Evolutionary Multiobjective Optimization: Past, Present and Future

Carlos A. Coello Coello

CINVESTAV-IPN

Depto. de Ingeniería Eléctrica Sección de Computación

Av. Instituto Politécnico Nacional No. 2508

Col. San Pedro Zacatenco

México, D. F. 07300, MEXICO

ccoello@cs.cinvestav.mx

Motivation

Most problems in nature have several (possibly conflicting) objectives to be satisfied. Many of these problems are frequently treated as single-objective optimization problems by transforming all but one objective into constraints.

What is a multiobjective optimization problem?

The Multiobjective Optimization Problem (MOP) (also called multicriteria optimization, multiperformance or vector optimization problem) can be defined (in words) as the problem of finding (Osyczka, 1985):

a vector of decision variables which satisfies constraints and optimizes a vector function whose elements represent the objective functions. These functions form a mathematical description of performance criteria which are usually in conflict with each other. Hence, the term "optimize" means finding such a solution which would give the values of all the objective functions acceptable to the decision maker.

A Formal Definition

The general Multiobjective Optimization Problem (MOP) can be formally defined as:

Find the vector $\vec{x}^* = [x_1^*, x_2^*, \dots, x_n^*]^T$ which will satisfy the m inequality constraints:

$$g_i(\vec{x}) \ge 0 \qquad i = 1, 2, \dots, m \tag{1}$$

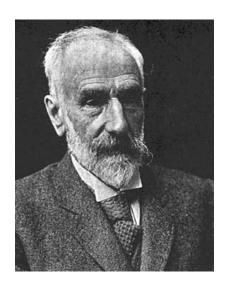
the p equality constraints

$$h_i(\vec{x}) = 0 \quad i = 1, 2, \dots, p$$
 (2)

and will optimize the vector function

$$\vec{f}(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), \dots, f_k(\vec{x})]^T$$
 (3)

What is the notion of optimum in multiobjective optimization?



Having several objective functions, the notion of "optimum" changes, because in MOPs, we are really trying to find good compromises (or "trade-offs") rather than a single solution as in global optimization. The notion of "optimum" that is most commonly adopted is that originally proposed by Francis Ysidro Edgeworth in 1881.

What is the notion of optimum in multiobjective optimization?



This notion was later generalized by Vilfredo Pareto (in 1896). Although some authors call *Edgeworth-Pareto optimum* to this notion, we will use the most commonly accepted term: *Pareto optimum*.

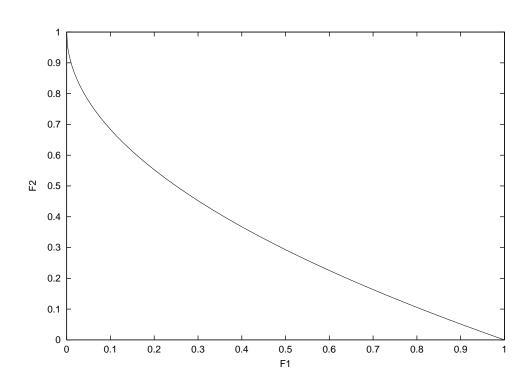
Definition of Pareto Optimality

We say that a vector of decision variables $\vec{x}^* \in \mathcal{F}$ is Pareto optimal if there does not exist another $\vec{x} \in \mathcal{F}$ such that $f_i(\vec{x}) \leq f_i(\vec{x}^*)$ for all i = 1, ..., k and $f_j(\vec{x}) < f_j(\vec{x}^*)$ for at least one j.

Definition of Pareto Optimality

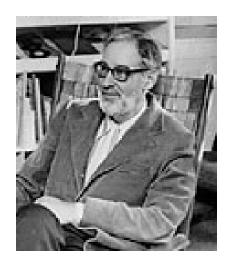
In words, this definition says that \vec{x}^* is Pareto optimal if there exists no feasible vector of decision variables $\vec{x} \in \mathcal{F}$ which would decrease some criterion without causing a simultaneous increase in at least one other criterion. Unfortunately, this concept almost always gives not a single solution, but rather a set of solutions called the *Pareto optimal set*. The vectors \vec{x}^* correspoding to the solutions included in the Pareto optimal set are called nondominated. The plot of the objective functions whose nondominated vectors are in the Pareto optimal set is called the Pareto front.

Sample Pareto Front





As early as 1944, John von Neumann and Oskar Morgenstern mentioned that an optimization problem in the context of a social exchange economy was "a peculiar and disconcerting mixture of several conflicting problems" that was "nowhere dealt with in classical mathematics".



In 1951 Tjalling C. Koopmans edited a book called *Activity Analysis of Production and Allocation*, where the concept of "efficient" vector was first used in a significant way.

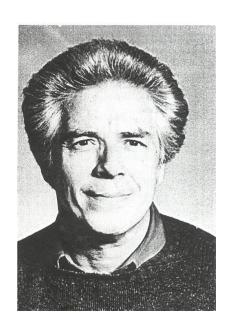


The origins of the mathematical foundations of multiobjective optimization can be traced back to the period that goes from 1895 to 1906. During that period, Georg Cantor and Felix Hausdorff laid the foundations of infinite dimensional ordered spaces.

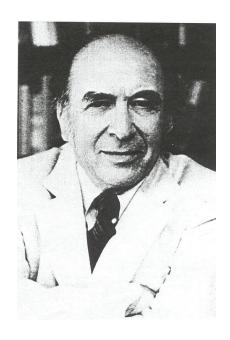


Cantor also introduced equivalence classes and stated the first sufficient conditions for the existence of a utility function.

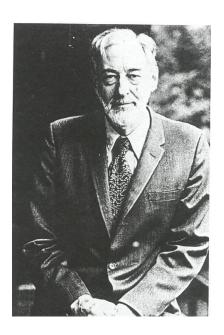
Hausdorff also gave the first example of a complete ordering.



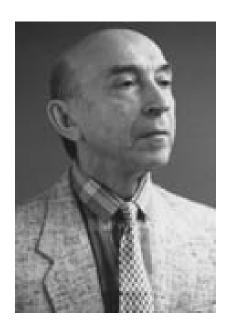
However, it was the concept of *vector maximum problem* introduced by Harold W. Kuhn and Albert W. Tucker (1951) which made multiobjective optimization a mathematical discipline on its own.



However, multiobjective optimization theory remained relatively undeveloped during the 1950s. It was until the 1960s that the foundations of multiobjective optimization were consolidated and taken seriously by pure mathematicians when Leonid Hurwicz generalized the results of Kuhn & Tucker to topological vector spaces.



The application of multiobjective optimization to domains outside economics began with the work by Koopmans (1951) in production theory and with the work of Marglin (1967) in water resources planning.



The first engineering application reported in the literature was a paper by Zadeh in the early 1960s. However, the use of multiobjective optimization became generalized until the 1970s.

Current State of the Area



Currently, there are over 30 mathematical programming techniques for multiobjective optimization. However, these techniques tend to generate elements of the Pareto optimal set one at a time. Additionally, most of them are very sensitive to the shape of the Pareto front (e.g., they do not work when the Pareto front is concave or when the front is disconnected).

Why Evolutionary Algorithms?

Evolutionary algorithms seem particularly suitable to solve multiobjective optimization problems, because they deal simultaneously with a set of possible solutions (the so-called population). This allows us to find several members of the Pareto optimal set in a single run of the algorithm, instead of having to perform a series of separate runs as in the case of the traditional mathematical programming techniques. Additionally, evolutionary algorithms are less susceptible to the shape or continuity of the Pareto front (e.g., they can easily deal with discontinuous or concave Pareto fronts), whereas these two issues are a real concern for mathematical programming techniques.

How popular are evolutionary algorithms in multiobjective optimization?

The potential of evolutionary algorithms in multiobjective optimization was hinted by Rosenberg in the 1960s, but the first actual implementation was produced in the mid-1980s (Schaffer, 1984). During ten years, the field remain practically inactive, but it started growing in the mid-1990s, in which several techniques and applications were developed.

How popular are evolutionary algorithms in multiobjective optimization?

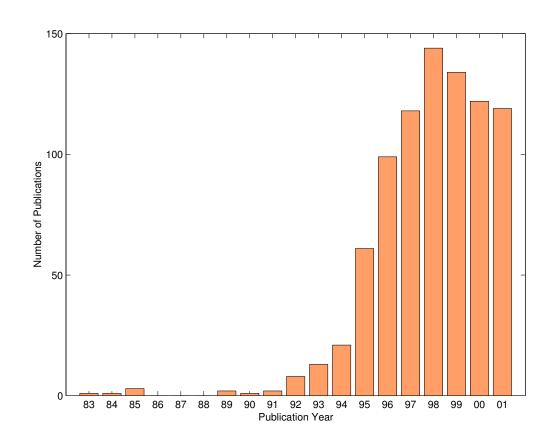


Figure 1: Citations by Year (up to year 2001)

Classifying Techniques

We will use the following simple classification of Evolutionary Multi-Objective Optimization (EMOO) approaches:

- First Generation Techniques
 - Non-Pareto approaches
 - Pareto-based approaches
- Second Generation Techniques

First Generation Techniques

Non-Pareto Techniques include the following:

- Aggregating approaches
- VEGA
- Lexicographic ordering
- The ε -constraint Method
- Target-vector approaches

First Generation Techniques

Pareto-based Techniques include the following:

- Pure Pareto ranking
- MOGA
- NSGA
- NPGA and NPGA2

Second Generation Techniques

- PAES, PESA and PESA II
- SPEA and SPEA 2
- NSGA-II
- MOMGA and MOMGA-II
- micro-GA

Non-Pareto Techniques

- Approaches that do not incorporate directly the concept of Pareto optimum.
- Incapable of producing certain portions of the Pareto front.
- Efficient and easy to implement, but appropriate to handle only a few objectives.

Aggregation Functions

- These techniques are called "aggregating functions" because they combine (or "aggregate") all the objectives into a single one. We can use addition, multiplication or any other combination of arithmetical operations.
- Oldest mathematical programming method, since aggregating functions can be derived from the Kuhn-Tucker conditions for nondominated solutions.

Aggregation Functions

An example of this approach is a sum of weights of the form:

$$\min \sum_{i=1}^{k} w_i f_i(\vec{x}) \tag{4}$$

where $w_i \geq 0$ are the weighting coefficients representing the relative importance of the k objective functions of our problem. It is usually assumed that

$$\sum_{i=1}^{k} w_i = 1 \tag{5}$$

Advantages and Disadvantages

- Easy to implement
- Efficient
- Linear combinations of weights do not work when the Pareto front is concave, regardless of the weights used (Das, 1997).

 Note however, that the weights can be generated in such a way the Pareto front is rotated (Jin et al., 2001). In this last case, concave Pareto fronts can be efficiently generated.

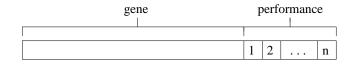
Sample Applications

- Design of DSP systems (Arslan, 1996).
- Water quality control (Garrett, 1999).
- System-level synthesis (Blickle, 1996).
- Design of optical filters for lamps (Eklund, 1999).

Vector Evaluated Genetic Algorithm (VEGA)

- Proposed by Schaffer in the mid-1980s (1984,1985).
- It uses subpopulations that optimize each objective separately. The concept of Pareto optimum is not directly incorporated into the selection mechanism of the GA.

VEGA



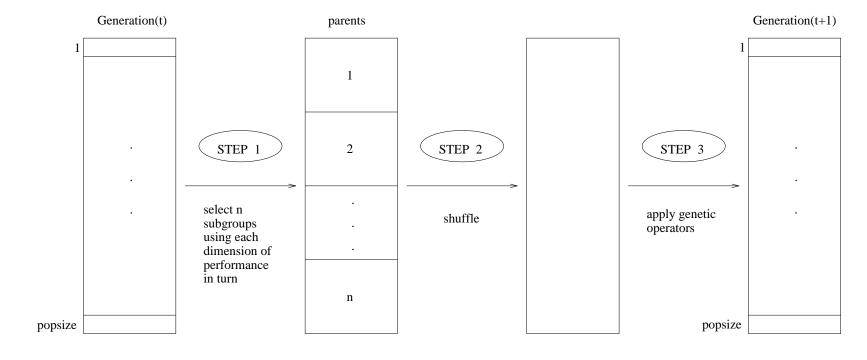


Figure 2: Schematic of VEGA selection.

Advantages and Disadvantages

- Efficient and easy to implement.
- It doesn't have an explicit mechanism to maintain diversity. It doesn't necessarily produce nondominated vectors.

Sample Applications

- Aerodynamic optimization (Rogers, 2000).
- Combinational circuit design at the gate-level (Coello, 2000).
- Design multiplierless IIR filters (Wilson, 1993).
- Groundwater pollution containment (Ritzel, 1994).

Lexicographic Ordering

In this method, the user is asked to rank the objectives in order of importance. The optimum solution is then obtained by minimizing the objective functions, starting with the most important one and proceeding according to the assigned order of importance of the objectives.

It is also possible to select randomly a single objective to optimize at each run of a GA.

Advantages and Disadvantages

- Efficient and easy to implement.
- Requires a pre-defined ordering of objectives and its performance will be affected by it.
- Inappropriate when there is a large number of objectives.

- Symbolic layout compaction (Fourman, 1985).
- Robot path planning (Gacôgne, 1999).
- Personnel scheduling (El Moudani et al., 2001).

The ε -Constraint Method

This method is based on minimization of one (the most preferred or primary) objective function, and considering the other objectives as constraints bound by some allowable levels ε_i . Hence, a single objective minimization is carried out for the most relevant objective function subject to additional constraints on the other objective functions. The levels ε_i are then altered to generate the entire Pareto optimal set.

Advantages and Disadvantages

- Easy to implement.
- Potentially high computational cost (many runs may be required).

- Preliminary design of a marine vehicle (Lee, 1997).
- Groundwater pollution containment problems (Ranjithan, 1992).
- Fault tolerant system design (Schott, 1995).

Target-Vector Approaches

- Definition of a set of goals (or targets) that we wish to achieve for each objective function. EA minimizes differences between the current solution and these goals.
- Can also be considered aggregating approaches, but in this case, concave portions of the Pareto front could be obtained.
- Examples: Goal Programming, Goal Attainment, min-max method.

Advantages and Disadvantages

- Efficient and easy to implement.
- Definition of goals may be difficult in some cases and may imply an extra computational cost.
- Some of them (e.g., goal attainment) may introduce a misleading selection pressure under certain circumstances.
- Goals must lie in the feasible region so that the solutions generated are members of the Pareto optimal set.

- Intensities of emission lines of trace elements (Wienke, 1992).
- Optimization of a fishery bioeconomic model (Mardle et al., 2000).
- Optimization of the counterweight balancing of a robot arm (Coello, 1998).

Pareto-based Techniques

- Suggested by Goldberg (1989) to solve the problems with Schaffer's VEGA.
- Use of nondominated ranking and selection to move the population towards the Pareto front.
- Requires a ranking procedure and a technique to maintain diversity in the population (otherwise, the GA will tend to converge to a single solution, because of the stochastic noise involved in the process).

Pure Pareto Ranking

Although several variations of Goldberg's proposal have been proposed in the literature (see the following subsections), several authors have used what we call "pure Pareto ranking". The idea in this case is to follow Goldberg's proposal as stated in his book (1989).

Advantages and Disadvantages

- Relatively easy to implement.
- Problem to scale the approach, because checking for nondominance is $O(kM^2)$, where k is the amount of objectives and M is the population size.
- Fitness sharing is $O(M^2)$.
- The approach is less susceptible to the shape or continuity of the Pareto front.

- Water distribution network (Halhal et al., 1997)
- Design of DSP systems (Bright, 1999).
- Design of electromagnetic devices (Weile et al., 1996).
- Optimal planning of an electrical power distribution system (Ramírez, 2001).

Multi-Objective Genetic Algorithm (MOGA)

- Proposed by Fonseca and Fleming (1993).
- The approach consists of a scheme in which the rank of a certain individual corresponds to the number of individuals in the current population by which it is dominated.
- It uses fitness sharing and mating restrictions.

Advantages and Disadvantages

- Efficient and relatively easy to implement.
- Its performance depends on the appropriate selection of the sharing factor.
- MOGA was the most popular first-generation MOEA and it normally outperformed all of its contemporary competitors.

Some Applications

- Fault diagnosis (Marcu, 1997).
- Control system design (Chipperfield 1995; Whidborne, 1995; Duarte, 2000).
- Design of antennas (Thompson, 2001).
- System-level synthesis (Dick, 1998).

Nondominated Sorting Genetic Algorithm (NSGA)

- Proposed by Srinivas and Deb (1994).
- It is based on several layers of classifications of the individuals. Nondominated individuals get a certain dummy fitness value and then are removed from the population. The process is repeated until the entire population has been classified.
- To maintain the diversity of the population, classified individuals are shared (in decision variable space) with their dummy fitness values.

NSGA

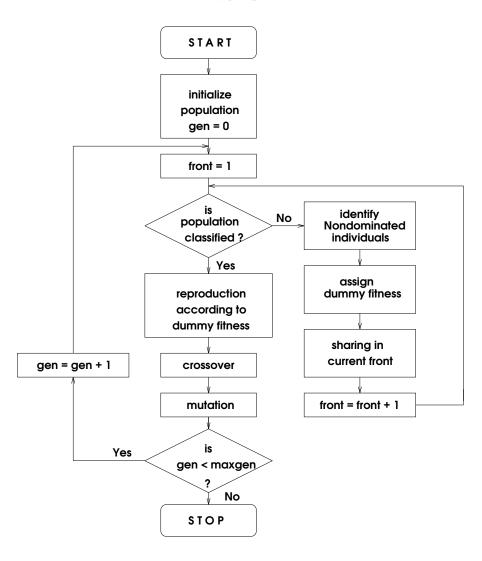


Figure 3: Flowchart of the Nondominated Sorting Genetic Algorithm (NSGA).

Advantages and Disadvantages

- Relatively easy to implement.
- Seems to be very sensitive to the value of the sharing factor.

- Water quality control (Reed et al., 2001).
- Design of control systems (Blumel, 2001).
- Constellation design (Mason, 1999).
- Computational fluid dynamics (Marco, 1999).

Niched-Pareto Genetic Algorithm (NPGA)

- Proposed by Horn et al. (1993,1994).
- It uses a tournament selection scheme based on Pareto dominance. Two individuals randomly chosen are compared against a subset from the entire population (typically, around 10% of the population). When both competitors are either dominated or nondominated (i.e., when there is a tie), the result of the tournament is decided through fitness sharing in the objective domain (a technique called equivalent class sharing was used in this case).

Advantages and Disadvantages

- Easy to implement.
- Efficient because does not apply Pareto ranking to the entire population.
- It seems to have a good overall performance.
- Besides requiring a sharing factor, it requires another parameter (tournament size).

- Analysis of experimental spectra (Golovkin, 2000).
- Feature selection (Emmanouilidis, 2000).
- Fault-tolerant systems design (Schott, 1995).
- Road systems design (Haastrup & Pereira, 1997).

NPGA 2

Erickson et al. (2001) proposed the NPGA 2, which uses Pareto ranking but keeps tournament selection (solving ties through fitness sharing as in the original NPGA).

Niche counts in the NPGA 2 are calculated using individuals in the partially filled next generation, rather than using the current generation. This is called continuously updated fitness sharing, and was proposed by Oei et al. (1991).

• Design of groundwater remediation systems (Erickson et al., 2001).

Second Generation Techniques

The second generation in evolutionary multiobjective optimization emphasizes efficiency. Secondary (external) populations become common and clever ways are devised to generate solutions that are both nondominated and uniformly distributed.

Ways around two problems: fitness sharing is $O(N^2)$, and checking for Pareto dominance is $O(kM^2)$ per generation.

The Pareto Archived Evolution Strategy (PAES)

PAES was introduced by Knowles & Corne (2000).

It uses a (1+1) evolution strategy together with an external archive that records all the nondominated vectors previously found.

It uses an adaptive grid to maintain diversity.

- Off-line routing problem (Knowles, 1999)
- Adaptive distributed database management problem (Knowles, 2000)

The Pareto Envelope-based Selection Algorithm (PESA)

PESA was proposed by Corne et al. (2000). This approach uses a small internal population and a larger external (or secondary) population. PESA uses the same hyper-grid division of phenotype (i.e., objective funcion) space adopted by PAES to maintain diversity. However, its selection mechanism is based on the crowding measure used by the hyper-grid previously mentioned. This same crowding measure is used to decide what solutions to introduce into the external population (i.e., the archive of nondominated vectors found along the evolutionary process).

• Telecommunications problems (Corne et al., 2000).

The Pareto Envelope-based Selection Algorithm-II (PESA-II)

PESA-II (Corne et al., 2001) is a revised version of PESA in which region-based selection is adopted. In region-based selection, the unit of selection is a hyperbox rather than an individual. The procedure consists of selecting (using any of the traditional selection techniques) a hyperbox and then randomly select an individual within such hyperbox.

• Telecommunications problems (Corne et al., 2001).

The Strength Pareto Evolutionary Algorithm (SPEA)

SPEA was introduced by Zitzler & Thiele (1999).

It uses an external archive containing nondominated solutions previously found.

It computers a strength value similar to the ranking value used by MOGA.

A clustering technique called "average linkage method" is used to keep diversity.

- Exploration of trade-offs of software implementations for DSP algorithms (Zitzler, 1999)
- Treatment planning (Petrovski, 2001)
- Allocation in radiological facilities (Lahanas, 2001)

SPEA2

A revised version of SPEA has been recently proposed: SPEA2 (Zitzler, 2001). SPEA2 has three main differences with respect to its predecessor: (1) it incorporates a fine-grained fitness assignment strategy which takes into account for each individual the number of individuals that dominate it and the number of individuals by which it is dominated; (2) it uses a nearest neighbor density estimation technique which guides the search more efficiently, and (3) it has an enhanced archive truncation method that guarantees the preservation of boundary solutions.

• Control code size and reduce bloat in genetic programming (Bleuler, 2001).

NSGA-II

Deb et al. (2000,2002) proposed a new version of the Nondominated Sorting Genetic Algorithm (NSGA), called NSGA-II, which is more efficient (computationally speaking), uses elitism and a crowded comparison operator that keeps diversity without specifying any additional parameters.

- Forest management (Ducheyne, 2001).
- Shape optimization (Deb, 2001)

The Multi-Objective Messy Genetic Algorithm (MOMGA)

MOMGA was proposed by Van Veldhuizen and Lamont (2000). This is an attempt to extend the messy GA to solve multiobjective optimization problems.

MOMGA consists of three phases: (1) Initialization Phase, (2) Primordial Phase, and (3) Juxtapositional Phase. In the *Initialization Phase*, MOMGA produces all building blocks of a certain specified size through a deterministic process known as partially enumerative initialization. The *Primordial Phase* performs tournament selection on the population and reduces the population size if necessary. In the *Juxtapositional Phase*, the messy GA proceeds by building up the population through the use of the cut and splice recombination operator.

Sample Applications

- Design of controllers (Herreros, 2000).
- Traditional benchmarks (Van Veldhuizen & Lamont, 2000).

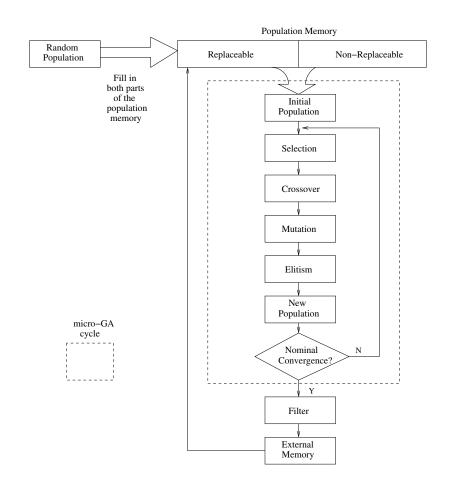
The Multi-Objective Messy Genetic Algorithm-II (MOMGA-II)

Zydallis et al. (2001) proposed MOMGA-II. In this case, the authors extended the fast-messy GA, which consists of three phases: (1) Initialization Phase, (2) Building Block Filtering, and (3) Juxtapositional Phase. Its main difference with respect to the original messy GA is in the two first phases. The *Initialization* Phase utilizes probabilistic complete initialization which creates a controlled number of building block clones of a specified size. The Building Block Filtering Phase reduces the number of building blocks through a filtering process and stores the best building blocks found. The Juxtapositional Phase is the same as in the MOMGA.



• Traditional benchmarks (Zydallis et al., 2001).

The Micro Genetic Algorithm for Multiobjective Optimization



Sample Applications

• It has been used only with traditional benchmarks (Coello, 2001)

Some Interesting Statistics

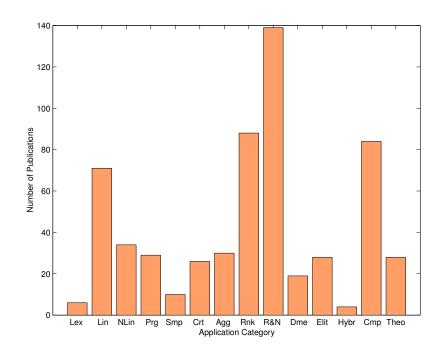


Figure 4: Lex = Lexicographic, Lin = Linear Fitness Combination, NLin = Nonlinear Fitness Combination, Prg = Progressive, Smp = Independent Sampling, Crt = Criterion Selection, Agg = Aggregation Selection, Rnk = Pareto Ranking, R&N = Pareto Rank- and Niche-Based Selection, Dme = Pareto Deme-Based Selection, Elit = Pareto Elitist-Based, Hybr = Hybrid Selection, Cmp = Technique Comparisons, Theo = Theory and Reviews.



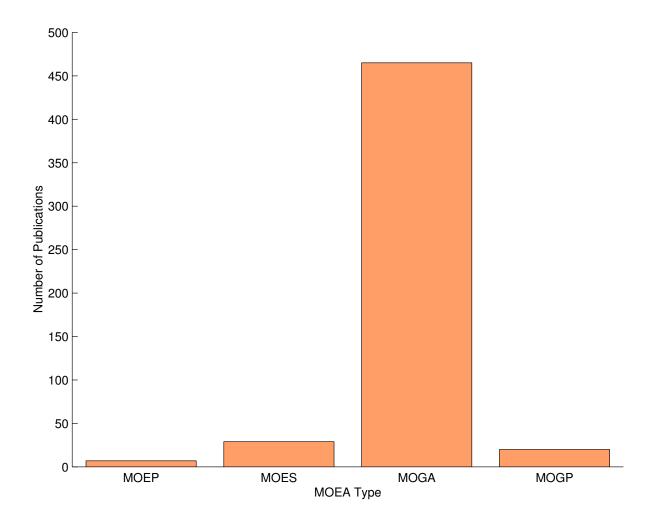


Figure 5: MOEA citations by type of technique (up to year 2001)

Some Interesting Statistics

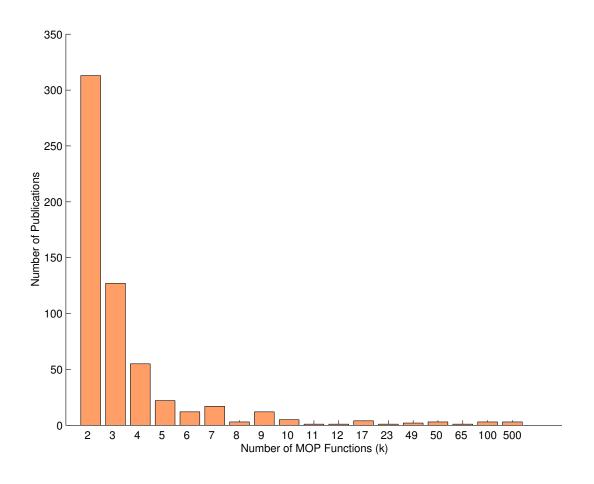


Figure 6: Number of objective functions reported in the literature (up to year 2001)

Theory

The most important theoretical work related to EMOO has concentrated on the following issues:

- Studies of convergence towards the Pareto optimum set (Rudolph, 1998, 2000, 2001; Hanne 2000, 2000a; Veldhuizen, 1998).
- Ways to compute appropriate sharing factors (or niche sizes) (Horn, 1997, Fonseca, 1993).

Theory

Much more work is needed. For example:

- To study the structure of fitness landscapes (Kaufmann, 1989) in multiobjective optimization problems.
- Convergence of parallel MOEAs.
- Theoretical limit to the number of objective functions that can be used in practice.

Theory

- Formal models of alternative heuristics used for multiobjective optimization (e.g., ant system, particle swarm optimization, etc.).
- Complexity analysis of MOEAs and running time analysis.
- Study of population dynamics for an MOEA.

Test Functions

- Good benchmarks were disregarded for many years.
- Recently, there have been several proposals to design test functions suitable to evaluate EMOO approaches.
- Constrained test functions are of particular interest.
- Multiobjective combinatorial optimization problems have also been proposed.

Test Functions

- After using biobjective test functions for a few years, test functions with three objectives and higher numbers of decision variables are becoming popular.
- What about dynamic test functions, uncertainty and real-world applications?

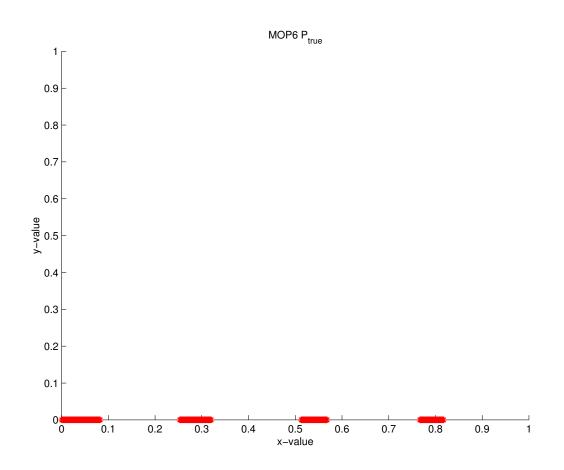


Figure 7: P_{true} of one of Deb's test functions.

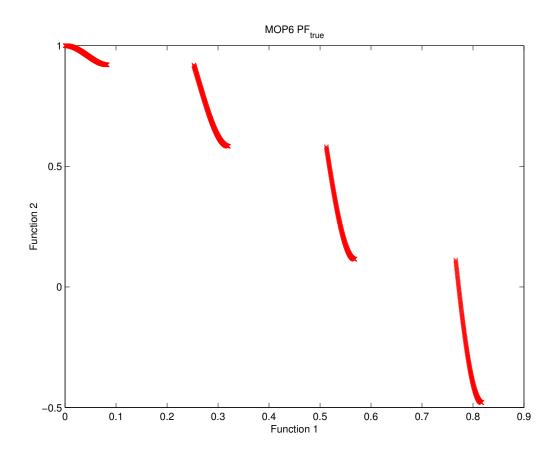


Figure 8: PF_{true} of one of Deb's test functions.

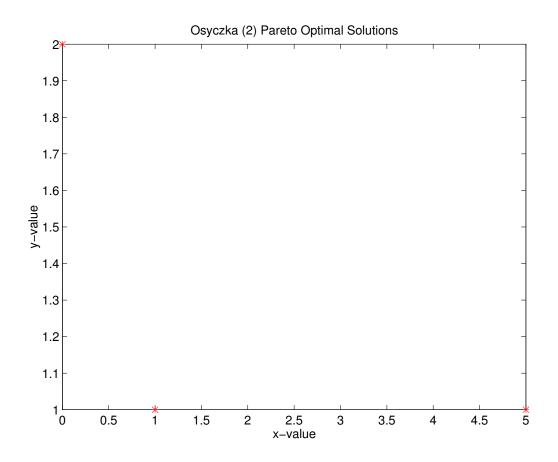


Figure 9: P_{true} of Osyczka's function.

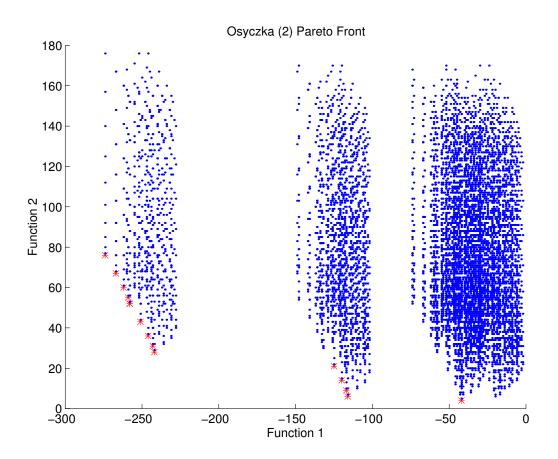


Figure 10: PF_{true} of Osyczka's function.

Metrics

Three are normally the issues to take into consideration to design a good metric in this domain (Zitzler, 2000):

- 1. Minimize the distance of the Pareto front produced by our algorithm with respect to the true Pareto front (assuming we know its location).
- 2. Maximize the spread of solutions found, so that we can have a distribution of vectors as smooth and uniform as possible.
- 3. Maximize the amount of elements of the Pareto optimal set found.

Error rate: Enumerate the entire intrinsic search space explored by an evolutionary algorithm and then compare the true Pareto front obtained against those fronts produced by any MOEA.

Obviously, this metric has some serious scalability problems.

Spread: Use of a statistical metric such as the chi-square distribution to measure "spread" along the Pareto front.

This metric assumes that we know the true Pareto front of the problem.

Attainment Surfaces: Draw a boundary in objective space that separates those points which are dominated from those which are not (this boundary is called "attainment surface").

Perform several runs and apply standard non-parametric statistical procedures to evaluate the quality of the nondominated vectors found.

It is unclear how can we really assess how much better is a certain approach with respect to others.

Generational Distance: Estimates how far is our current Pareto front from the true Pareto front of a problem using the Euclidean distance (measured in objective space) between each vector and the nearest member of the true Pareto front.

The problem with this metric is that only distance to the true Pareto front is considered and not uniform spread along the Pareto front.

Coverage: Measure the size of the objective value space area which is covered by a set of nondominated solutions.

It combines the three issues previously mentioned (distance, spread and amount of elements of the Pareto optimal set found) into a single value. Therefore, sets differing in more than one criterion cannot be distinguished.

Promising areas of future research

- Incorporation of preferences.
- Emphasis on efficiency (alternative data structures and clever algorithms that minimize Pareto dominance checkings).
- More test functions and metrics.

Promising areas of future research

- More theoretical studies.
- New approaches (hybrids with other heuristics) and extensions of alternative heuristics (e.g., particle swarm optimization, the ant system, cultural algorithms, etc.).
- New applications.
- What to expect for the third generation?

Promising areas of future research

- Tackling dynamic (multiobjective) test functions, handling uncertainty and high epistasis.
- Answering fundamental questions such as: what makes difficult a multiobjective optimization problem for an EA? Can we really produce reliable metrics for multiobjective optimization? Can we design robust MOEAs? Is there a way around the dimensionality curse in multiobjective optimization? Can we benefit from coevolutionary schemes?

Please visit our EMOO repository located at:

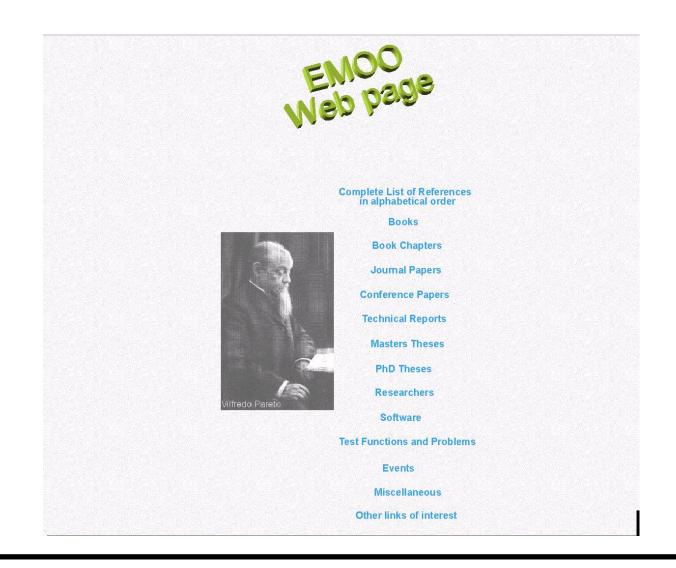
http://delta.cs.cinvestav.mx/~ccoello/EMOO

with mirrors at:

http://www.jeo.org/emo

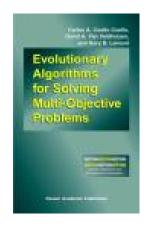
and:

http://www.lania.mx/~ccoello/EMOO



The EMOO repository currently contains:

- Over 940 bibliographic references including 36 PhD theses
- Contact info of about 50 EMOO researchers
- Public domain implementations of SPEA, NSGA, NSGA-II, the microGA, and PAES, among others.



You can consult the following book recently published:

Carlos A. Coello Coello, David A. Van Veldhuizen and Gary B. Lamont, **Evolutionary Algorithms for Solving**Multi-Objective Problems, Kluwer Academic Publishers, New York, May 2002, ISBN 0-3064-6762-3.