

Hybrid evolutionary algorithm and gradient-based algorithm for multi-objective optimization

Zhang Kenneth, Wang Gui Zheng, and Xu Kuang Jie
Southern University of Science and Technology

Traditional multi-objective optimization problems (MOP) have been rigorously studied and various algorithms based on the principles of evolution, such as NSGA-II, SPEA-II, MOEA/D, and SMS-EMOA, have been proposed. An alternative approach to solving the MOP includes the calculation of gradients and guiding the search towards some local optimum. In this paper, we present a hybrid algorithm combining the effects of gradient-descent algorithms as well as evolutionary algorithms.

I. AN OVERVIEW OF THE MOP AND ITS RELATED CONCEPTS

A multiobjective optimization problem (MOP) can be stated as follows: [?]

$$\begin{aligned} &\text{minimize } F(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T \\ &\text{subject to } \mathbf{x} \in \Omega \end{aligned} \quad (1)$$

where Ω is the *decision (variable) space*, $F : \Omega \rightarrow \mathbb{R}^m$ consists of m real-valued objective functions and \mathbb{R}^m is called the objective space. The *attainable objective set* is defined as the set $\{F(\mathbf{x}) \mid \mathbf{x} \in \Omega\}$.

Assuming minimization of the objective functions, among solution vectors $\mathbf{y}, \mathbf{y}' \in \mathbb{R}^m$, a partial order is defined by the *dominance relation* “ \prec ”: [?]

$$\begin{aligned} &\mathbf{y} \prec \mathbf{y}' \text{ (y dominates y')} \\ \iff &\forall i \in \{1, \dots, m\} : y_i \leq y'_i \\ &\text{and } \exists j \in \{1, \dots, m\} : y_j < y'_j \end{aligned} \quad (2)$$

For $\mathbf{x}, \mathbf{x}' \in \Omega$, we write $\mathbf{x} \prec \mathbf{x}' \iff F(\mathbf{x}) \prec F(\mathbf{x}')$. The minimal elements of the dominated relation are called *Pareto-optimal* and the set of Pareto-optimal decision vectors is called *Pareto(-optimal) set*:

$$\{\mathbf{x} \in \Omega \mid \nexists \mathbf{x}' \in \Omega : \mathbf{x}' \prec \mathbf{x}\} \quad (3)$$

The corresponding image under F in the solution space \mathbb{R}^m is called *Pareto front*. A point in the search space is called *non-dominated* within a set $A \subset \Omega$ if there is no point in A dominating it.

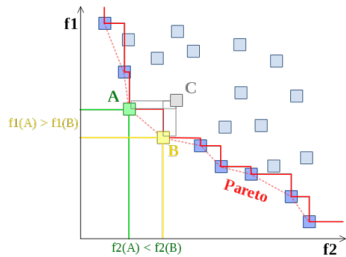


FIG. 1. Concept of pareto-optimality. Blue squares are the non-dominated solutions, white squares are dominated solutions.

A. Elitism

[?] Elitism has always been a major topic of discussion in the evolutionary computation community. This mainly concerns cross-generational interaction. Though it is clear that all natural individuals have a limited lifespan, there is no general answer to how this concept should be applied to evolutionary algorithms. Interpretations are diverse: Canonical genetic algorithms, for instance, allow no parental survivability at all. In generation gap methods at least a fraction of the next generation is reserved for good parents, and in the $(\mu + \lambda)$ evolution strategies parents live as long as they are surpassed by better offspring.

The question of when to accept a newly generated solution, in spite of previously found ones, arises in all kinds of iterative melioration processes, e.g. simulated annealing, where the probability of accepting (temporary) worsening is controlled by a ‘cooling’ schedule.

As opposed to single-objective optimization, the question of elitism becomes even more complicated with MOEAs. Here, the set of decision alternatives is no longer totally ordered, and alternatives might be incomparable. This makes an ‘elitist’ decision difficult, so a number of different MOEA implementations emerge with elitistic features. These features may be classified along different criteria:

- Using a secondary ‘elitist’ population vs. implicit elitism through a ‘plus’ selection as in evolution strategies;
- The elitism strategy, or how the elitist population is updated;
- The evaluation strategy, or how the elite individuals affect the fitness assignment of the current generation and vice versa;
- The re-insertion strategy, or how elite individuals take part in the production of offspring;
- The control flow, or when archiving and re-insertion take place.

As this classification is not hierarchical, algorithms can arbitrarily combine different features.

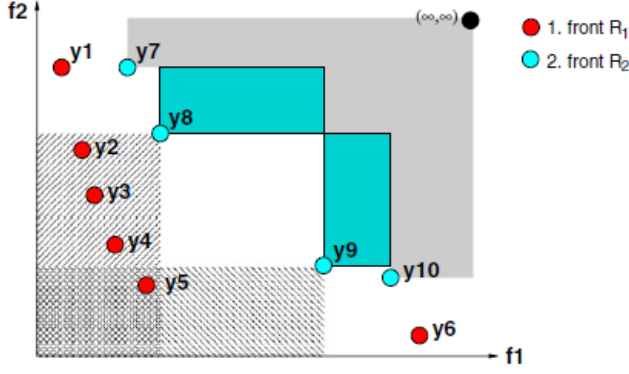


FIG. 2. Concept of hypervolume. Red circles are non-dominated front whereas blue circles are the next front. The hypervolume for the second front is indicated as the union of the gray shade and the blue shade.

B. Hypervolume

[?] A variety of quality indicators is used in order to measure the quality of Pareto front approximations. Among them, the *hypervolume measure* or \mathcal{S} *metric* is of outstanding importance. It is a quality indicator that rewards the convergence towards the Pareto front as well as the representative distribution of points along the front. The hypervolume measure was originally proposed by Zitzler and Thiele, who called it the *size of dominated space*. Let Λ denote the Lebesgue measure, then the \mathcal{S} metric is defined as:

$$\mathcal{S}(B, \mathbf{y}_{\text{ref}}) := \Lambda \left(\bigcup_{\mathbf{y} \in B} \{\mathbf{y}' \mid \mathbf{y} \prec \mathbf{y}' \prec \mathbf{y}_{\text{ref}}\} \right), B \subset \mathbb{R}^m \quad (4)$$

Here, $\mathbf{y}_{\text{ref}} \in \mathbb{R}^m$ denotes a reference point that should be dominated by all Pareto-optimal solutions.

Fleischer [5] proved that, given a finite search space and a reference point, maximisation of the hypervolume measure is equivalent to finding the Pareto set. Therefore, a set of bounded size which obtains the maximal \mathcal{S} metric value that is possible for the given size, only consists of Pareto-optimal solutions. It has been empirically observed that for a fixed number of points, the maximisation of the hypervolume metric yields subsets of the Pareto front which are well-distributed (cf. [6,7]). [?]

II. A DISCUSSION OF CONTEMPORARY MOEAS

Multi-Objective Evolutionary Algorithms (MOEAs) have emerged as a powerful and broadly applicable subset of evolutionary computation techniques that effectively address multi-objective optimization problems (MOPs). MOPs are characterized by the need to optimize more than one conflicting objective, necessitating a search for

a set of optimal solutions, known as Pareto optimal solutions. The advent of Vector Evaluated Genetic Algorithm (VEGA) by Schaffer et al. in 1985 pioneered the generalization of Genetic Algorithms (GAs) to handle vector-valued (multi-objective) fitness functions. However, it was constrained to a single feature optimization, prompting a surge in the development of more sophisticated approaches.

Building upon the foundational work of VEGA, Horn et al.'s Niche Pareto Genetic Algorithm (NPGA) in 1994 introduced the concept of niching and the sharing parameter, which allows the algorithm to maintain a diverse population of solutions. The addition of a comparison set was particularly novel, allowing the algorithm to filter out inferior solutions efficiently.

Further refinement came with Fonseca & Fleming's Multiple Objective Genetic Algorithms (MOGA) in 1993, which pioneered the use of domination-based ranking for fitness assignment. MOGA also discussed niche formation using both phenotypic and genotypic distances, highlighting the importance of genetic encoding in representing phenotype similarities.

The integration of random weight in Murata & Ishibuchi's 1995 study introduced an essential aspect of diversity maintenance, allowing for broader exploration of the solution space. This concept was enhanced by Ishibuchi et al. in 1998 with Multi-Objective Genetic Local Search (MOGLS), which combined random weight vectors with local search techniques.

Deb et al.'s NSGA-I in 1994 and its successor, NSGA-II in 2002, marked significant improvements in computational efficiency and solution quality. NSGA-II's introduction of elitism, non-dominated sorting, and crowding distances provided an elegant method of maintaining diversity while being computationally tractable, a considerable advancement from its predecessors.

The Pareto Archive Evolution Strategy (PAES) by Knowles et al. in 1999 and the Pareto Envelope-based

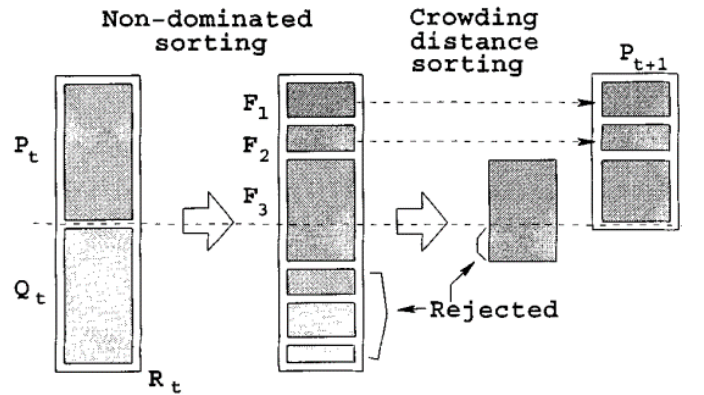


FIG. 3. NSGA-II utilizes an external archive (elitism) to guide the selection process of the next group of population.

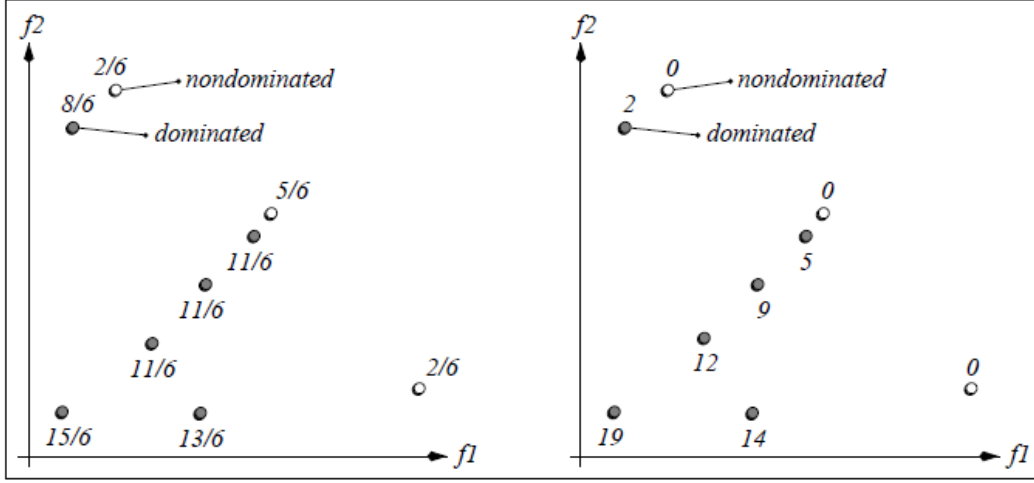


FIG. 4. SPEA-II using a fitness assignment scheme based on the concept of Pareto strength.

Selection Algorithm (PESA) by Corne et al. in 2000, took the evolution strategies route, focusing on archiving and hypergrid techniques to preserve solution quality across generations.

Zitzler et al.'s Strength Pareto Evolutionary Algorithm (SPEA) in 1999 and SPEA-II in 2001 provided a holistic approach that assimilated characteristics of VEGA, MOGA, NPGA, and NSGA. The introduction of fine-grained strength fitness assignment and nearest neighbor density estimation in SPEA-II represented a significant step towards the maintenance of boundary solutions and the enhancement of solution quality.

In 2007, Zhang et al. introduced Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D), which represented a paradigm shift in handling MOPs. By decomposing a multi-objective problem into several single-objective subproblems and optimizing them simultaneously, MOEA/D provided an effective means of exploring and exploiting the problem space.

Somewhat recently, SMS-EMOA by Beume et al. in 2006 presented an approach centered around the hypervolume measure, offering a fitness metric that quantified the “size of the dominated space.” This metric provided an intuitively appealing and direct method of assessing solution quality in terms of the volume of the objective space that is dominated by the Pareto front. This performance indicator shall be the foundation of our proposed work.

III. A SHORT REVIEW OF GRADIENT-BASED MULTI-OBJECTIVE ALGORITHMS

The Multiple-Gradient-Descent Algorithm (MGDA) is an innovative approach in multi-objective optimization that capitalizes on the concept of considering multiple gradients simultaneously. Unlike traditional gradient de-

scendent, which is tailored for single-objective problems, MGDA identifies a search direction that is descent for all objectives. This is accomplished through the computation of a common descent direction derived from the gradients of all objectives, seeking a compromise between them. The algorithm is particularly adept at navigating the trade-off surface defined by conflicting objectives, steering the solution towards Pareto optimality.

Pareto-Set Learning (PSL) represents a more nuanced approach to multi-objective optimization. It extends beyond the mere pursuit of Pareto front by learning the underlying structure of the Pareto set. PSL employs a collaborative filtering method that harnesses the gradients of the objectives to learn a model of the Pareto set. This model is then utilized to predict Pareto-optimal solutions. PSL's strength lies in its ability to incorporate the geometry of the problem, which often results in a more efficient convergence to high-quality Pareto-optimal solutions compared to algorithms that do not leverage such information.

To our great interest is the usage of the hypervolume indicator in the SMS-EMOA algorithm. A recent algorithm called Hypervolume Indicator Gradient Ascent Multi-objective Optimization (HIGA-MO) combined the idea of gradient descent with the hypervolume indicator. HIGA-MO introduces a novel perspective by focusing on the hypervolume indicator, a set performance measure that reflects the size of the space dominated by the Pareto front. By ascending the gradient of the hypervolume indicator, HIGA-MO seeks to expand the dominated space, effectively pushing the Pareto front towards better trade-offs. This approach is particularly powerful as the hypervolume indicator inherently balances the spread and convergence of the solutions. HIGA-MO is adept at uncovering a set of non-dominated solutions that provides a comprehensive portrayal of the trade-offs involved in multi-objective scenarios.

IV. BACKGROUND CONCEPTS IN CONDUCTING GRADIENT ALGORITHMS

To utilize gradient information, we assume that each objective function f_i is continuous differentiable *almost everywhere* in S_i . Thus, the MOP can be formulated as follows: [?]]

$$\max_{\mathbf{x} \in S} F(\mathbf{x}), \quad \Omega = \bigcap_{i=1}^m \Omega_i \subseteq \mathbb{R}^d \quad (5)$$

where $F(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_m(\mathbf{x})]^T$ is a vector-valued function composed of m objective functions. This aligns with our earlier definition of the MOP. Due to the continuous differentiability assumption on each objective function, F is again continuous differentiable almost everywhere in S . The gradient information is expressed as transpose of the Jacobian matrix as follows:

$$\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}} = [\nabla f_1(\mathbf{x}), \dots, \nabla f_m(\mathbf{x})] \quad (6)$$

In addition, it is assumed that each gradient vector above (column vector) can be computed either analytically or numerically. In MOPs, a set of decision vectors are moved in *decision space* Ω to approximate the Pareto efficient set.

V. AN EFFICIENT OPERATOR UTILIZING GRADIENT INFORMATION OF PARETO-OPTIMAL SOLUTIONS

Usually gradient descent information is used on a single point to determine the direction of search space. We

propose an efficient searching method using the ‘migration’ operator. We define some terms and make analogies to a natural phenomena: In classical theories of anthropology and sociology, a ‘clan’ consists of a ‘clan leader’ and dominated ‘clan members’. We derive the concept of ‘leader’ by referring to non-dominated solutions as ‘leading solutions’, and the set of populations dominated by

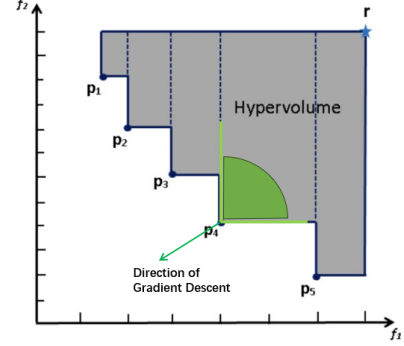


FIG. 5. The concept of migration operator using gradient information of a Pareto-optimal solution

this ‘leading solutions’ is named a ‘clan’. We introduce a ‘migration’ operator where the ‘leader’ determines first its local Gradient information, and establishes a search direction in which the entire ‘clan’ will move towards; however, during the process an indicator needs to be evaluated to determine the efficiency and utility of the migration step. Here we use the HV-indicator to determine solutions that improves the overall hypervolume. An elitist approach is conducted by maintaining a Pareto Archive which the ‘migration step’ can compare the hypervolume with, and we collectively discard individuals that degrade the level of hypervolume after the migration step.