Approximating the Pareto Set: Concepts, Diversity Issues, and Performance Assessment

Marco Laumanns   Günter Rudolph   Hans-Paul Schwefel

March 15, 1999

Abstract

This paper adresses the problem of diversity in multiobjective evolutionary algorithms and its implications for the quality of the approximated set of efficient solutions (Pareto set). Current approaches for maintaining diversity are classified and related to the overall fitness assignment strategy. The resulting groups of complex selection operators are presented and tested on different objective functions exhibiting different levels of difficulty. For the assessment of the algorithmic performance a quality measure based on the notion of dominance is applied that reflects gain of information produced by the algorithm. This allows an on-line and time-dependent evaluation in order to characterize the dynamic behavior of an algorithm.

1 Introduction

Evolutionary algorithms (EA) have been applied to optimization problems with multiple criteria since the 1980s. Schaffer's VEGA [Schaffer, 1985] has been the first algorithm that was able to cope with different objectives simultaneously without scalarization by aggregation and solving a single objective surrogate problem instead. Since then many different approaches and algorithms have been proposed and tested both on mathematical test problems and on real world applications. The concept of populations of distinct individuals motivates the use of EA for finding a set of compromize solutions for a multi-objective problem. Some algorithms let different criteria dictate the selection or deletion or different parts of the population while others make use of Pareto's [Pareto, 1972] notion of dominance to compare invidiual solutions. An extensive survey can be found in [Fonseca and Fleming, 1995] and [Fonseca and Fleming, 1997].

Though the choice of a specific algorithm may depend on the preferences of the decision maker as well as on technical preconditions (such as availability of software, representational issues or applicability on parallel machines), there will still be a variety of alternatives left to choose from. These can be used "as is" or "recombined" and refined further to create new algorithms. In order to compare different algorithms – or rather methods – performance criteria need to be defined.

In most cases selection is the operator that receives the lion's share of the researchers' interest. This is certainly due to the fact that selection must somehow reveal a preference relation on the set of individuals and that the case of multiple criteria usually implies the existence of non-comparable individuals.

Another major issue is the expected outcome of the algorithm. In contrast to the single objective case, where usually the global optimum has to be found, we now face a set of efficient or Pareto-optimal solutions. This raises the question of convergence as well as of the distribution of solutions. If we consider the Pareto set as the "solution" of a multi-objective optimization problem we want an appropriate algorithm to approximate it as good as possible. Hence, preservation of the individuals' diversity is crucial, not only to avoid premature convergence, but also not to loose any potentially efficient solution.

This study tries to classify different selection schemes together with different methods to maintain diversity. For each class a representative algorithm is implemented and tested on highdimensional mathematical test problems according to a performance criterion based on the size of the objective space that an algorithm identifies as being dominated. Here, the main focus will be on the development of the trade-off surface over time.

2 Concepts and approaches

2.1 Basic definitions

Let $f : \mathcal{X} \to \mathbb{R}^q$ with $\mathcal{X} \subseteq \mathbb{R}^\ell$ and $q \geq 2$ be a vector-valued function that maps a decision vector $x \in \mathcal{X}$ to an objective vector $y = f(x) \in \mathbb{R}^q$. In the ideal case the objective functions $f_i : \mathcal{X} \to \mathbb{R}$ should be minimized simultaneously for $i = 1, \ldots, q$. The problem is, however, that the set of objective vectors is not totally ordered. Let $\mathcal{F} = \{f(x) : x \in \mathcal{X} \subseteq \mathbb{R}^\ell\} \subset \mathbb{R}^q$ be the set of objective vectors that are attainable under the mapping f. An objective vector $y^* \in \mathcal{F}$ is said to dominate another objective vector $y \neq y^*$, written $y^* \prec y$, if $y_i^* \leq y_i$ for all $i = 1, \ldots, q$. If $\nexists y \in \mathcal{F}$ with $y \prec y^*$, then y^* in called Pareto-optimal with respect to f. The set $\mathcal{F}^*$ of all Pareto-optimal objective vectors is called the Pareto set. Each decision vector $x^* \in \mathcal{X}^* = \{x \in \mathcal{X} \subseteq \mathbb{R}^\ell : f(x) \in \mathcal{F}^*\}$ is termed an efficient solution or a Pareto-optimal decision vector of the multi-objective optimization problem. Since the sets $\mathcal{X}^*$ and $\mathcal{F}^*$ can be analytically determined only in exceptional cases and since the dimension of $\mathcal{X}^*$ as well as $\mathcal{F}^*$ may be as large as $\min\{\ell, q-1\}$, numerical methods for finding the set of Pareto-optimal decisions are generally restricted to approximating the shape of the set $\mathcal{X}^*$.

2.2 Performance criteria

In most cases collecting information about the trade-offs between objectives prior to decision making is very important, or even necessary, for solving optimization problems with multiple objectives. Ideally the algorithm should find the set of all nondominated solutions, which will constitute the global trade-off (hyper)surface. But since resources are finite, all algorithms are restricted to approximate the Pareto set by a number of representative solutions. Yet even these samples give us very limited insight into the real trade-off surface, because we cannot assume anything about the location of Pareto-optimal points in between two currently nondominated points. Of course interpolation is prohibited, though we should be aware that this is done automatically by a human eye confronted with an algorithmic sample.

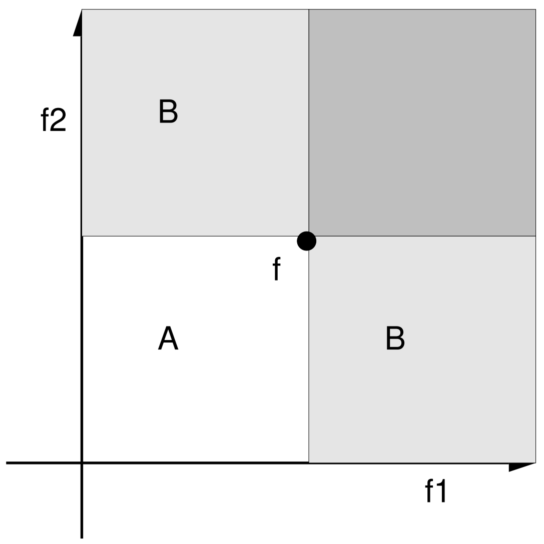

Figure 1: Dominated, dominating and incomparable points in objective space

The knowledge of a single vector in objective space $\mathcal{F}$ just gives us information about which objective vectors are dominated by f (darker grey), which vectors would dominate f (A) and which vectors would be incomparable to f (B, see figure 1). As we do not know from this sample, if there are any vectors in $\mathcal{F} = f(\mathcal{X})$ that belong to the latter two groups, the only piece of information that really helps is the one mentioned first.

Let $D(y) = \{y' \in \mathcal{F} : y \prec y'\}$ be the set of objective vectors that are dominated by an individual y (dominated set). Then we refer to the union of all dominated sets of the individuals within a population P as the global dominated set:

$$D_g = \bigcup_{y \in P} D(y) \tag{1}$$

Thus gain of information can be measured by the size of the global dominated set. Measuring the size of the global dominated set as a performance criterion was first proposed by [Zitzler and Thiele, 1998].

Fonseca and Fleming introduced a related concept called "attainment surface" which can be regarded as the "family of tightest goal vectors" known to be attainable given the algorithmically produced data [Fonseca and Fleming, 1996]. This surface constitutes the boundary of the global dominated set in objective space. In order to produce statistically sound results the authors recommend a superposition of attainment surfaces of different runs and to produce unidimensional samples by intersecting these with straight lines running diagonally to the axes.

Another study that focuses on the distribution of the samples on the Pareto set is given by Srinivas and Deb [Srinivas and Deb, 1994]. These authors measure the distribution by means of the χ^2-statistics. Naturally, this is only possible when the location of the Pareto set is known in advance.

Here, we presume only the single objective optima as being known.[1] and define the vector of there values as $y^* = (y_1^*, y_2^*, \ldots, y_m^*)^T$ and the m Pareto-optimal objective vectors with y_i^* as the i-th component as y^{*i}. Needless to say y^* will usually not be attainable, and y^{*i} constitute the outer points of the Pareto set. Starting from this and adding the information we get from every search point we define gain of information as the size of the intersection of the hypercube $H = \{y \in \mathbb{R}^q : y = y^* + \sum_{i=1}^m a_i \cdot (y^{*i} - y^*), a_i \in [0,1]\}$ and the global dominated set D_g, normalized by the size of H:

$$I := \frac{\lambda(H \cap D_g)}{\lambda(H)} \tag{2}$$

where $\lambda(A)$ denotes the Lebesgue measure of the bounded set A. This will serve as the measure of performance in the following.

2.3 Selection modes

Fonseca and Fleming classified multi-objective EAs according to the underlying fitness assignment concept. They differentiate aggregational, population-based and Pareto-based approaches. Since aggregation of the different objectives into a scalar surrogate function implies a search for a distinct optimal point rather than for the whole set of efficient points, only the latter two classes will be considered in this study. The conceptional difference is whether or not the notion of dominance is considered when assigning (scalar) fitness values and comparing individuals.

Population-based approaches consider all objectives as different selection criteria of the environment that will affect the population either simultaneously or iteratively. This can be implemented as filling different parts of the next generation or the mating pool by individuals according to the different components of their respective objective function values (like in VEGA) or by deleting fractions of the next generation according to different criteria (like in [Kursawe, 1990] or [Laumanns et al., 1998]).

Pareto-based approaches can further be divided into those which classify the whole population at once by applying a specific fitness assignment strategy based on the notion of either dominance level or dominance grade and in those which don't. The dominance grade of an individual simply is a function of the number of individuals that dominate it and is applied by the strategies of [Fonseca and Fleming, 1993]. On the other hand, if one successively removes the subsets of non-dominated individuals from the populations, the stage at which an individual becomes non-dominated constitutes the dominance level it belongs to. This method was proposed by Goldberg

[1] At last one could run a preliminary single objective EA on every component of the objective function.

[Goldberg, 1989]. As an example for the latter class, tournament selection can be applied without global fitness assignment. In that case, like in the Pareto GA proposed by [J. Horn and Goldberg, 1994], individuals are compared mutually using the dominance relation.

2.4 Diversity

In single-objective optimization preservation of diversity during the course of evolution is necessary to prevent premature convergence. Methods range from modifications of the selection operator in panmictic populations, such as niching by fitness sharing or crowding, to structuring populations with local interactions of individuals in order to receive genetic diversity by isolation-by-distance.

In the multi-objective case the issue of diversity is raised by another dimension. Diverse single solutions need to be maintained until the algorithm terminates to get a good approximation of the Pareto set. Current approaches to preserve diversity during the evolution are based on methods developed for the single-objective case. Another concept to keep track of the whole Pareto set is to store all non-dominated points in an continuously updated archive. These stored representatives may or may not take part in the further evolution. Clustering algorithms can further be applied to reduce the size of the set of non-dominated solutions.

In this study it will be investigated how the most common methods for maintaining diversity behave when they are applied to a multi-objective problem and how they are influenced by the underlying selection mode.

3 Perfomance assessment

Previous studies on the performance of multi-objective EAs always considered the final output or the results of an algorithm. This implies defining an appropriate termination criterion that should take different convergence velocities into account. Furthermore there is no insight into the way these results are produced and whether an algorithm could have been stopped earlier with nearly no loss of quality or should have run for longer.

In this study the on-line performance of different approaches will be examined, i. e., the evolution of (non-dominated) search points over time. Thereby it should be possible to characterize the behavior of different approaches in a more detailed manner. In addition there will be no archive of non-dominated solutions, because an algorithm's ability to maintain them itself is one major issue of interest. This may be justified though some algorithms really depend on storing all non-dominated points externally, i. e., not as members of the current population. However, an archive can easily be supplied to any algorithm.

3.1 Choice of representative algorithms

The test candidates for this comparative evaluation should rather reflect the principles of a method than algorithmic details. To guarantee a fair comparison and an isolated evaluation of the fitness assignment and selection principles, the algorithms differ only in this respect. Representation of solutions, mutation, and recombination will remain identical.

For our purpose methods are classified along two dimensions (see table 1). The first one is the familiar distinction between Pareto-based and non-Pareto population-based selection, while the second refers to the manner diversity is enhanced or to the niching principle. Here, we restrict ourselves to the most frequently used fitness sharing concept for panmictic populations in contrast to geographical EAs with structured populations, which have been extended to multi-objective problems during the last three years. In the following the choice of appropriate algorithms to represent each class is motivated and described.

Population — Selection	Pareto	non-Pareto
panmictic	PP	PN
structured	SP	SN

Table 1: Two-dimensional classification of selection methods

3.1.1 Pareto-based selection with fitness sharing

The representative for Pareto-based selection with fitness sharing is derived from the MOGA presented in [Fonseca and Fleming, 1993]. Fitness is assigned using the dominance grade plus the sharing function value, in which a normalized fitness of 1 is shared in objective space. This guarantees that after sharing the fitness an individual of grade n is always better than the fitness of a grade $n + 1$ individual which implies that in the case of dominance the dominating individual's fitness will always be better than the dominated one's. The final fitness values are then used to perform truncation selection.

3.1.2 Population-based non-Pareto selection in panmictic populations

This class will be represented by a simplified version of Kursawe's variant of Evolution Strategies for vector optimization [Kursawe, 1990]. Here, truncation selection is applied in several steps. Each step consists of ranking the population according to a (randomly chosen) objective, and removing the worst individuals from the population until the final size of the mating pool for creating the next generation is reached.

3.1.3 Pareto-based selection in a structured population

Poloni [Poloni, 1995] has been the first one to use structured populations for multi-objective problems to maintain diversity in a more natural way. In his approach he applied local tournament selection to a population where individuals are placed on a two-dimensional torus. Specifically, each individual produces its offspring by recombination with the best individual from a given neighborhood according to the dominance relation. If multiple non-dominated individuals exists, one is chosen at random.

3.1.4 Non-Pareto selection in a structured population

Another approach using a spatially distributed population is the predator-prey-model proposed in [Laumanns et al., 1998]. Here, selection is performed locally by predator individuals, which move across the population according to a random walk. Each predator corresponds to one of the objectives and deletes the individual from its neighborhood which fulfills this criterion worst. Empty spaces are then refilled by recombination of adjacent individuals.

3.2 Choice of test problems

Though the choice of test problems is always arbitrary we tried again to cover certain groups by specific functions. In contrast to the case study reported in [**?**], where discrete constrained optimization problems have been examined, we focus on continuous, real valued, and unconstrained problems. Since the influence of the convexity of the trade-off surface on the optimizer has been subject to many discussions, we included problems with convex (F1) and concave (F2 [Quagliarella and Vicini, 1997]) trade-off surfaces. A more sophisticated problem (F3 [Kursawe, 1990]) even exhibits convex and concave parts in the trade-off surface as well as an isolated efficient point. In detail the test functions are defined as follows, for all compenents minimization is assumed:

$$F1(x) = \begin{pmatrix} (x)^2 \\ (x-c)^2 \end{pmatrix}, \qquad c = (2, 0, 0, \dots, 0)^T \tag{3}$$

$$F2(x) = \begin{pmatrix} (\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i) + 10))^{\frac{1}{4}} \\ (\frac{1}{n}\sum_{i=1}^{n}((x_i - 1.5)^2 - 10\cos(2\pi(x_i - 1.5)) + 10))^{\frac{1}{4}} \end{pmatrix} \tag{4}$$

$$F3(x) = \begin{pmatrix} \sum_{i=1}^{n-1}(-10e^{-0.2\sqrt{x_i^2 + x_{i+1}^2}}) \\ \sum_{i=1}^{n}(|x_i|^{0.8} + \sin(x_i)^3) \end{pmatrix} \tag{5}$$

3.3 Representation and parameter setting

Since all test problems are functions of 20 variables each individual is represented by a vector of 20 floating point numbers. Mutation is carried out by adding a normal-distributed random number to each of the decision variables with zero mean and standard deviation $\sigma = 0.05$, which remains fixed during the evolution. Offspring are produced by recombining two of the possible parent individuals in each turn using discrete recombination on the decision variables.

The total population size p should be identical for all algorithms. In the case of the structured population a two-dimentional torus is used with $\sqrt{p} \times \sqrt{p}$ individuals. For the non-structured populations truncation selection will lead to a number of $\mu = \frac{1}{5}p$ possible parents for the next generation. In the predator-prey model $pred$ individuals will be replaced in each iteration, which corresponds to the number of predators and should therefore be a multiple of m to assure equal importance to each objective. For the tournament selection in SP all individuals found during a s-step random walk starting from the first (fixed) parent's position will be considered, while in this case $s = \frac{1}{2}\sqrt{p}$ is chosen. Finally the niche radius for PP will be calculated for each generation using the formula proposed in [Fonseca and Fleming, 1993].

4 Test results

For all test runs the individuals have been initialized using the uniform distribution on $[-5.12, 5.12]^{20}$. The runs have been terminated after $102,400$ function evaluations in the case of $p = 1024$ and $20,000$ evaluations in the case of $p = 100$. Figure 2 displays the results for the measure I as the average values of 10 independent runs for each function and model.

Figure 3 shows the final populations for all models in the case of $F3$.

4.1 Discussion

The first obvious observation that can be conducted from the plots in figure 2 is the fast gain of information of the Pareto-based selection in the panmictic population (PP) in all three test problems. In the beginning PN shows similar behavior, but begins to stagnate very early to be overtaken by the structured populations in each case. This is certainly due to the loss of diversity as soon as the vicinity of the trade-off surface is reached.

The structured populations both reveal slower gain of information, and are able to reach approximately 90 per cent of PP's maximal value at last. Surprizingly, the final state of non-Pareto selection seems to be slightly better than the (local) Pareto-based selection in general. While on F1 and F2 PPs's performance is monotonously increasing, it reaches its peak on F3 at about 20,000 function evaluations and gradually falls to 70 per cent of its maximum afterwards. On the concave problem F3 only PPs it able to maintain its maximal value, while PP without sharing again loses its performance. In these cases the structured populations maintain diversity for longer and manage to overtake PP on F2 as well as both PP and PPs on F3.

On F1 no significant differences can be observed betweenn PP and PPs. On F2 PPs performs slightly better after the maximum is reached, but on F3 fitness sharing seems to be crucial not to lose performance. This leads us to the assumption that only on concave problems, where it is very difficult to get into the "corners" of the trade-off surface, niching by sharing is of importance. In the other cases the fitness assignment strategy based on the dominance grade turns out to be sufficient to guide the search from densely populated areas to those represented by only few individuals.

The test runs with $p = 100$ show that the differences between structured and non-sturctured population models diminish for small population sizes. This is certainly due to the fact that isolation-by-distance does not work properly for a small diameter of the underlying spatial structure.

All in all those models using structured populations generally evolve slower concerning the quality measure applied in this study and, especially the predator-prey selection method, need large population sizes. However, spatially distributed models are usually better suited for parallel computation which could in turn lead to a further improvement with respect to the computation time used in contrast to the panmictic models.

5 Conclusions

The application of a dynamic performance measure provides insight into the behavior of different selection schemes in multi-objective evolutionary algorithms. It has been shown empirically that Pareto-based selection usually exhibits fast convergence towards the vicinity of the Pareto set, and that a fitness assignment strategy based on the dominance grade leads to a good distribution of individuals on convex functions. On more difficult functions, however, further enhancement of the selection operator is necessary to maintain diversity, which can partially be achieved by fitness sharing. Spatially distributed populations are able to maintain diversity for a longer timespan, but generally evolve slower than their panmictic counterparts.

These investigations may motivate the analysis of the underlying principles of the different fitness assignment and selection methods. This could, in return, lead to the constuction of advanced operators that provide both convergence velocity and maintainence of diversity.

Further research should include the influence of mutation and recombination in the case of multiple objectives. Especially adaptive operators, like the self-adaptation of mutation rates, should be of interest.

Acknowledgments

This work is a result of the *Collaborative Research Center "Computational Intelligence" (SFB 531)* supported by the German Research Foundation (DFG).

References

[Fonseca and Fleming, 1993] Fonseca, C. M. and Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In Forrest, S., editor, *Int'l Conf. Genetic Algorithms (ICGA'93)*, pages 2–8, University of Illinois at Urbana-Champaign IL. Morgan Kaufmann, San Mateo CA.

[Fonseca and Fleming, 1995] Fonseca, C. M. and Fleming, P. J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation, 3(1)*, pages 1–16.

[Fonseca and Fleming, 1996] Fonseca, C. M. and Fleming, P. J. (1996). On the performance assessment and comparison of stochastic multiobjective optimizers. In Goos, G., Hartmanis, J., and van Leeuwen, J., editors, *Parallel Problem Solving from Nature - PPSN IV*, pages 584–593, Berlin. Springer.

[Fonseca and Fleming, 1997] Fonseca, C. M. and Fleming, P. J. (1997). Multiobjective optimization. In B"ack, T., Fogel, D. B., and Michalewicz, Z., editors, *Handbook of Evolutionary Computation*, pages C4.5:1–9, Bristol, UK. IOP Publishing and Oxford Univertity Press.

[Goldberg, 1989] Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley Publishing Company,Inc., Reading.

[J. Horn and Goldberg, 1994] J. Horn, N. N. and Goldberg, D. E. (1994). A niched pareto genetic algorithm for multiobjective optimization. In Michalewicz, Z., Schaffer, J. D., Schwefel, H.-P., Fogel, D. B., and Kitano, H., editors, *Proc. First IEEE Conf. Evolutionary Computation, IEEE World Congress on Computational Intelligence*, volume 2, pages 82–89, Orlando FL. IEEE Press, Piscataway NJ.

[Kursawe, 1990] Kursawe, F. (1990). A variant of evolution strategies for vector optimization. In Schwefel, H.-P. and Männer, R., editors, *Parallel Problem Solving from Nature – Proc. First Workshop PPSN*, pages 193–197, University of Dortmund. Springer, Berlin, 1991.

[Laumanns et al., 1998] Laumanns, M., Rudolph, G., and Schwefel, H.-P. (1998). A spatial predator-prey approach to multi-objective optimization. In Eiben, A. E., Bäck, T., Schoenauer, M., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature – PPSN V, Fifth Int'l Conf., Amsterdam, The Netherlands, September 27–30, 1998, Proc.* Springer, Berlin.

[Pareto, 1972] Pareto, V. (1972). *Manuale di economia politica (engl.)*. Macmillan, London.

[Poloni, 1995] Poloni, C. (1995). Hybrid GA for multi objective aerodynamic shape optimization. In Winter, G., Périaux, J., Galán, M., and Cuesta, P., editors, *Genetic Algorithms in Engineering and Computer Science, Proc. First Short Course EUROGEN'95*, pages 397–415, Las Palmas de Gran Canaria, Spain. Wiley, New York.

[Quagliarella and Vicini, 1997] Quagliarella, D. and Vicini, A. (1997). Coupling genetic algorithms and gradient based optimization techniques. In Quagliarella, D., Périaux, J., Poloni, C., and Winter, G., editors, *Genetic Algorithms and Evolution Strategy in Engineering and Computer Science — Recent advances and industrial applications*, pages 289–309. Wiley, Chichester.

[Schaffer, 1985] Schaffer, D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In Grefenstette, J., editor, *Proceedings of an International Conference on Genetic Algorithms and their Applications, Pittsburgh*, pages 93–100.

[Srinivas and Deb, 1994] Srinivas, N. and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation, 2(3)*, pages 221–248.

[Zitzler and Thiele, 1998] Zitzler, E. and Thiele, L. (1998). Multiobjective optimization using evolutionary algorithms – a comparative case study. In Eiben, A. E., Bäck, T., Schoenauer, M., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature – PPSN V, Fifth Int'l Conf., Amsterdam, The Netherlands, September 27–30, 1998, Proc.* Springer, Berlin.

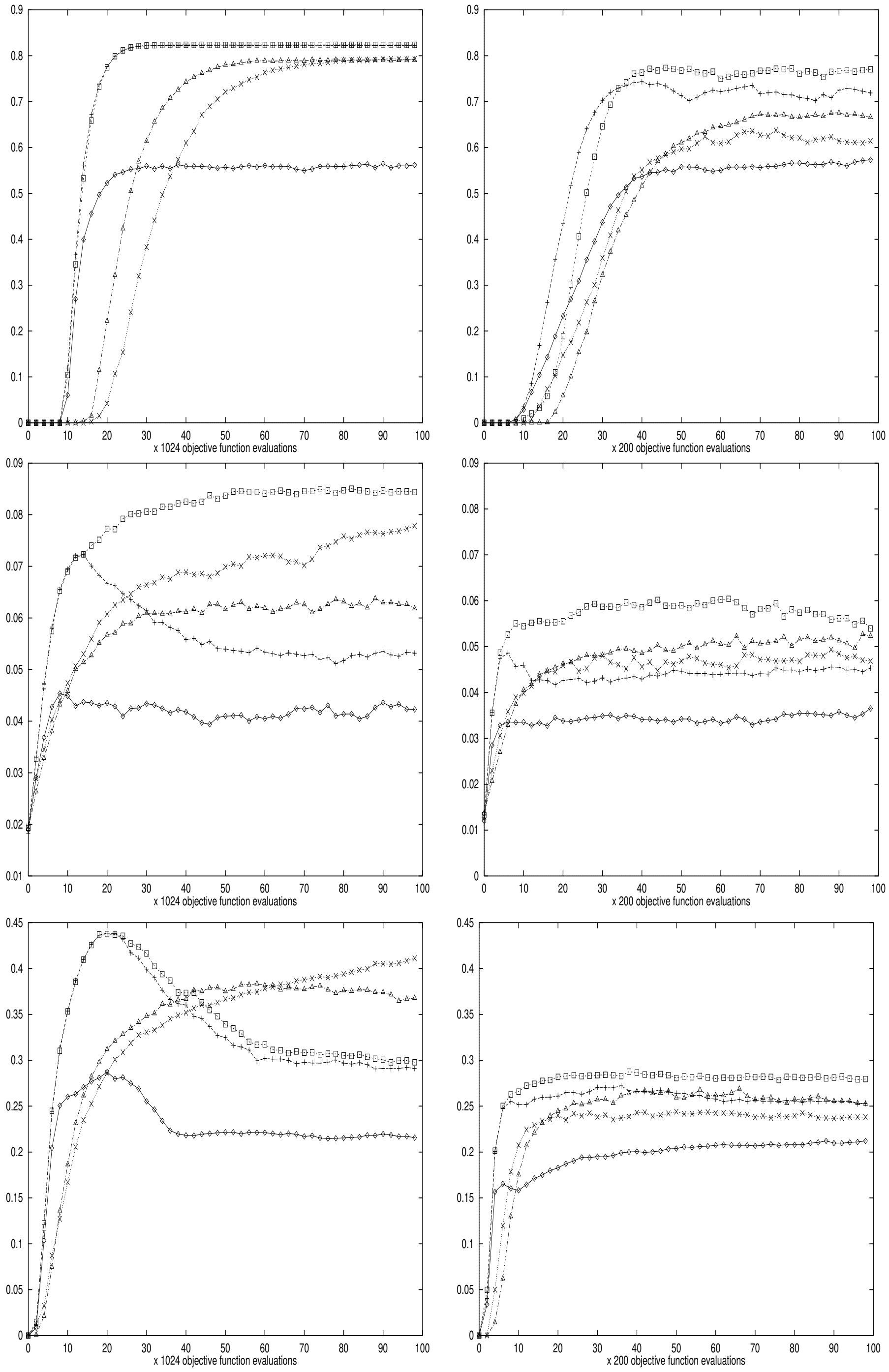

Figure 2: Development of the measure I during the evolution, displayed on the number of function evaluations, in the case of $p = 1024$ (left) and $p = 100$ (right). The first row shows F1, the second F2, and the last F3. The marker $\square$ corresponds to PPs, $+$ to PP, $\diamond$ to PN, $\triangle$ to SP, and $\times$ to SN.

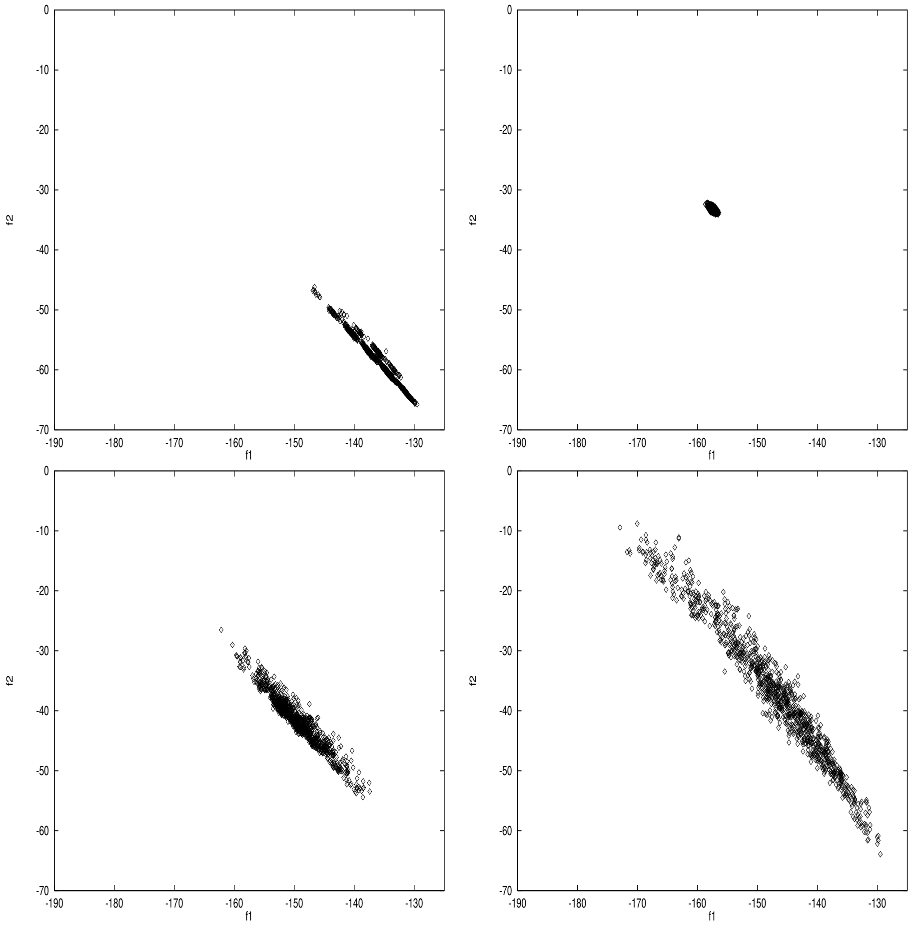

Figure 3: Population after $100,000$ function evaluations of F3 using PP (top left), PN (top right), SP (bottom left) and SN (bottom right)