact, in order to increase the energy resolution, we examine methods for using a larger set for energy calculations. There are a couple of straightforward, but ultimately inadequate, methods for artificially increasing the size of F , which we now briefly discuss before describing a method using the attainment surface.

A. Conditional Removal of Dominated Points

A straightforward method for increasing the size of the archive is not to delete solutions known to be dominated if deleting them would reduce $|F|$ below some predefined minimum. However, the existence of old solutions in F , may lead to desirable proposals (i.e., not nondominated solutions) being rejected. In addition, the old solutions may bias the search away from regions of the front that were previously well populated.

A further disadvantage of this method is that the retained solutions may be positioned so that they are dominated by the

archive and indeed by the current point and the vast majority of proposals. In this case, they serve to increase the resolution of the energy at the expense of the range. By contrast, the interpolation method using the attainment surface that we propose below insists that interpolating points are only weakly dominated by the archive.

B. Linear Interpolation

Another apparently suitable method of augmenting F is linear interpolation (in objective space) between the solutions in F . In this method, when the archive is smaller than some predefined size, new points in objective space are generated on the simplices defined by an element of F and its $D - 1$ nearest neighbors in F . This overcomes the limitations of the previous method: Since new solutions are generated "on" the current estimated Pareto front, the problems which could occur with using old, dominated elements of F in the energy calculations are avoided. The interpolated points generated can also be evenly distributed between the current estimated Pareto-optimal solutions, which is beneficial as it does not deter the algorithm from exploring any region of the estimated front which is not already densely populated. The principal disadvantage of this method is that proposals may be dominated by an interpolated point, but not by any of the real elements of F , meaning that the proposal may erroneously be disregarded.

C. Attainment Surface Sampling

Consideration of the previous two methods of augmenting the estimated Pareto front suggests that the augmenting points should have the following properties.

- 1) The augmenting points must be sufficiently close to the current estimation of the Pareto front that they can affect the energy of solutions generated near to the current estimated Pareto front.
- 2) They must be evenly distributed across the currently estimated Pareto front so as to not discourage the algorithm from accepting proposals in poorly populated regions of the front.
- 3) They must not dominate any proposal which is not dominated by any member of F , so that potential entrants to the archive are not discarded. A consequence of this is that they must all be dominated by at least one member of F .

The *attainment surface*, which has previously been used for estimated Pareto front visualization [27] and is closely related to the attainment function [28], is an interpolating surface between the elements of F that has the requisite properties. The attainment surface S_F corresponding to F is a conservative interpolation of the elements of F so that every point of S_F is dominated by an element of F . The attainment surface for an F comprising four two-dimensional elements is sketched in Fig. 3. More formally, the attainment surface is the boundary of the region in objective space which is dominated by elements of F . If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^D$, we say that \mathbf{u} *properly dominates* \mathbf{v} (denoted $\mathbf{u} \prec \mathbf{v}$) iff $u_i < v_i \forall i = 1, \dots, D$. Then, if

$$\mathcal{F} = \{\mathbf{y} | \mathbf{u} \prec \mathbf{y} \text{ for some } \mathbf{u} \in F\} \quad (13)$$

$$\mathcal{U} = \{\mathbf{y} | \mathbf{u} \prec \mathbf{y} \text{ for some } \mathbf{u} \in F\} \quad (14)$$

the attainment surface is $S_F = \mathcal{F} \setminus \mathcal{U} = \partial \mathcal{U}$.

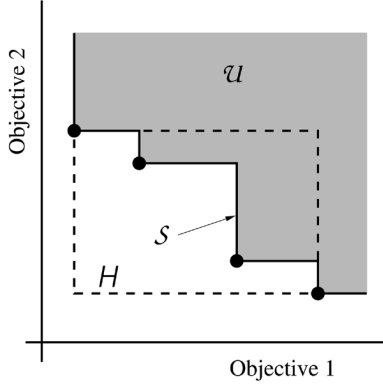


Fig. 3. Attainment surface \mathcal{S}_F is the boundary of the region, \mathcal{U} dominated by the nondominated set F , whose elements are marked as dots. Dashed lines denote H the minimum rectangle containing F .

Algorithm 2: Sampling a point from the attainment surface

Inputs:

$\{L_d\}_{d=1}^D$ *Elements of F , sorted by increasing coordinate d*

```

1:  for  $i := 1, \dots, D$     Generate a random point,  $\mathbf{v}$ 
2:     $v_i := \text{rand}(\min(L_i), \max(L_i))$ 
3:  end
4:   $d := \text{randint}(1, D)$     Choose a dimension,  $d$ 
5:  for  $i := 1, \dots, |F|$     Find smallest  $v_d$  s.t.  $\mathbf{v}$  is
6:     $\mathbf{u} := L_{d,i}$     dominated by an element of  $F$ 
7:     $v_d := u_d$ 
8:    if  $F \prec \mathbf{v}$ 
9:      return  $\mathbf{v}$ 
10:   end
11: end
```

Let H_F be the minimum axis-parallel hyper-rectangle containing F ; that is, the hyper-rectangle defined by

$$H_F = [\min_{\mathbf{y} \in F}(y_1), \max_{\mathbf{y} \in F}(y_1)] \times \dots \times [\min_{\mathbf{y} \in F}(y_D), \max_{\mathbf{y} \in F}(y_D)]. \quad (15)$$

Then, as illustrated in Fig. 3, we interpolate F with random samples uniformly distributed on $\mathcal{S}_F \cap H_F$, the attainment surface restricted to H_F . From the definition of \mathcal{S}_F , it is apparent that interpolated points are dominated by an element of F , thus satisfying the third criterion. Uniform random sampling ensures that the second criterion is met, as is the first criterion because \mathcal{S}_F interpolates F .

Sampling from \mathcal{S}_F may be performed using Algorithm 2, which works by sampling a point from a uniform distribution on the surface of H_F , and then restricting one coordinate so that the point is dominated by an element of F . This is facilitated by the use of lists $L_d, d = 1, \dots, D$ which comprise the elements of F sorted in increasing order of coordinate d . Determining whether an element of F dominates \mathbf{v} on line 8 may be efficiently implemented using binary searches of the lists L_d , in which case the algorithm requires $\mathcal{O}(|F| \log(|F|))$ time for the

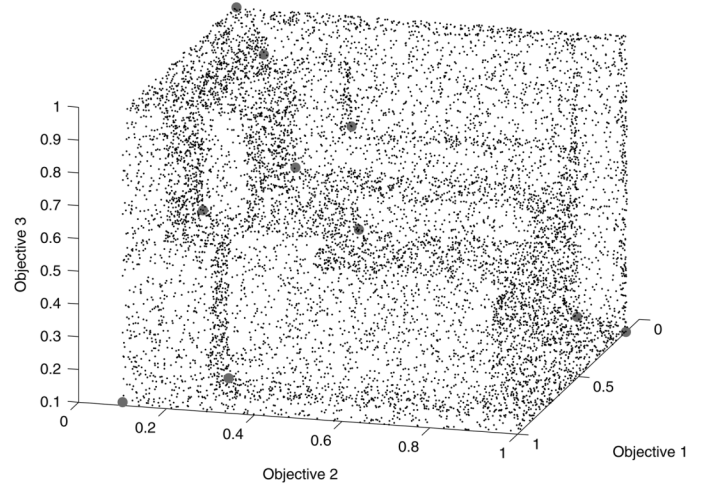


Fig. 4. Ten thousand samples from the attainment surface for an archive of ten points, which are marked with heavy dots.

generation of each sample. Fig. 4 illustrates the sampled attainment surface for a set of ten three-dimensional points; 10 000 samples are shown for visualization. In the experiments reported in Section VII, F was augmented with 100 samples from \mathcal{S}_F before calculating the energy of the proposal. With more objectives, the energy resolution can be beneficially increased by sampling more interpolating points. It is important to note that the purpose of attainment surface sampling is to uniformly increase the resolution of the energy function across F and that, if performed extensively, this will partially negate the benefit of the energy function guiding search towards lesser-populated regions of F . For this reason, the number of sampled points should not be too high, and it is advisable to only sample when it is necessary to increase the resolution. The results presented here, where 100 samples from \mathcal{S}_F are always taken, demonstrate that sampling when not strictly necessary does not prevent convergence.

V. MULTIOBJECTIVE SIMULATED ANNEALING (MOSA) ALGORITHM

Having discussed sampling from the attainment surface to increase the energy resolution, we are now in a position to summarize the main points of our proposed MOSA algorithm. As shown in Algorithm 3, the multiobjective algorithm differs from the uniojective algorithm in that an archive F of nondominated solutions found so far is maintained, and the energy difference between the proposed and current solution is calculated based on the current archive or its attainment surface.

The archive is initialized with the initial feasible point (line 1 of Algorithm 3). At each stage, the current solution \mathbf{x} is perturbed to form the proposed solution \mathbf{x}' . In the work reported here, in which the parameters \mathbf{x} are continuous and real valued, we perturb each element of \mathbf{x} singly, drawing the perturbations from a Laplacian distribution centred on the current value.

If there are sufficiently many solutions in F , the augmented archive \hat{F} is constructed by adding \mathbf{x} and \mathbf{x}' (line 9) to F and the energy difference between \mathbf{x}' and \mathbf{x} is calculated using (12). If there are fewer than S solutions, then additional samples are drawn from the attainment surface \mathcal{S}_F using Algorithm 2 (line

6); the energy difference is then calculated based on the sampled attainment surface \mathbf{x} and \mathbf{x}' . In the work reported here, we always augment F with 100 samples from \mathcal{S}_F . As even when there are a large number of solutions in the archive of the estimated Pareto front it is worthwhile sampling from \mathcal{S}_F as this samples evenly across the front, providing greater resolution in sparsely populated areas of the front.

If the proposal is accepted (line 14), the archive must be updated. If \mathbf{x} is not dominated by any of the archival solutions, all archival solutions that are dominated by \mathbf{x} are deleted from the archive (line 16) and \mathbf{x} is added to the archive (line 17). Clearly F is always a nondominated set, although note that \mathbf{x}' may be dominated by members of F .

Algorithm 3: Multiobjective simulated annealing

Inputs:

$\{L_k\}_{k=1}^K$ *Sequence of epoch durations*
 $\{T_k\}_{k=1}^K$ *Sequence of temperatures, $T_{k+1} < T_k$*
 \mathbf{x} *Initial feasible solution*

```

1:   $F := \{\mathbf{x}\}$     Initialize archive
2:  for  $k := 1, \dots, K$ 
3:    for  $i := 1, \dots, L_k$ 
4:       $\mathbf{x}' := \text{perturb}(\mathbf{x})$ 
5:      if  $|F| < S$     If  $F$  is small
6:         $\mathcal{S}_F := \text{interpolate}(F)$ 
7:         $\tilde{F} := \mathcal{S}_F \cup F \cup \{\mathbf{x}\} \cup \{\mathbf{x}'\}$ 
8:      else
9:         $\tilde{F} := F \cup \{\mathbf{x}\} \cup \{\mathbf{x}'\}$ 
10:    end
11:     $\delta E(\mathbf{x}', \mathbf{x}) := E(\mathbf{x}') - E(\mathbf{x})$ 
12:     $u := \text{rand}(0, 1)$ 
13:    if  $u < \min(1, \exp(-\delta E(\mathbf{x}', \mathbf{x})/T_k))$ 
14:       $\mathbf{x} := \mathbf{x}'$     Accept new current point
15:      if  $\mathbf{x}$  is not dominated by any
16:        element of  $F$ 
17:         $F := F \cup \mathbf{x}$     Add  $\mathbf{x}$  to  $F$ 
18:    end
19:  end
20: end
21: end
```

VI. REAL-TIME ALGORITHM PARAMETER OPTIMIZATION

The performance of this algorithm, in common with other SA systems, depends upon parameters for the initial temperature, the annealing schedule and the size of perturbations made to so-

lutions when generating new proposals. Here, we give details of methods which permit automatic setting of the initial temperature, and which adjust the scale of perturbations made to maximize the quality of proposed solutions.

A. Annealing Schedule

If the initial computational temperature is set too high, all proposed solutions will be accepted, irrespective of their relative energies, and if set too low proposals with a higher energy than the current solution will never be accepted, transforming the algorithm into a greedy search. As a reasonable starting point, we set the initial temperature to achieve an initial acceptance rate of approximately 50% on derogatory proposals. This initial temperature T_0 can be easily calculated by using a short “burn-in” period during which all solutions are accepted and setting the temperature equal to the average positive change of energy divided by $\ln(2)$.

In the work reported here, all epochs L_k are of equal length, $L_k = 100$, and we adjust the temperature according to $T_k = \beta^k T_0$, where β is chosen so that T_k is 10^{-5} after two thirds of the evaluations are completed.

B. Perturbation Scalings

For simplicity, a proposal is generated from \mathbf{x} by perturbing only one parameter or decision variable of \mathbf{x} . The parameter to be perturbed is chosen at random and perturbed with a random variable ϵ drawn from a Laplacian distribution $p(\epsilon) \propto e^{-|\sigma\epsilon|}$, where the scaling factor σ sets magnitude of the perturbation. The Laplacian distribution has tails that decay relatively slowly, thus ensuring that there is a high probability of exploring regions distant from the current solutions.

We maintain two sets of scaling factors, since the perturbations generating moves to a nondominated proposal within a front (we call these *traversals*) may potentially be very different from those required to locate a front closer to \mathcal{P} , which we call *location* moves. We maintain a scaling factor for each dimension of parameter space for each of the location perturbations and the traversal perturbations, and adjust these independently to increase the probability of such a move being generated. When perturbing a solution, it is chosen randomly with equal probability whether the location scaling set will be used, or the traversal scaling set. Statistics are kept on perturbations generating traversal and location moves; clearly these can be updated only after the proposal has been generated so that the type of move is known. The scalings are adjusted throughout the optimization, whenever a suitably large statistic set is available to reliably calculate an appropriate scaling factor. These scalings are initially set large enough to sample from the entire feasible space.

Traversal Scaling: The traversal rescaling for a particular decision variable x_j is performed whenever approximately 50 traversal perturbations have been made to x_j since the last rescaling.

In order to ensure wide coverage of the front, we wish to maximize the distance (in objective space) covered by the traversals to ensure the entire front is evenly covered. Generating traversals traveling a small distance will concentrate the estimated front around the point at which the current front was discovered, an effect we aim to avoid.

We seek to generate proposals on approximately the scale that has previously been successful in generating wide-ranging traversals. To achieve this, the perturbations are sorted by absolute size of perturbation in parameter space, and then trisected in order, giving three groups, one of the smallest third of perturbations, the largest third of perturbations, and the remaining perturbations. For each group, the mean traversal size caused by the perturbations is calculated. The traversal size is measured as the Euclidean distance traveled in objective space when the current solution and the proposed solution are mutually non-dominating. If a perturbation and the current solution are not mutually nondominating, the traversal size is counted as being 0. The traversal perturbation scaling for this decision variable is then set to the average perturbation of the group which generated the largest average traversal.

This heuristic is open to the criticism that it depends upon measuring distances in objective space while the relative weighting of the D objective functions is unknown. To alleviate this difficulty, however, the objectives may be renormalized during optimization so the front has approximately the same extent in each objective. We emphasize that, of course, the use of metric information for setting the approximate scale of perturbations does not affect the dominance-based energy.

Location Scaling: Drawing from methods widely used in evolutionary algorithms (see [30]–[32] for recent work in this area), we aim to adjust the scale of location perturbations to keep the acceptance rate for \mathbf{x}' that have a higher energy than \mathbf{x} to approximately 1/3, so that exploratory proposals are made and accepted at all temperatures.

The location perturbation scaling is recalculated for each parameter for which 20 proposals having $\delta E(\mathbf{x}', \mathbf{x}) > 0$ have been generated, after which the count is reset. Location perturbation rescaling is omitted in two cases: First, when the archive of the estimated Pareto front F has fewer than ten members. Second, when the combined size of F augmented by the samples from the attainment surface when multiplied by the temperature does not exceed 1. This is because we adjust the scalings to attempt to keep the acceptance rate of derogatory moves approximately a third; when this value is too small, it becomes impossible to generate such a scaling, and so the scalings are kept at the most recent valid value.

Counting only moves generated from perturbations to a particular dimension of parameter space, the acceptance rate of derogatory moves α is the fraction of proposals to a greater energy which are accepted. If σ denotes the location perturbation scaling for a particular dimension, the new σ is set as

$$\sigma := \begin{cases} \sigma(1 + 2(\alpha - 0.4)/0.6), & \text{if } \alpha > 0.4 \\ \sigma, & \text{if } 0.3 \leq \alpha \leq 0.4 \\ \sigma/(1 + 2(0.3 - \alpha)/0.3), & \text{if } \alpha < 0.3 \end{cases} \quad (16)$$

This update works because, in general, smaller perturbations in parameter space are more likely to generate small changes in objective space, resulting in smaller changes in energy.

VII. EXPERIMENTS

We illustrate the performance of this annealer on some well-known test functions from the literature, namely, the DTLZ test

TABLE I
TEST PROBLEM DEFINITION OF DTLZ1–DTLZ7 OF [29] FOR THREE OBJECTIVES (USING THE SUGGESTED PARAMETER SIZES). (DEFINITION OF DTLZ5 CORRECTED.)

| Problem | Definition |
|---------|---|
| DTLZ1 | $f_1(\mathbf{x}) = \frac{1}{2}x_1x_2(1 + g(\mathbf{x}))$ $f_2(\mathbf{x}) = \frac{1}{2}x_1(1 - x_2)(1 + g(\mathbf{x}))$ $f_3(\mathbf{x}) = \frac{1}{2}(1 - x_1)(1 + g(\mathbf{x}))$ $g(\mathbf{x}) = 100(\mathbf{x} - 2) + \sum_{i=3}^P (x_i - 0.5)^2 - \cos(20\pi(x_i - 0.5))$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 7$ |
| DTLZ2 | $f_1(\mathbf{x}) = \cos(x_1\pi/2) \cos(x_2\pi/2)(1 + g(\mathbf{x}))$ $f_2(\mathbf{x}) = \cos(x_1\pi/2) \sin(x_2\pi/2)(1 + g(\mathbf{x}))$ $f_3(\mathbf{x}) = \sin(x_1\pi/2)(1 + g(\mathbf{x}))$ $g(\mathbf{x}) = \sum_{i=3}^P (x_i - 0.5)^2$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 12$ |
| DTLZ3 | $f_1(\mathbf{x}) = \cos(x_1\pi/2) \cos(x_2\pi/2)(1 + g(\mathbf{x}))$ $f_2(\mathbf{x}) = \cos(x_1\pi/2) \sin(x_2\pi/2)(1 + g(\mathbf{x}))$ $f_3(\mathbf{x}) = \sin(x_1\pi/2)(1 + g(\mathbf{x}))$ $g(\mathbf{x}) = 100(\mathbf{x} - 2) + \sum_{i=3}^P (x_i - 0.5)^2 - \cos(20\pi(x_i - 0.5))$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 12$ |
| DTLZ4 | $f_1(\mathbf{x}) = \cos(x_1^\alpha\pi/2) \cos(x_2^\alpha\pi/2)(1 + g(\mathbf{x}))$ $f_2(\mathbf{x}) = \cos(x_1^\alpha\pi/2) \sin(x_2^\alpha\pi/2)(1 + g(\mathbf{x}))$ $f_3(\mathbf{x}) = \sin(x_1^\alpha\pi/2)(1 + g(\mathbf{x}))$ $g(\mathbf{x}) = \sum_{i=3}^P (x_i - 0.5)^2$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 12$ |
| DTLZ5 | $f_1(\mathbf{x}) = \cos(\theta_1\pi/2) \cos(\theta_2)(1 + g(\mathbf{x}))$ $f_2(\mathbf{x}) = \cos(\theta_1\pi/2) \sin(\theta_2)(1 + g(\mathbf{x}))$ $f_3(\mathbf{x}) = \sin(\theta_1\pi/2)(1 + g(\mathbf{x}))$ $g(\mathbf{x}) = \sum_{i=3}^P (x_i - 0.5)^2$ $\theta_1 = x_1$ $\theta_2 = \frac{\pi}{4(1+g(\mathbf{x}))} (1 + 2g(\mathbf{x})x_2)$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 12$ |
| DTLZ6 | $f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = x_2$ $f_3(\mathbf{x}) = (1 + g(\mathbf{x}))h(f_1, f_2, f_3, g)$ $g(\mathbf{x}) = \frac{9}{P-2} \sum_{i=3}^P x_i$ $h(f_1, f_2, f_3, g) = 3 - \sum_{i=1}^3 \left(\frac{f_i}{1+g} (1 + \sin(3\pi f_i)) \right)$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 22$ |
| DTLZ7 | $f_1(\mathbf{x}) = \frac{1}{10} \sum_{i=1}^{10} x_i$ $f_2(\mathbf{x}) = \frac{1}{10} \sum_{i=11}^{20} x_i$ $f_3(\mathbf{x}) = \frac{1}{10} \sum_{i=21}^{30} x_i$ s.t. $g_1(\mathbf{x}) = f_3(\mathbf{x}) + 4f_1(\mathbf{x}) - 1 \geq 0$ s.t. $g_2(\mathbf{x}) = f_3(\mathbf{x}) + 4f_2(\mathbf{x}) - 1 \geq 0$ s.t. $g_3(\mathbf{x}) = 2f_3(\mathbf{x}) + f_1(\mathbf{x}) + f_2(\mathbf{x}) - 1 \geq 0$ $0 \leq x_i \leq 1$, for $i = 1, 2, \dots, P$, $P = 30$ |

functions of Deb *et al.* [29], [33], and compare them to the performance of the well established NSGA-II evolutionary algorithm [11] (using the PISA reference implementation [34]) and Nam and Park's multiobjective simulated annealer [8] which we discuss in Section I. The benefit of using the DTLZ test functions is that the true Pareto front, \mathcal{P} , is known, so we can discover how close our estimated archive F is to \mathcal{P} , as well as compare results from each algorithm. Note that we rectify a couple of minor typographical errors in the description of DTLZ5 and DTLZ6 here, as the formulae published in [29] and [33] do not yield the Pareto fronts described.¹ For completeness, we give the problem definitions in Table I; in all the experiments, we use $D = 3$ objectives.

¹In equation (25) of [29], only θ_1 should be multiplied by $\pi/2$ when calculating f_1, \dots, f_M . In (27), the calculation of $g(\mathbf{x}_M)$ is inconsistent with the results provided, meaning all f_3 values in the figure in [33] are halved.

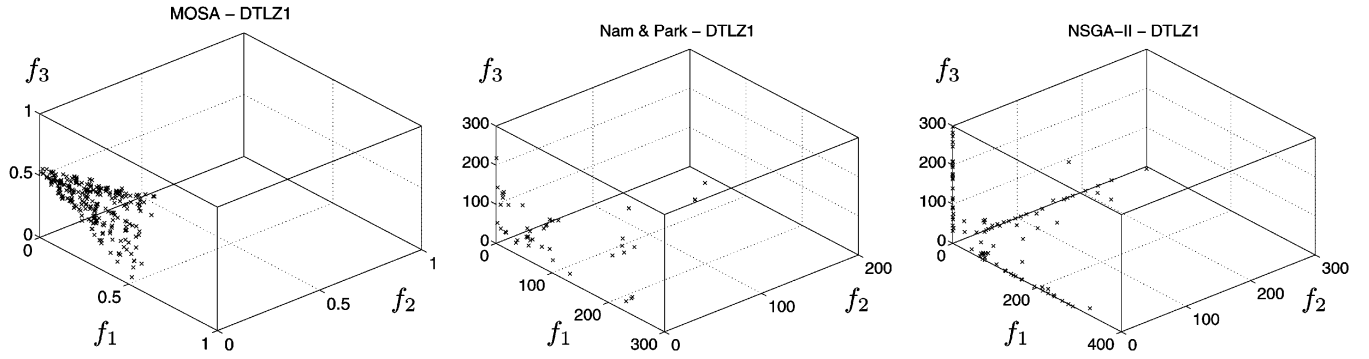


Fig. 5. Archives on test problem DTLZ1 after 5000 function evaluations for each of the three algorithms.

In the work reported here, all epochs L_k are of equal length, $L_k = 100$, and we adjust the temperature according to $T_k = \beta^k T_0$, where β is chosen so that T_k is 10^{-5} after approximately two thirds of the evaluations are completed; run lengths and the exact number of evaluations before T_k is 10^{-5} are given in Table II. For MOSA, the parameter perturbations are controlled using the scheme described in Section VI-B. The perturbations for Nam and Park's annealer are performed using a scheme similar to that for MOSA but without the automatic rescaling feature novel to MOSA; the scalings are fixed at 0.1 (determined from a small empirical study, although the results are only mildly dependent on the scaling). The parameters for the NSGA-II algorithm used were those suggested as the default values in the PISA [34] package.² We use 100 simultaneous chains for the Nam and Park implementation and a population of size 100 for NSGA-II.

We first discuss the performance of the algorithms on each of the DTLZ test problems, after which we present statistical results summarizing the performance over 20 runs. We use the nonparametric Mann–Whitney rank-sum test (at the 0.05 level) to test for significant differences between the algorithms in the hypervolume and true front distance comparison measures. Files containing the final archives located by MOSA for each of these problems are available online at <http://www.secam.ex.ac.uk/people/staff/reversion/research/mosa.html>.

A. DTLZ 1

Fig. 5 shows views in objective space of the archive obtained from a single run of each of the algorithms on test problems DTLZ1 after 5000 objective evaluations. For each algorithm, the plotted results are those which have the median distance of solutions to the true front out of a series of 20 runs; this ensures that the results presented are representative of the series. The true front for DTLZ1 is the segment of the plane passing through 0.5 on each of the objective space coordinate axes, and it can be seen that the majority of solutions generated by MOSA lie very close to the front. This test problem has a large number ($\approx 11^5$) of local fronts which lie as planes parallel to and further from the origin than \mathcal{P} ; the existence of these fronts is evident from the histogram (not shown) of the distances from \mathcal{P} which

²The values for the PISA variator parameters are: individual_mutation_probability=1, individual_recombination_probability=1, variable_mutation_probability=1, variable_swap_probability=0.5, variable_recombination_probability=1, eta_mutation=15, eta_recombination=5.

TABLE II
ANNEALING SCHEDULES

| Problem | Run length | Time to $T_k = 10^{-5}$ |
|---------|------------|-------------------------|
| DTLZ1 | 5000 | 3000 |
| DTLZ2 | 1000 | 500 |
| DTLZ3 | 15000 | 10000 |
| DTLZ4 | 5000 | 3000 |
| DTLZ5 | 1000 | 500 |
| DTLZ6 | 5000 | 3000 |
| DTLZ7 | 9000 | 6000 |

shows solutions clustered at two distinct distances for MOSA and several for NSGA-II (this effect is less marked on the Nam and Park front, where the solutions are distributed more evenly across many fronts which are close in objective space). It seems likely that it is these local fronts which prevent Nam and Park's annealer and NSGA-II from converging on the true front, since in later problems without this feature the difference in performance between the three algorithms is, while still significant, much less extreme. Fig. 14 provides, for each test problem, box plots comparing the average distance of the archive to the true front, the volume measure of the archive and the number of solutions in the archive (which is a fixed value for NSGA-II due to the constrained nature of the algorithm). For this DTLZ1 problem, the figure clearly illustrates that MOSA has not only converged to a set very close to the true front but that the front is also well covered as shown by the volume measure results; the number of solutions in the MOSA archive is unconstrained, so the algorithm has been able to generate a large archive close to, and with good coverage of, the true front. We observe that the annealer on this problem converges to a local front, spreads across it until a perturbation “breaks through” to a front closer to \mathcal{P} after which the annealer explores the nearer local front, adding solutions on this front to the archive and removing solutions on the previous local front as they are dominated during the exploration. Fig. 6 shows the median, maximum, and minimum (over 20 runs) of the distance of the current point \mathbf{x} to the true front \mathcal{P} versus iteration, together with the median (over 20 runs) of the median distance of members of the archive F from \mathcal{P} on a much longer set of runs. The presence of local fronts is apparent from the “steps” in the median archive distance. The current solution clearly leads the archive, particularly at later iterations when the computational temperature is low and the search is effectively a greedy search.

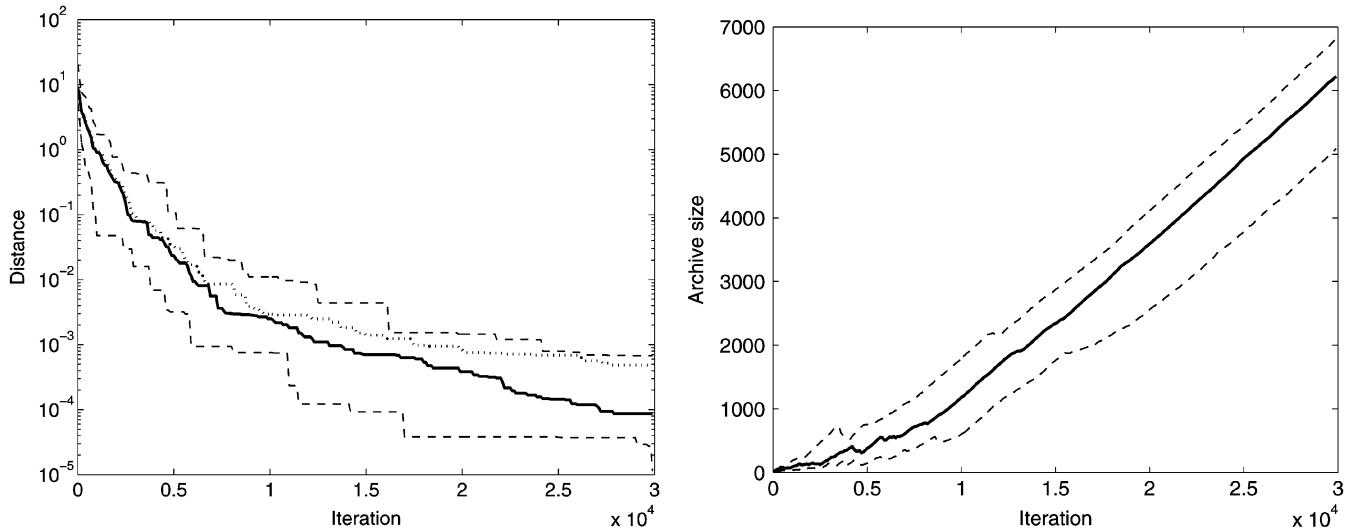


Fig. 6. Left: Distance of current point, \mathbf{x} , and archive F from the true Pareto front, \mathcal{P} , versus iteration for DTLZ1. The dotted line shows median over 20 runs of distance of \mathbf{x} from \mathcal{P} ; dashed lines show maximum and minimum (over the 20 runs) distances at each iteration. The thick line shows the median (over 20 runs) of the median distance of archive members to \mathcal{P} . Right: Archive growth versus iteration. Thick line shows median (over 20 runs) archive size and dashed lines show maximum and minimum.

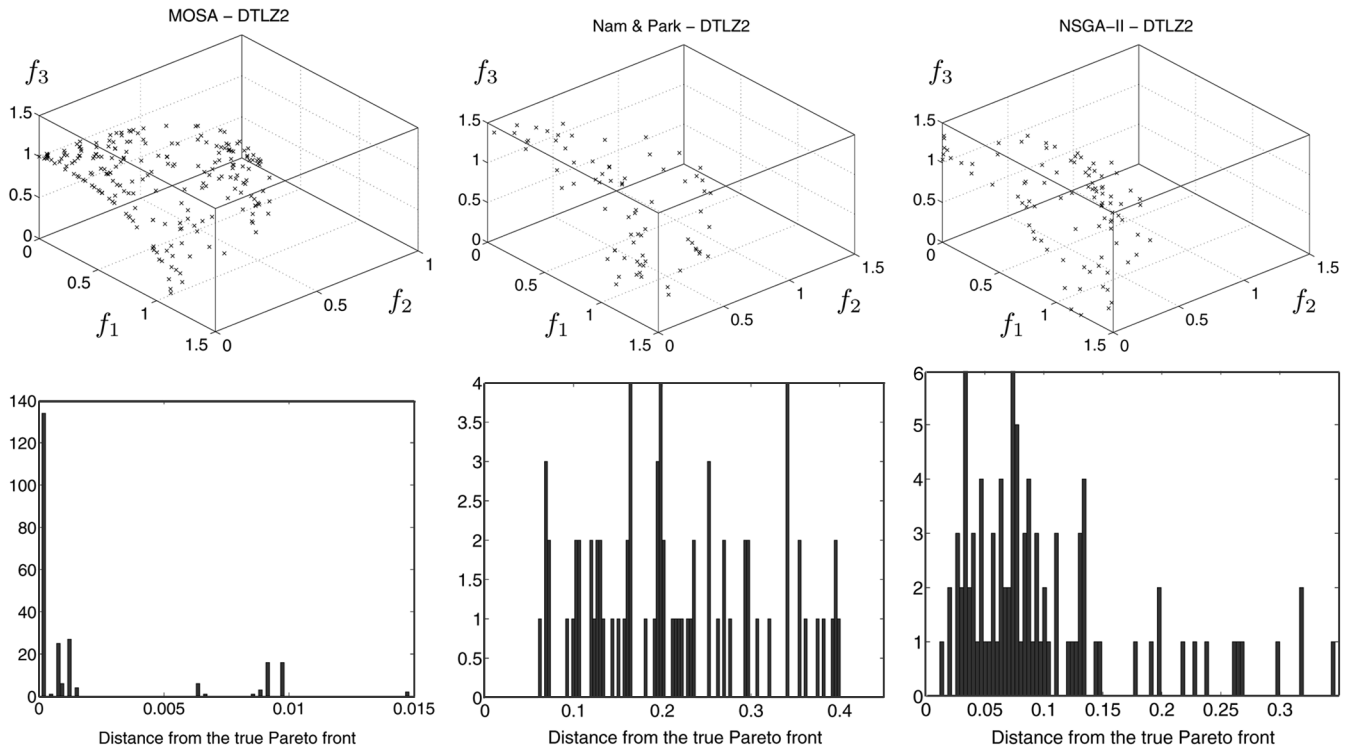


Fig. 7. Top: Archives on test problem DTLZ2 after 1000 function evaluations. Bottom: Histograms of archive member distances from the true Pareto front (the 5% most distant have been omitted to aid visualization).

B. DTLZ 2

Fig. 7 presents the archive resulting from a representative run of the algorithms on problem DTLZ2 for 1000 function evaluations and a plot of the distances from the true front, which is the eighth of a spherical shell of radius 1, centered on the origin, lying in the positive octant. As the figure shows, the archive lies close to the optimal front for each of the algorithms, with MOSA significantly closer than the other algorithms.

We remark that this problem, and several others of the DTLZ suite without a plethora of local fronts, can be successfully

treated with a rapid cooling schedule, as used here. Due to the ease of convergence to the true front on this problem, we anticipate that any multiobjective optimizer will be able to produce a set of solutions close to the true front although the density and coverage may vary significantly, as is the case here. Fig. 14 illustrates that, while all three algorithms have converged close to the true front, MOSA is significantly closer than NSGA-II or Nam and Park's annealer. The volume measure plot shows that the archive produced by MOSA also has a greater coverage/density of solutions; even after only 1000 evaluations, the

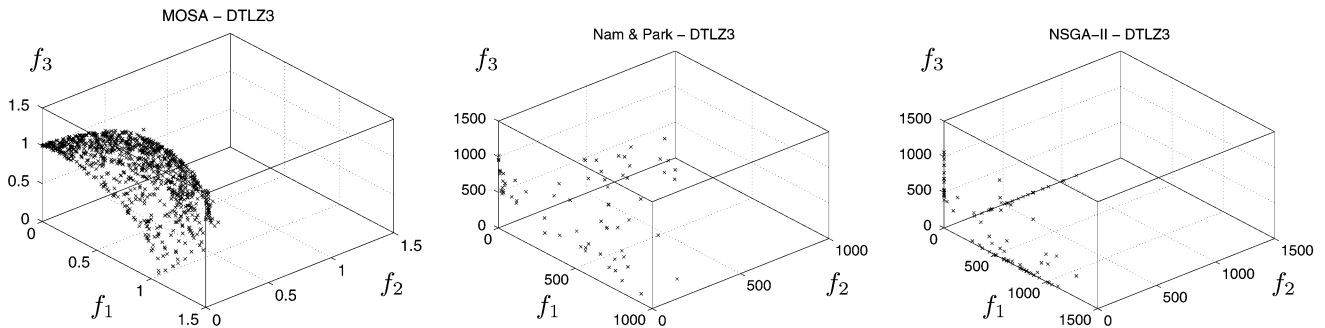


Fig. 8. Archives on test problem DTLZ3 after 15 000 function evaluations.

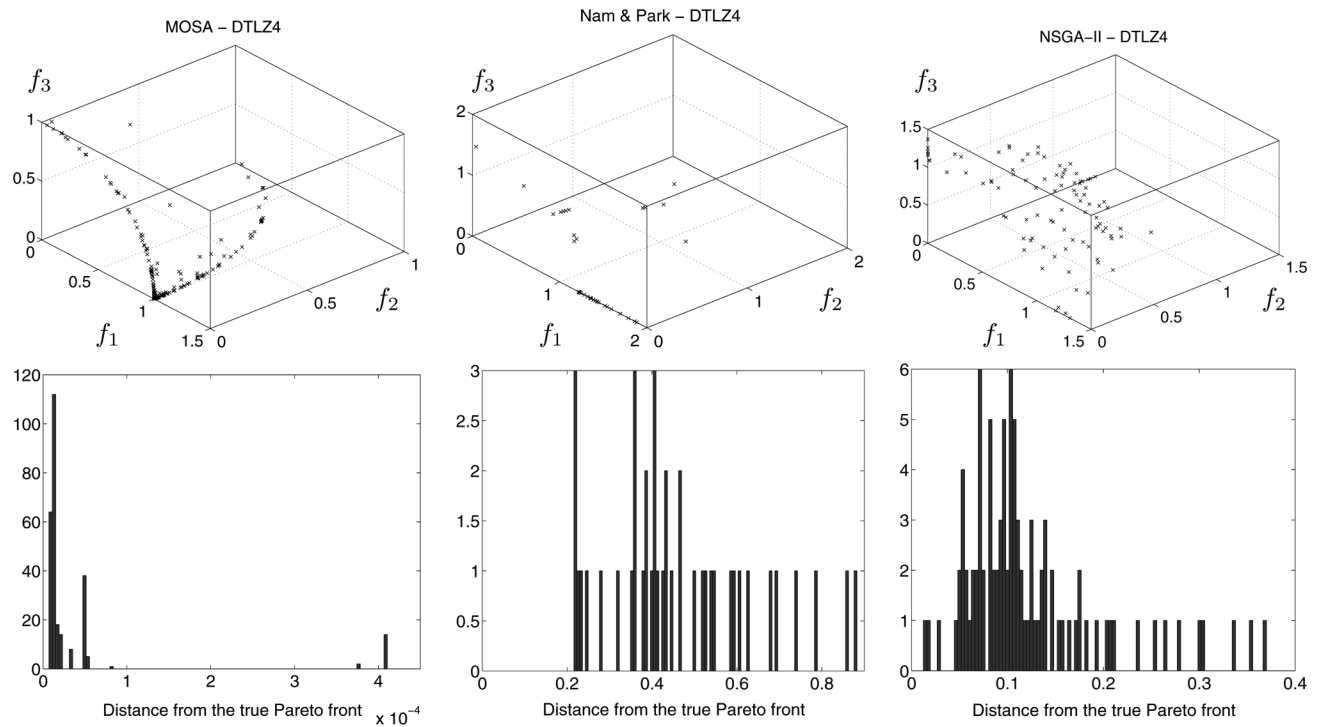


Fig. 9. Top: Archives on test problem DTLZ4 after 5000 function evaluations. Bottom: Histograms of the distance from the true Pareto front of the archive members (the 5% most distant have been omitted to aid visualization).

archive size plot clearly illustrates that MOSA has already converged very close to the true front and is searching across the front improving the coverage and density.

While knowledge about the applicability of a short annealing schedule would not be initially available for typical real-world problems, we anticipate that, for real-world problems, the annealer would be run with a very rapid annealing schedule initially to discover if the problem were searchable in this manner.

C. DTLZ 3

A striking example of the annealer's performance is provided in Fig. 8, where its evaluation on DTLZ3 is shown for 15 000 function evaluations. The Pareto front here is again an eighth of a spherical shell, preceded by multiple local fronts of the same order as DTLZ1. The computational archive is converged to within 0.01 of the true front. Consistent with the findings by Deb *et al.* [33], NSGA-II had failed to converge (Deb *et al.* comment that in their experiments that NSGA-II had still failed to

converge after 50 000 function evaluations) and Nam and Park's annealer yields performance similar to NSGA-II (as illustrated in Fig. 14). Consistent with the previous problems, MOSA's archive is shown to be large, dense, and well covering in Fig. 14.

D. DTLZ 4

The true Pareto front for this problem is again an eighth of a spherical shell, but the solutions are unevenly distributed across it. Fig. 9 shows the algorithms' archives after 5000 function evaluations, showing that solutions are concentrated close to the $f_1 - f_3$ and $f_1 - f_2$ planes together with a less dense covering of the shell between them for MOSA and Nam and Park's algorithm, while NSGA-II achieves an even coverage. Though the distribution of points across the front is more even with NSGA-II than MOSA, MOSA produced solutions which were far closer to the true front. Fig. 14 shows that the solutions generated by MOSA have a much lower volume measure; although visually the solutions from the NSGA-II runs seem superior to

MOSA's, the performance metrics suggest that MOSA has produced a better estimation of the true front. Deb *et al.* [29] observe that each run of NSGA-II in their experiments converged to a different part of the Pareto front; either to the $f_1 - f_2$ plane, the $f_3 - f_1$ plane, or distributed across the curved region of the front between these planes. The reason for the improved coverage of the PISA NSGA-II implementation is that the clustering close to the rims characteristic of the problem increases as solutions approach the true front. It is much more likely for solutions situated increasingly far from the true Pareto front to lie behind the central region of the front, although also to be dominated by the rims.

E. Density of Solutions on the Front

MOSA solutions on the front located by the annealer for problem DTLZ4 are close to the true Pareto front, but they are clearly inhomogeneously distributed across the front. Likewise, it is apparent from Figs. 5, 7, and 8, for problems DTLZ1, DTLZ2, and DTLZ3, that the density of solutions is greater close to the $f_1 - f_2$ plane than distant from it. Here, we discuss in some detail the reasons for this inhomogeneity; related work may be found in [35] and [36].

As we alluded to in Section III, when \mathbf{x} and \mathbf{x}' both lie on or very close to \mathcal{P} , then $\delta E(\mathbf{x}', \mathbf{x}) = 0$ and all proposals lying on the front are accepted, so that the trajectory of the current solution is a random walk in *parameter* space. The density of solutions on this front in *objective* space is governed by the mapping of area or volume from parameter space to objective space. Assuming that the $f_i(\mathbf{x})$ are continuous in a neighborhood of \mathbf{x} , the mapping is locally linear and is described the D by N Jacobian matrix of partial derivatives³

$$\mathbf{J}_{ij}(\mathbf{x}) = \frac{\partial f_i}{\partial x_j}(\mathbf{x}). \quad (17)$$

It is useful to write \mathbf{J} in terms of its singular value decomposition (SVD; see, for example, [38])

$$\mathbf{J} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (18)$$

Here, \mathbf{U} is a D by D matrix whose orthonormal columns \mathbf{u}_i ($i = 1, \dots, D$) form a local basis for objective space at $\mathbf{f}(\mathbf{x})$. Likewise, the D columns of $\mathbf{v} \in \mathbb{R}^{N \times D}$, denoted \mathbf{v}_i , ($i = 1, \dots, D$) are orthonormal N -dimensional vectors forming a local basis for the D -dimensional subspace of parameter space that locally maps to objective space. The matrix $\mathbf{\Sigma} \in \mathbb{R}^{D \times D}$ is diagonal, whose diagonal elements $\sigma_i \geq 0$ are known as singular values and are conventionally listed in descending order so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_D \geq 0$. The singular value σ_i quantifies the magnification of a perturbation in direction \mathbf{v}_i in parameter space: thus a small perturbation about \mathbf{x} of $\epsilon \mathbf{v}_i$ in parameter space yields a change in objective space from $\mathbf{f}(\mathbf{x})$ to $\mathbf{f}(\mathbf{x}) + \epsilon \sigma_i \mathbf{u}_i$.

If \mathbf{x} lies on the Pareto front, no parameter space perturbation can result in a change in objectives normal to the front, implying

that one of the singular values is zero and the rank of \mathbf{J} is at most $(D - 1)$. Assuming for simplicity that the Pareto front is $(D - 1)$ -dimensional, the direction normal to the front corresponds to \mathbf{u}_D and \mathbf{v}_D in objective and parameter spaces, respectively, and $\sigma_D = 0$. Perturbations lying in the span of $\mathbf{v}_1, \dots, \mathbf{v}_{D-1}$ result in traversal movements along the front and the (infinitesimal) volume in parameter space ν_p lying in $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{D-1})$ is magnified to volume

$$\nu_o = \nu_p \prod_{i=1}^{D-1} \sigma_i \quad (19)$$

on the Pareto front.

These ideas are illustrated in Fig. 10, which shows the volume magnification factor on the front for DTLZ1, DTLZ3, and DTLZ4. These were calculated by evaluating the Jacobian matrix at a large number of points in parameter space using a symbolic algebra package, and then numerically finding the singular values. Comparison with Figs. 5 and 6 for DTLZ1 and DTLZ3 makes it apparent that the magnification factors correspond to the density of solutions generated by the simulated annealer. If $\mathcal{X}_{\mathcal{P}} = \mathbf{f}^{-1}(\mathcal{P})$ is the $(D - 1)$ -dimensional manifold in parameter space that maps to the Pareto front, then this may be understood in terms of the annealer performing a random walk on $\mathcal{X}_{\mathcal{P}}$ which it covers fairly uniformly, producing a high density of solutions in objective space where the magnification factor is low, but a low density of solutions where the magnification factor is high because here solutions in parameter space are spread more thinly in objective space.

The bottom panels of Fig. 10 show the local volume magnification factors for DTLZ4, but with $\alpha = 2$ and $\alpha = 10$, rather than $\alpha = 100$ as recommended by Deb *et al.* [29], [33]. As the figure indicates, the magnification factor at points on the front even for $\alpha = 10$ is almost two orders of magnitude greater than the magnification factors for DTLZ1 and DTLZ3, when $\alpha = 100$ the pattern of magnification factors is similar but the range of magnifications is too great for sensible visualization. The magnification is least close to the $f_1 - f_2$ and $f_1 - f_3$ planes, corresponding precisely to the regions in which plenty of solutions are located by the annealer (Fig. 9) and greatest on the section of the front close to the $f_2 - f_3$ plane where few solutions are located. We infer that the annealer is locating and exploring $\mathcal{X}_{\mathcal{P}}$ in this case, but we see few solutions on parts of the front because the magnification factors are extremely high.

These deliberations lead us to consider again the question of what is an appropriate natural measure on the Pareto front. In our formulation of a multiobjective simulated annealer, we used an approximation to the Lebesgue measure, namely, the number of solutions in the archive, to evaluate the energy of a solution (9). However, this measure is defined in objective space and it might be argued that a more natural measure in objective space is the one induced by Lebesgue measure on $\mathcal{X}_{\mathcal{P}}$. In fact, as our experiments show, once the vicinity of the Pareto front has been located, it is (approximately) this induced measure that governs the density of solutions located. One may envisage that the singular value decomposition of \mathbf{J} may be used to counteract the inhomogeneity produced in objective space by the magnification factor by biasing the perturbations along the singular vectors \mathbf{v}_i

³In real problems, the Jacobian matrix may be estimated by finite differences or computer-aided differentiation packages, e.g., [37].

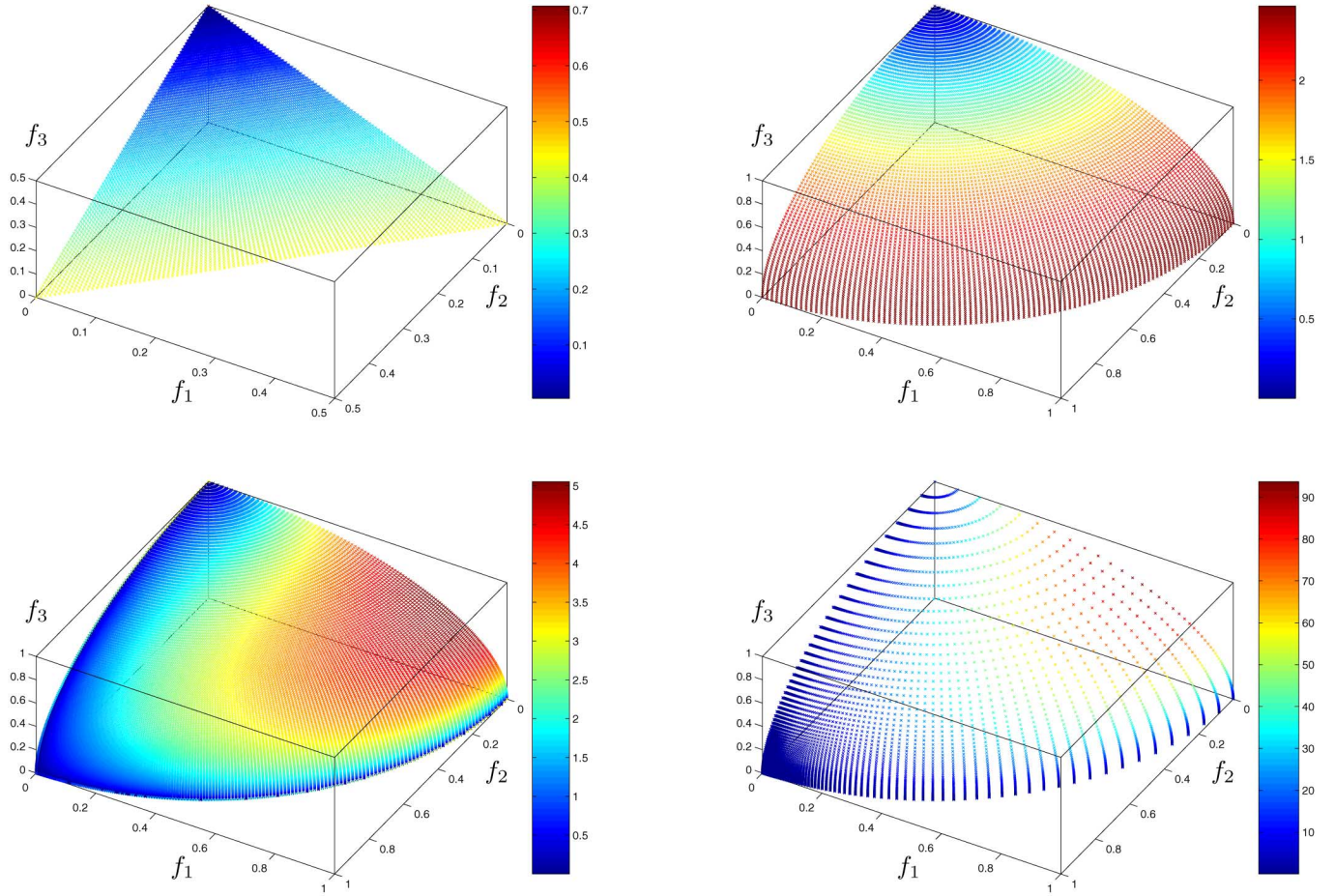


Fig. 10. Magnification factors on the Pareto front. Top left: DTLZ1; Top right: DTLZ3; Bottom left: DTLZ4 with $\alpha = 2$; Bottom right: DTLZ4 with $\alpha = 10$. Color indicates the local volume magnification factor from parameter space to objective space.

associated with large singular values σ_i . This is the subject of current research.

F. DTLZ 5

Fig. 11 shows the archives generated by the algorithms after 1000 function evaluations on test problem DTLZ5 for which the front is a one-dimensional curve rather than a full two-dimensional surface. As the distance plots show, the annealer has successfully located the one-dimensional front, while the other two algorithms generate sets which reside some distance behind this front; Deb *et al.* [29] also report that NSGA-II had not fully located the curve and yields a surface a little above the curve even after 20 000 function evaluations in their experiments. Fig. 14 shows that the MOSA archive dominates $\approx 90\%$ of the volume which is dominated by \mathcal{P} ; the true front is almost completely covered by the archive. This is the only test problem for which MOSA's archive does not grow larger (in the allowed iteration count) than NSGA-II's (enforced) set of 100 results, this is not especially significant however, as the NSGA-II set is significantly less well converged than MOSA's archive.

G. DTLZ 6

The front for DTLZ6 consists of four disjoint components.⁴ As Fig. 12 shows, the annealer is able to successfully locate each

⁴We use the formula given in [29] and [33]; the figures in these publications appear to have been generated with the f_3 objective scaled by a factor of 2.

of these components during a single run, that NSGA-II is able to generate solutions close to each front, and that Nam and Park's annealer does not converge in the allowed number of evaluations. Fig. 14 shows that, again, MOSA's coverage of the front, as well as the distance from the true front, dominates almost all the feasible search space. During optimization (and once the archive is close to the true Pareto front), we observe that the current solution \mathbf{x} of MOSA explores one component of the front for a few proposals before "jumping" to another component. If the regions of parameter space corresponding to each of the components of the front were widely separated, then it might be considerably more difficult for the annealer to simultaneously locate all components.

H. DTLZ 7

The DTLZ7 test problem is constructed using multiple constraint surfaces to yield a Pareto front consisting of a triangular planar section and a line segment. Fig. 13 shows the algorithm archives after 9000 function evaluations. The particular way in which DTLZ7 is constructed means that a perturbation of a single parameter of a solution lying on the front makes the perturbed parameter vector infeasible because it violates one of the constraints. Our schemes, described in Section VI-B, for adjusting the perturbation scalings rely on perturbing a single parameter at a time in order to keep track of the effect of the perturbation. However, this renders them ineffective for this problem:

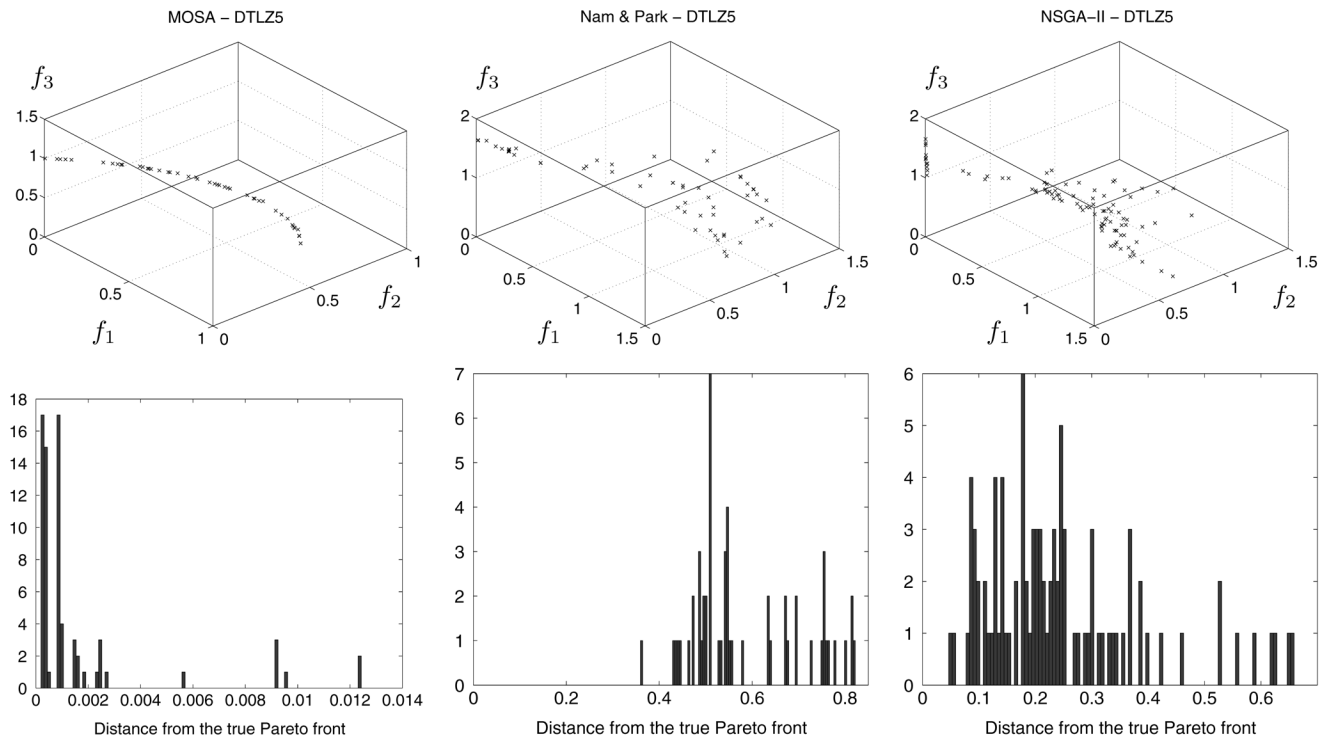


Fig. 11. Top: Archives on test problem DTLZ5 after 1000 function evaluations. Bottom: Histograms of the distance from the true Pareto front of the archive members (the 5% most distant have been omitted to aid visualization).

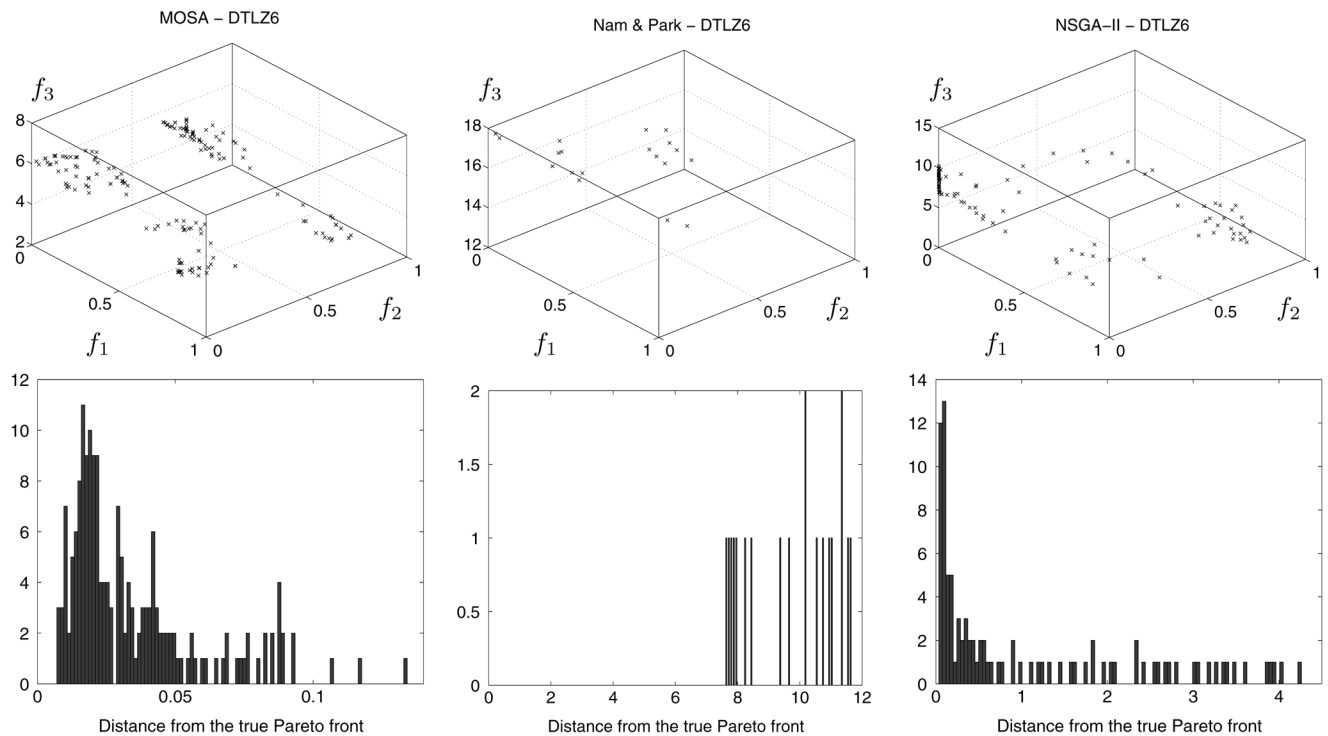


Fig. 12. Top: Archives on test problem DTLZ6 after 5000 function evaluations for each of the three algorithms. Bottom: Histograms of the distance from the true Pareto front of the archive members (the 5% most distant have been omitted in each of the six figures to aid visualization).

a single solution on the front is rapidly located, but the annealer is unable to explore the front because all perturbations result in infeasible proposals. For this reason the archive shown in Fig. 13 was generated by perturbing a randomly chosen number of parameters for each proposal; for simplicity the perturbation scales were kept constant at 0.1 of the feasible region throughout the

optimization. While more efficient perturbation schemes could probably be devised, the figure shows that the annealer is reasonably successful in locating the central portion of the front, although the extremities of the front have not been explored and there remain some extraneous solutions close to constraint surfaces bounding the front, but still quite distant from \mathcal{P} itself.

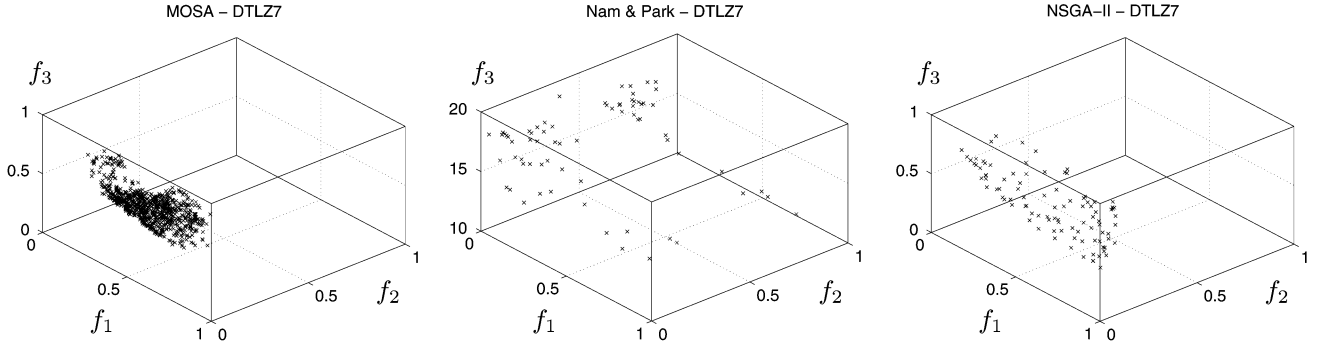


Fig. 13. Archives on test problem DTLZ7 after 9000 function evaluations for each of the three algorithms.

We also modified the single parameter perturbation scheme used in our implementation of Nam and Park's annealer to perform the same multiple point perturbations as MOSA. NSGA-II, the PISA implementation of which already used a (more advanced) multiple parameter perturbation, did not need to be modified for this problem. Figs. 13 and 14 show that, while MOSA has again converged well, and generates the solutions closest to, the true front, NSGA-II demonstrates the best coverage of solutions over the front towards the extremes of the constraints. It should be noted that the need to adapt to a multiple parameter perturbation scheme will be present for all algorithms which employ a specialized single parameter perturbation scheme (conversely, problems can be constructed that would prevent a multiple parameter perturbation scheme from converging to the true front).

I. Statistical Performance Measures

Unlike single-objective problems, solutions to multiobjective optimization problems can be assessed in several different ways. Therefore, in order to quantify the convergence of the algorithms, we measure two distinct properties. First, we calculate the average distance of the archived solutions discovered from the true front to ascertain how close on average solutions found are to the true front. Rather than using the root mean square distance which is susceptible to outliers, here we use the median distance of solutions in the archive

$$\bar{d}(F) = \text{median}_{\mathbf{x} \in F} [d(\mathbf{x})] \quad (20)$$

where $d(\mathbf{x})$ is the minimum Euclidean distance between \mathbf{x} and the true front \mathcal{P} . Clearly, this measure depends on the relative scaling of the objective functions, however, it yields a fair comparison here because the objectives for the DTLZ test functions have similar ranges.

Second, since we are concerned with finding solutions spread across the true Pareto front, we also use a variant of the volume \mathcal{V} measure [24] which is conceptually similar to the performance measure used in [39]. The idea is to calculate the amount of objective space that is dominated by the true front, but not by the calculated archive. To make this precise, let H be the minimum axis-parallel hypercube in objective space which contains \mathcal{P} . Then, $\mathcal{V}(\mathcal{P}, F)$ is the fraction of H which is dominated by \mathcal{P} but not by F . Clearly, this measure is zero when F covers the entire Pareto front and it approaches zero as F approaches \mathcal{P} . Importantly however, an archive comprised of a few solutions clustered together on the true front will have a larger $\mathcal{V}(\mathcal{P}, F)$ than an archive of solutions well spread across the front and,

therefore, dominating a larger fraction of objective space. This measure is straightforwardly calculated by Monte Carlo sampling (10^5 samples here) of H and counting the fraction of samples dominated exclusively by \mathcal{P} and not F ; see [24] for details.

Fig. 14 shows box plots over 20 runs, from different randomly selected initial solutions, of the median Euclidean distance $\bar{d}(F)$ fractional volume measures and archive size of the results for each algorithm on each test problem.

The distance of \mathcal{P} to the objective space origin is $\mathcal{O}(1)$ for all of these problems, so it can be seen from Fig. 14 that the annealer is able to converge very close to the front for all seven problems. In fact, MOSA is significantly closer to the front, (as described in Section VII) than both NSGA-II and Nam and Park's annealer. NSGA-II was able to converge to a set near to the true front for five of the problems (with two of those being very near) and Nam and Park's annealer was able to generate an archive near the true front on one of the problems.

The middle row of Fig. 14 shows $\mathcal{V}(\mathcal{P}, F)$, the fractional volume dominated by \mathcal{P} and not by F . As the figure indicates, the annealer both converges well to \mathcal{P} and also covers it reasonably well for all the problems. MOSA dominates significantly more volume than NSGA-II for six of the seven cases although NSGA-II is significantly better on DTLZ7. NSGA-II achieved a good coverage on those problems for which it could converge near to the true front; the diversity maintenance in the algorithm encourages this. NSGA-II performed particularly well on DTLZ7, where the coverage was better than MOSA's. Nam and Park's algorithm was unable to effectively cover the true front for any problem.

The results for DTLZ4 effectively demonstrate why it is necessary to measure convergence in terms of both distance and coverage, with MOSA having converged close to \mathcal{P} , but yielding a poor coverage of the front (in objective space), an artifact of the large range of volume magnification factors, as discussed earlier, also demonstrating that the visually appealing NSGA-II results were less well converged than it seems upon inspection. Confirming the impression given by the single run depicted in Fig. 13, on average, the annealer does not completely cover the true front for DTLZ7. As discussed above, this could probably be improved by designing particular perturbation strategies for this particular problem; the NSGA-II implementation has a multiple point mutation scheme which performs very well on this problem.

Fig. 14 also shows how the final archive size varies across the 20 runs for each of the DTLZ problems used here. For the MOSA results, it is clear that even the fronts generated by

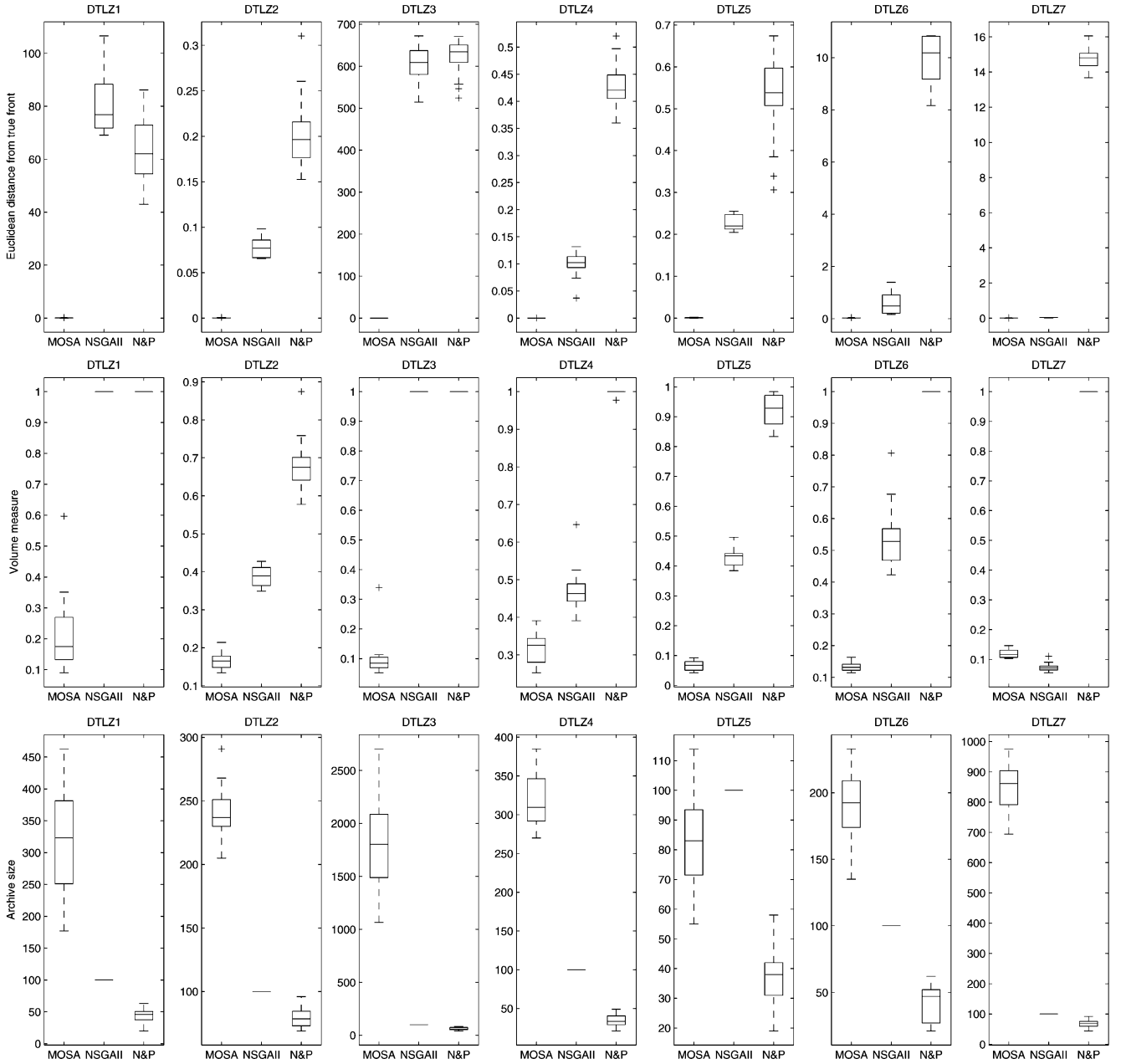


Fig. 14. Top: Box plots of the average distance $\bar{d}(F)$ of the archive from the true Pareto front for 20 runs of each DTLZ test problems, using the documented run lengths. Middle: Box plots of the volume measure $\mathcal{V}(\mathcal{P}, F)$ of the archive for each run. Bottom: Box plots of the size of the archive for each run. Each figure shows the results for MOSA, NSGA-II, and Nam and Park's annealer.

the least well-covered runs for each problem contain a large quantity of solutions relative to the run length. Furthermore, the number of solutions generated for each problem is consistent across runs, although, as may be expected, problems with multiple local fronts (DTLZ1 and DTLZ3) have a larger spread. The NSGA-II algorithm is constrained to a predefined size (100 solutions in the work presented here) and Nam and Park's annealer does not generate large sets of solutions as it does not converge close to the true front.

In these comparisons, we have allowed relatively small numbers of evaluations to each algorithm in order to test rapid convergence. It could be claimed, however, that this prejudices the results against the population based search of NSGA-II and in favor of MOSA, as it might be expected that MOSA

would demonstrate rapid convergence and slow coverage, while NSGA-II would converge slowly but demonstrate superior coverage subsequent to convergence. While the results presented earlier show that MOSA does not demonstrate this behavior, additional experiments were undertaken, allowing NSGA-II 100 000 function evaluations for each of DTLZ1, DTLZ2, and DTLZ3 (DTLZ1 and DTLZ3 being the two most difficult to converge to with multiple local fronts, and DTLZ2 being the least difficult). Over the course of the experiments, the archives generated by MOSA shown earlier for the low evaluation counts were significantly closer to the true front than those of NSGA-II after 100 000 evaluations (this is unsurprising given the previously published results of NSGA-II on these problems [33]) and also had a greater dominated volume.

VIII. CDMA NETWORK OPTIMIZATION

Mobile telephone subscribers are allocated to one of a number of distinct *cells* or *sectors* comprising the telephone network. Cells may vary in extent from a few tens of meters (in a large office building) to several kilometers (in rural areas). Each cell is served by a single antenna and as the phone subscriber moves to a new location a “handover” is made to a new cell in which the radio signal is stronger. The performance of the network as a whole and the quality-of-service enjoyed by individual subscribers is dependent upon many operating parameters, some associated with the antenna and radio interface itself (such as the antenna azimuth and downtilt) and others associated with the network as a whole, such as the handover policy [40]. In addition, performance itself may be evaluated in terms of several different metrics, for example: the network capacity (number of simultaneous calls); coverage (area served); and mean cell traffic channel power. The simultaneous optimization of all these competing objectives is generally impossible and here we use the MOSA to investigate the tradeoffs between them. We draw attention to recent work using multiobjective optimization in the mobile telecommunications domain: Ben Jamaa *et al.* [41] have used multiobjective genetic algorithms for cell planning in order to optimize the cost and coverage of a network, and Szabó *et al.* [42] have used multiobjective evolutionary algorithms for discovering the cost-interference tradeoff when allocating transmitter placement and assigning transmission frequencies in time-division multiple-access (TDMA) networks.

We treat as parameters to be optimized the pilot powers of a CDMA network. The pilot power may be loosely thought of as the power with which the cell transmits to establish initial communication with phones in its cell. Pilot power is a particularly important parameter in CDMA networks because cells transmit continuously and if the pilot power is too great a cell may drown out its neighbors, but will not be heard if the pilot power is too low. In this study, there were 94 pilot powers as parameters and we optimized three objectives: the pilot pollution factor, defined as the number of pilots that each subscriber receives within 5 dB of the dominant pilot; the mean downlink traffic channel (TCH) outage factor defined as the number of subscribers attempting to exceed their TCH power limit; and the mean reverse link server penalty which quantifies the unbalancing of the reverse link. This is done by comparing the propagation loss between each subscriber and its serving cell and the smallest propagation loss between that subscriber and any cell—the average difference between these values across all subscribers is defined as the reverse link penalty.

Unlike optimization of test problems, as in Section VII, the properties of the CDMA search space are not known in advance. Particularly, it is not known if the problem exhibits local front behavior, where an optimizer must make several successive movements out of a locally optimal region of parameter space in order to locate the globally optimal region which corresponds to the Pareto front in objective space.

The results reported here are for an operational CDMA-1X network consisting of 94 sectors. Computational optimization is feasible for this system due to the employment of a proprietary mathematical model of the downlink air interface which permits rapid evaluation of new configurations. In this study, the pilot

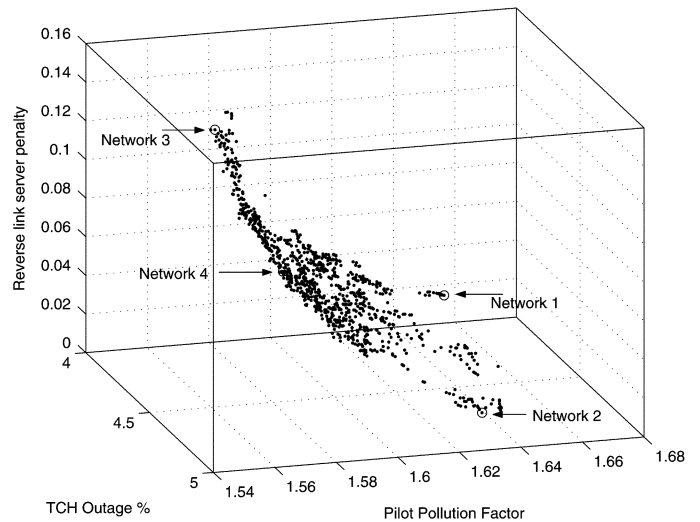


Fig. 15. Estimated Pareto front for network pilot power optimization.

power of each sector was allowed to vary over the range from 1.0 to 3.0 W in 0.5 W intervals. Initially, the pilot powers were set to their minimum feasible values and the MOSA was initialized to a temperature that yielded a 50% acceptance rate for derogatory moves, as described in Section VI-A. The computational temperature was then reduced every $L_k = 100$ proposals by a factor of $\beta_k = 0.958$. The annealer was run for 100 000 evaluations of the objective function.

Fig. 15 shows the estimated Pareto front obtained, which consists of 965 solutions. We remark that a standard genetic algorithm optimizing a composite objective function locates after 100 000 function evaluations a single solution which is dominated by almost all of the nondominated archive. The GA solution is distant from the archive but it is likely that the genetic algorithm would have located a point in the vicinity of the front generated by the simulated annealer if it had been permitted a greater number of objective evaluations. Of course, the principal advantage of the annealer is not the reduction in the time taken to find a desirable solution (although this is considerable), but the frontal nature of the results generated by the simulated annealer. The front which has been located is clearly curved in objective space and displays to the network engineer the range of tradeoffs which may be made in configuring the network.

The central portion of the network configuration corresponding to each of the solutions circled in Fig. 15 is shown in Fig. 16. In this figure, the pilot power for a sector is indicated by the length of an arrow rooted at the antenna location (antennae masts frequently support two or three antennae serving different sectors). It is interesting to note that each of the network configurations is very similar, despite their extreme relative frontal locations. This figure provides important information about the network to a network engineer, since some pilot powers seem to have single optimal values, particularly those in the less populated areas of the network with fewer interactions. However, as might be expected, it can be seen that the configuration with low pilot pollution (network 3) restricts pilot powers in sectors that face each other. The interaction between pilot powers and the other objectives is less clear, but the MOSA methodology provides a mechanism for locating

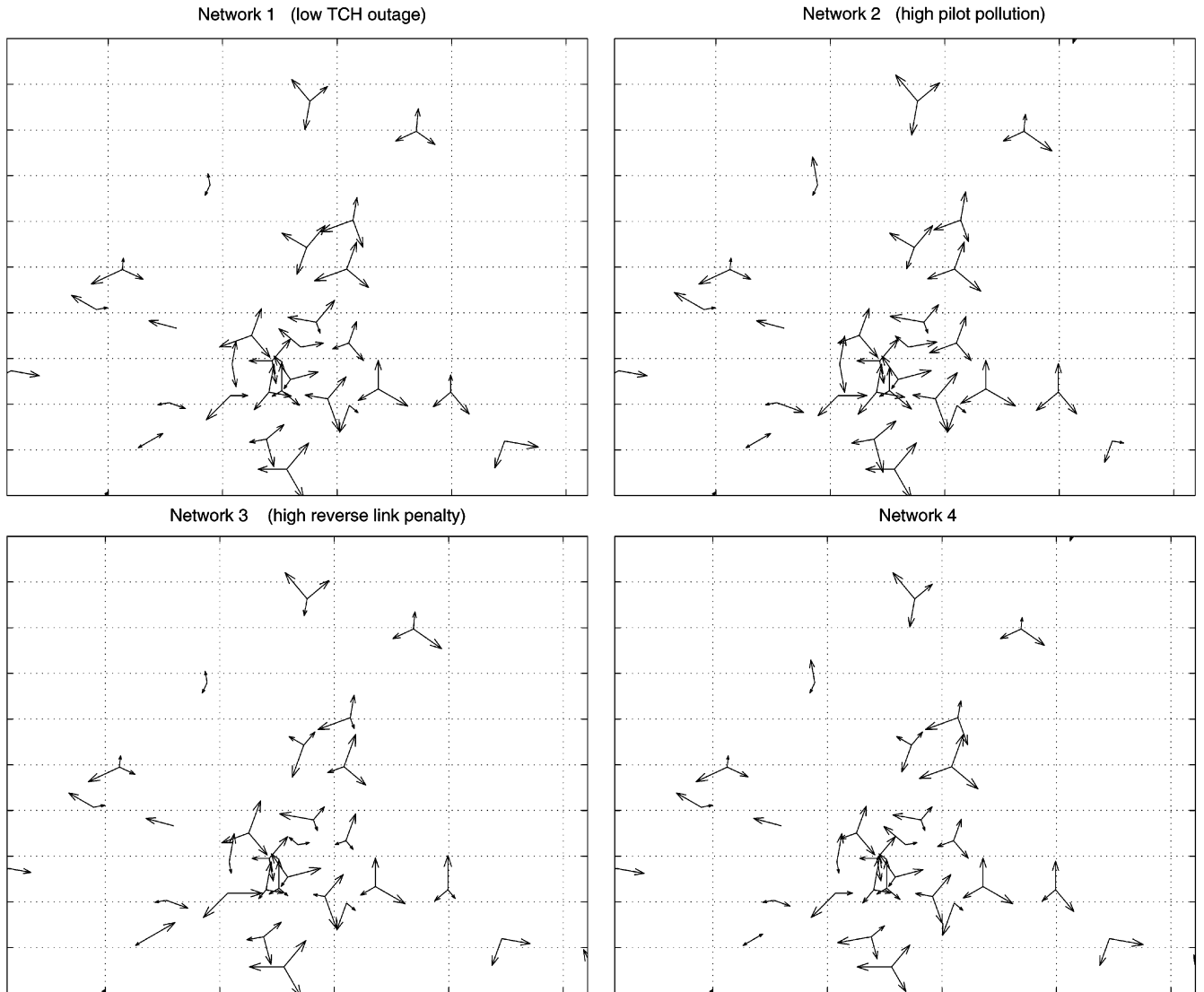


Fig. 16. Pareto-optimal central network configurations corresponding to the labelled points on the Pareto front in Fig. 15.

these optimal configurations. Current work involves applying this methodology to the optimization of antennae pilot powers, azimuths, and downtilts in larger networks.

IX. CONCLUSION

We have presented an energy measure for use in multiobjective SA, which is based on the fundamental notion of dominance, rather than employing a weighted combination of the objectives. Simulated annealers employing this measure were shown to have good convergence properties on the first seven DTLZ test functions [29], [33]. An extensive comparison with the evolutionary algorithm NSGA-II and the multiobjective annealer proposed by Nam and Park [8] on these problems shows that the annealer consistently generates archives closer to the true front than NSGA-II and Nam and Park's annealer and that in all but one case produces a significantly better coverage (on DTLZ7 NSGA-II generates fronts with a fuller coverage of the front, possibly due to the more specialized multiple point perturbations used).

We emphasize that the MOSA algorithm was not "tuned" for each of these problems, but run from a randomly chosen initial condition. More rapid convergence on many of these problems can be achieved by careful tuning of the annealing schedule but, of course, this requires many runs to discover a reasonable schedule; while tuning the annealing schedule is important in industrial applications where the annealer is to be run many times on similar problems (e.g., CDMA network optimization), we have refrained from tuning the annealer to particular *test* problems; here we simply limit all the algorithms to an evaluation count approximately at which the first algorithm converges on the true front. The only instance in which it was necessary to alter the annealers was for DTLZ7, for which single-point perturbations of solutions close to the Pareto front result in infeasible proposals, however, it should be noted that the problem with single-point perturbations will afflict all stochastic searches (evolutionary algorithms, GAs, etc.) that perturb a solution to generate a new candidate solution and that in this problem, MOSA performed almost as well with a very basic multiple-point perturbation scheme as NSGA-II did, which uses a more advanced scheme.

One possible limitation of the MOSA scheme is that the repeated perturbation of the single solution might make it difficult to explore a Pareto front which corresponds to several disjoint regions of parameter space (cf DTLZ6 in Section VII-G). However, work on schemes that permit perturbations across the front suggests that, in general, they do not converge more rapidly [43].

An advantage of the dominance based energy measure is that it is not *a priori* biased towards any part of the front. Weighted sum optimizers implicitly use distance information in objective space, which renders them sensitive to the relative scalings of the objectives, whereas our algorithm is robust to rescalings of the objectives. Indeed, if the relative importance or scales of the objectives were known in advance it might be more straightforward to optimize a single, appropriately weighted, sum of the objectives. Notions of dominance and Pareto optimality are well suited to handling competing objectives whose relative importance is *a priori* unknown, and it is therefore natural to eschew metric information in favor of dominance concepts in order to guide the search. Indeed, we have argued that the dominance-based energy tends to promote exploration in sparsely populated regions and, in practice, we have shown that estimated fronts evenly and widely cover the true front. An area of current investigation is to use the singular values and vectors of the Jacobian matrix to guide the search on the front towards areas that would otherwise be sparsely populated.

Determining an efficient scale on which to make proposals is more complicated in the multiobjective case than the uniojective case, because some proposals work to advance the front, while others traverse the front. We have proposed simple heuristics to adapt the perturbation scales and future work involves applying machine learning techniques to learn the local mapping between parameter and objective space in order to more sensitively control the search direction.

When applied to the optimization of a CDMA network, the annealer is successful in generating a front with a large number of mutually nondominating solutions, the vast majority of which are superior to the single solution located by a genetic algorithm optimizing a composite objective function. This allows a network engineer to make an informed decision regarding network configurations with additional knowledge of the costs of the tradeoffs involved. Further work in this direction will be focused on optimizations involving both more parameters, such as antenna azimuths and downtilts, and additional objectives relating to the quality of service for subscribers.

Our $E(\mathbf{x})$ is a measure of a portion of the dominating set, namely, $\mu(\hat{F}_{\mathbf{x}})$, which is a close relation to Fleischer's recently proposed measure [21]; loosely, our measure deals with the area of the dominating surface—the attainment surface—while Fleischer's considers the dominated volume. It would be interesting to investigate the convergence of an annealer based on Fleischer's measure but, as shown in [44], the complexity of the calculation is polynomial in the number of archived solutions and exponential in the number of objectives; this makes calculation for even three objectives infeasibly slow for use as the energy calculation for an annealer. Although a proof of convergence for simulated annealers based on our measure remains to be completed, this is an area of current work, together with the application of the annealer to other large scale problems.

ACKNOWLEDGMENT

The authors would like to thank M. Fisher and the anonymous reviewers for their useful comments. They would also like to thank the authors of the PISA code [34] for making their implementation of the NSGA-II algorithm available for use by the academic community.

REFERENCES

- [1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [2] N. Metropolis, A. W. Rosenbluth, M. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1092, 1953.
- [3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.
- [4] C. A. Coello Coello, "A comprehensive survey of evolutionary-based multiobjective optimization techniques," *Knowledge and Information Systems: An Int. J.*, vol. 1, no. 3, pp. 269–308, 1999.
- [5] D. van Veldhuizen and G. Lamont, "Multiobjective evolutionary algorithms: Analyzing the state-of-the-art," *Evol. Comput.*, vol. 8, no. 2, pp. 125–147, 2000.
- [6] P. Engrand, "A multi-objective approach based on simulated annealing and its application to nuclear fuel management," in *Proc. 5th Int. Conf. Nuclear Eng.*, Nice, France, 1997, pp. 416–423.
- [7] P. Czyzak and A. Jaskiewicz, "Pareto simulated annealing—A metaheuristic technique for multiple-objective combinatorial optimization," *J. Multi-Criteria Decision Anal.*, vol. 7, pp. 34–47, 1998.
- [8] D. K. Nam and C. H. Park, "Multiobjective simulated annealing: A comparative study to evolutionary algorithms," *Int. J. Fuzzy Syst.*, vol. 2, no. 2, pp. 87–97, 2000.
- [9] M. Hapke, A. Jaskiewicz, and R. Slowinski, "Pareto simulated annealing for fuzzy multi-objective combinatorial optimization," *J. Heuristics*, vol. 6, no. 3, pp. 329–345, 2000.
- [10] A. Suppaitarm, K. A. Seffen, G. T. Parks, and P. J. Clarkson, "A simulated annealing algorithm for multiobjective optimization," *Eng. Opt.*, vol. 33, pp. 59–85, 2000.
- [11] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Proc. Parallel Problem Solving from Nature—PPSN VI*, 2000, pp. 849–858.
- [12] K. I. Smith, R. M. Everson, and J. E. Fieldsend, "Dominance measures for multi-objective simulated annealing," in *Proc. Congr. Evol. Comput., CEC04*, 2004, pp. 23–30.
- [13] L. Ingber, "Simulated annealing: Practice versus theory," *Math. Comput. Modelling*, vol. 18, pp. 29–57, 1993.
- [14] P. Salamon, P. Sibani, and R. Frost, "Facts, conjectures, and improvements for simulated annealing," Society for Industrial and Applied Mathematics, 2002, ser. Monographs on Mathematical Modeling and Computation, no. 7.
- [15] P. Serafini, "Simulated annealing for multiobjective optimization problems," *Multiple Criteria Decision Making. Expand and Enrich the Domains of Thinking and Application*, pp. 283–292, 1994.
- [16] E. L. Ulungu, J. Teghaem, P. Fortemps, and D. Tuytens, "MOSA method: A tool for solving multiobjective combinatorial decision problems," *J. Multi-Criteria Decision Analysis*, vol. 8, pp. 221–236, 1999.
- [17] D. Tuytens, J. Teghem, and N. El-Sherbeny, "A particular multiobjective vehicle routing problem solved by simulated annealing," in *Metaheuristics for multiobjective optimisation*, ser. Lecture Notes in Economics and Mathematical Systems, X. Gandibleux, M. Sevaux, K. Sörensen, and V. T'kindt, Eds. New York: Springer, 2003, vol. 535, pp. 133–152.
- [18] M. Hapke, A. Jaskiewicz, and R. Slowinski, "Pareto simulated annealing for fuzzy multi-objective combinatorial optimization," *J. Heuristics*, vol. 6, no. 3, pp. 329–345, 2000.
- [19] I. Das and J. Dennis, "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," *Structural Optimization*, vol. 14, no. 1, pp. 63–69, 1997.
- [20] A. Jaskiewicz, "Comparison of local search-based metaheuristics on the multiple objective knapsack problem," *Foundations of Comput. Decision Sci.*, vol. 26, no. 1, pp. 99–120, 2001.
- [21] M. Fleischer, "The measure of Pareto optima: Applications to multi-objective metaheuristics," in *Evolutionary Multi-Criterion Optimization, Second International Conference, EMO2003*, ser. Lecture Notes in Computer Science, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. New York: Springer, 2003, vol. 2632, pp. 519–533.

- [22] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multi-objective optimization: Formulation, discussion, and generalization," in *Proc. 5th Int. Conf. Genetic Algorithms*, 1993, pp. 416–423.
- [23] D. Greening, "Simulated annealing with inaccurate costs functions," in *Proc. Int. Congr. Math. Comput. Sci., IMACS*, Dublin, Ireland, 1993, Trinity College.
- [24] J. E. Fieldsend, R. M. Everson, and S. Singh, "Using unconstrained elite archives for multi-objective optimisation," *IEEE Trans. Evol. Comput.*, vol. 7, no. 3, pp. 305–323, Jun. 2003.
- [25] M. Jensen, "Reducing the run-time complexity of multi-objective EAs: The NSGA-II and other algorithms," *IEEE Trans. Evol. Comput.*, vol. 7, no. 5, pp. 502–515, Oct. 2003.
- [26] T. Hanne, "On the convergence of multiobjective evolutionary algorithms," *Eur. J. Oper. Res.*, vol. 117, pp. 553–564, 1999.
- [27] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: Methods and applications," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, ETH, 1999, diss ETH No. 13398.
- [28] V. Grunet Da Fonseca, C. M. Fonseca, and A. O. Hall, "Inferential performance assessment of stochastic optimisers and the attainment function," in *Proc. 1st Int. Conf. Evol. Multi-Criterion Opt.*, E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, Eds. Berlin, Germany: Springer-Verlag, 2001, vol. 1993, Lecture Notes in Computer Science, pp. 213–225.
- [29] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," *Proc. Cong. Evol. Comput. (CEC'2002)*, vol. 1, pp. 825–830, 2002.
- [30] M. Laumanns, G. Rudolph, and H. P. Schwefel, "Mutation control and convergence in evolutionary multi-objective optimisation," in *Proc. 7th Int. Mendel Conf. Soft Comput. (MENDEL 2001)*, R. Matousek and P. Osmera, Eds., Brno, Czech Republic, 2001, pp. 24–29.
- [31] I. F. Szalzarini, S. Müller, and P. Koumoutsakos, "Microchannel optimization using multiobjective evolution strategies," in *Proc. 1st Int. Conf. Evol. Multi-Criterion Opt. (EMO)*, 2001, Lecture Notes in Computer Science, pp. 516–530, ser. .
- [32] D. Büche, S. D. Müller, and P. Koumoutsakos, "Self-adaptation for multi-objective evolutionary algorithms," in *Proc. 2nd Int. Conf. Evol. Multi-Criterion Opt., EMO 2003*, ser. Lecture Notes in Computer Science, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer, 2003, vol. 2632, pp. 267–281.
- [33] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," Institute für Technische Informatik und Kommunikationsnetze, ETH Zurich, Tech. Rep. 112, 2001.
- [34] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, "PISA—A platform and programming language independent interface for search algorithms," in *Proc. Evol. Multi-Criterion Opt. (EMO 2003)*, ser. Lecture Notes in Computer Science, C. Fonseca, P. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. New York: Springer, 2003, pp. 494–508.
- [35] T. Okabe, Y. Jin, and B. Sendhoff, "On the dynamics of evolutionary multi-objective optimisation," in *Proc. Genetic Evol. Comput. Conf.*, 2002, pp. 247–256.
- [36] T. Okabe, Y. Jin, M. Olhofer, and B. Sendhoff, "On test functions for evolutionary multi-objective optimization," in *Parallel Problem Solving from Nature, VIII*, ser. Lecture Notes in Computer Science 3242. New York: Springer, 2004, pp. 792–802.
- [37] M. Berz, C. Bischof, G. Corliss, and A. Griewank, Eds., *Computational Differentiation: Techniques, Applications, and Tools*. Philadelphia, PA: SIAM, 1996.
- [38] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Oxford, U.K.: North Oxford Academic, 1983.
- [39] M. Laumanns, E. Zitzler, and L. Thiele, "A unified model for multi-objective evolutionary algorithms with elitism," in *Proc. Congr. Evol. Comput.*, 2000, pp. 46–53.
- [40] J. Korhonen, *Introduction to 3G mobile communications*, ser. Mobile communications series. Norwood, MA: Artech, 2001.
- [41] S. Ben Jamaa, Z. Altman, J. M. Picard, and B. Fourestié, "Multi-objective strategies for automatic cell planning of UMTS networks," in *Proc. Vehicle Technol. Conf.*, Milan, Italy, 2004.
- [42] G. Szabó, K. Weicker, N. Weicker, and P. Widmayer, "Evolutionary multiobjective optimization for base station transmitter placement with frequency assignment," *IEEE Trans. Evol. Comput.*, vol. 7, pp. 189–203, 2003.
- [43] K. I. Smith, R. M. Everson, and J. E. Fieldsend, "Simulated annealing and greedy searches for multi-objective optimisation problems," School of Eng., Comput. Sci. Math., The Univ. Exeter, Exeter, U.K., Tech. Rep., 2006. [Online]. Available: <http://www.secam.ex.ac.uk/people/ksmith/mosa/>
- [44] L. While, "A new analysis of the LeBMeasure algorithm for calculating hypervolume," in *Proc. 3rd Int. Conf. Evol. Multi-Criterion Opt., EMO*, 2005, vol. 3410, Lecture Notes in Computer Science, pp. 326–340.



Kevin I. Smith graduated with a degree in computer science and received the Ph.D. degree in computer science, studying simulated annealing techniques for multiobjective optimization from the University of Exeter, Exeter, U.K., in 2002 and 2007, respectively.

He is currently with the University of Exeter in a collaboration with the ai Corporation investigating credit card fraud classification. His current interests lie in multiobjective optimization and pattern recognition for classification.



Richard M. Everson (M'00) graduated with a degree in physics from Cambridge University, Cambridge, U.K., in 1983 and the Ph.D. degree in applied mathematics from Leeds University, Leeds, U.K., in 1988.

He worked at Brown and Yale Universities on fluid mechanics and data analysis problems until moving to Rockefeller University, Rockefeller, NY, to work on optical imaging and modeling of the visual cortex. After working at Imperial College, London, U.K., he joined the Department of Computer Science, Exeter

University, where he is now an Associate Professor of Machine Learning. Current research interests are in statistical pattern recognition, multiobjective optimization and the links between them.



Jonathan E. Fieldsend (S'00–M'02) received the B.A. degree (Hons.) in economics from the University of Durham, Durham, U.K., in 1998 and received an EPSRC studentship and received the M.Sc. degree in computational intelligence from the University of Plymouth, Plymouth, U.K. In November 1999, he received a further studentship from the University of Exeter, Exeter, U.K., to undertake Ph.D. research, completing a thesis on new algorithms for multiobjective search and their application to evolutionary neural network training.

He is currently a Lecturer in Computer Science at the School of Engineering, Computer Science and Mathematics, University of Exeter. He has previously held Research and Business Fellowships at the same institution. His current research interests include reversible jump MCMC methods, multiobjective optimization, pattern recognition, and machine learning.

Dr. Fieldsend is a Member of the IEEE Computational Intelligence Society and the IEEE Signal Processing Society.



Chris Murphy received the Ph.D. degree in satellite remote sensing from Aston University, Birmingham, U.K., during which time he was commissioned by the European Space Agency to calibrate satellite instrumentation.

He subsequently spent two years developing models of the human respiratory system for clinical decision support tools. He has been at Motorola for the last seven years, where he defined and commercialized innovative services for 3G and 4G cellular network operators. He is an author of several

conference and journal papers and has filed four patents.



Rashmi Misra received the M.B.A. degree with Distinction from the University of Warwick, Warwick, U.K., and the Ph.D. degree in artificial intelligence from the University of Exeter, Exeter, U.K.

She has been heading the Seamless Mobility Consulting Services Initiative in EMEA and Asia PAC for Motorola Global Services and has been working at the cutting edge of mobile technology for almost ten years. She has several patents in the area of mobile telecommunications. She has a track record of taking innovative and early technology

projects to market.