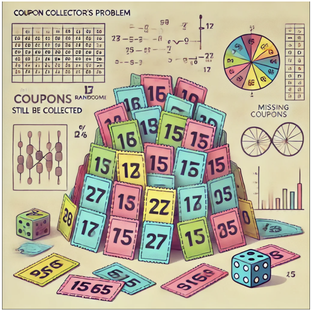
Overcoming Coupon Collector's Syndrome when Optimizing Multimodal Domains



PPSN-2024

workshop on "Multimodal Multiobjective Optimisation"

Ofer M. Shir (Tel-Hai College & Migal Institute, ISRAEL)
ofersh@telhai.ac.il

problem statement: singleobjective multimodal

Search over multimodal domains may induce different targets / research questions:

(i) **multiple solutions**: usually for engineering purposes
(ii) **global only!** as in the global optimization challenge
(iii) **landscape research**: assessing structure and topology

These 3 targets are typically addressed by different researchers and clearly result in different algorithms design.

attaining multiple solutions

The engineering motivation is usually strong, sometimes even to the extreme – it is worth mentioning the so-called *Second Toyota Paradox*:[1]

"Delaying decisions, communicating ambiguously, and pursuing an excessive number of prototypes, can produce better cars faster and cheaper."

This engineering incentive provides a clear practical motivation to the area of **Niching Methods**,[2] whose mission statement is stated as *Attaining the optimal interplay between partitioning the search space into niches occupied by stable subpopulations, by means of population* ***diversity preservation****, and exploiting the search in each niche by means of a highly efficient optimizer with local-search capabilities.*

niching in a nutshell

A robust measure to assess the number of species is the Solow-Polasky diversity:[3,4] let $\Psi := (\psi_{ij}) \in \mathbb{R}^{n \times n}$ represent the exponential decay of the mutual distances,

$$\psi_{ij} = \exp\left(-\gamma \cdot d_{\mathcal{X}}(\vec{x}_i, \vec{x}_j)\right)$$

Then, this diversity measure is defined as the following scalar:

$$D_{SP} = \vec{1}^T \Psi^{-1} \vec{1} \tag{1}$$

Niching techniques either "penalize dense subpopulations",

- fitness sharing (Holland, 1975)
- crowding (de Jong, 1975)
- clearing (Petrowski, 1996)

or "enforce separation"

- islands
- clustering
- **repulsion**

repelling subpopulations

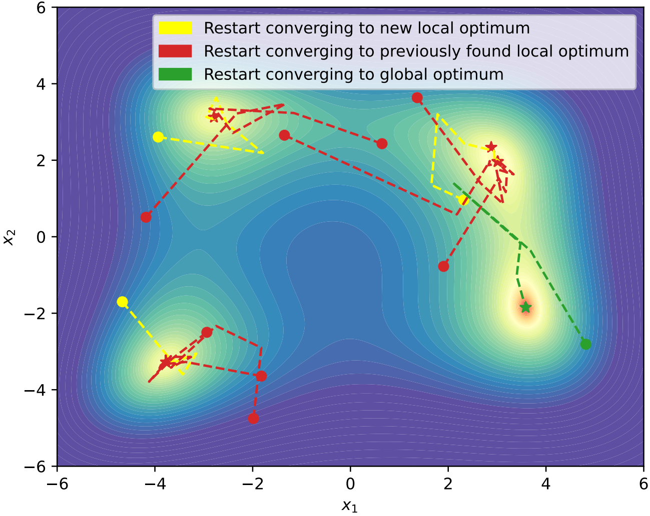
The idea of “avoiding duplicates” is a fundamental concept in heuristic search – recall the classical **Tabu Search**.

The CMSA with *Repelling Subpopulations* (RS-CMSA) has been proposed as a niching technique (Ahrari+Deb+Preuss, 2017) and proved successful as such.
It introduced the Tabu Rejection Probability to eliminate candidate search points in areas that have been already populated.

This repulsion concept has proved beneficial in solving niching problems, and advanced the field also by removing assumptions on niche radius or landscape structure.

alternative: going sequential

A straightforward approach of *iteration* can be used to sequentially locate multiple peaks in the landscape via an *iterative local search*:[5]



the coupons collector's problem

Revisiting optima resembles the so-called **Coupon Collector's Problem** – a classic problem in Probability Theory:

Suppose there are q distinct types of coupons, and each time you collect a coupon, it is equally likely to be any one of these q types. How many coupons do you need to collect, on average, to obtain at least one of each type?

cost of retrials' naïvity

Given q unique coupons, the **expected number of trials** needed to collect them all, with replacement, is

$$q \cdot H_q \quad \text{with } H_q := \sum_{t=1}^{q} \frac{1}{t} \quad \text{(being the } q^{th} \text{ harmonic number).}$$

This can be approximated by:

$$E(q) \approx q \ln q + \gamma q$$

where γ is the *Euler-Mascheroni constant*, $\gamma \approx 0.577$.

For instance, given a target set of 50 distinct coupons, the expected number of coupons needed to cover this set is ≈225.

naïve restarts

If the procedure is blind to any information accumulated throughout previous runs, and it sequentially restarts stochastic search processes, the ambition to hit a different peak in every run resembles the collector's hope to obtain all the coupons in only q trials.
Overall, it is likely to encounter *redundancy*, and the number of expected iterations is then increased by a factor.

A **redundancy factor** can be derived **if the peaks are of equal fitness** (the probability to converge into any of the q peaks is uniform and equal to $1/q$): H_q.[6]

but optima are often non-uniform

The attraction strengths of different basins play a role in landscapes with non-uniform optima. Definitions of attraction basins vary:

(i) **Region of Convergence** focuses on the set of initial points that cause an algorithm to converge to a local minimum.

(ii) **Set of Points** formalizes the basin of attraction using the sequence of iterates from the optimization algorithm.

(iii) **Downhill Region** provides a geometric criterion based on the gradient to ensure movement towards the local minimum.

Hill-Valley: a “low-cost” heuristic to determine whether two points belong to the same basin of attraction.

pap300: avoiding redundant restarts in multimodal global optimization

Motivated by Tabu search, and by adopting the niching-repulsion concept, the idea is to prevent scenarios of re-sampling already-visited basins of attraction.

This study, to be presented on Tuesday (Session-4), introduces the so-called RR-CMA-ES.

The results demonstrate improved performance (decreased redundancy factors) over the BBOB+CEC2013 test-sets when compared to standard restart schemes (IPOP, BIPOP).

and now let's arrive at multiobjective optimization

The *coupons collector's multiobjective analogy*:
Given a **set-oriented method** searching over a multiobjective space, the problem occurs when the algorithm picks Pareto optimal points with low decision-space diversity w.r.t. the existing set.

Let $\mathcal{X} \subset \mathbb{R}^n$ denote the set of feasible solutions, $\mathcal{Y} \subset \mathbb{R}^m$ its image in the objective space. If $x \in \mathcal{X}$, then $y = f(x) \in \mathcal{Y}$, and the i^{th} objective function value is $y^{(i)} = f^{(i)}(x)$.
Given a subset E of $\mathcal{Y}$ and a point $y \in \mathcal{Y}$, we quantify the *diversity measure* of y with respect to this subset ($y_e \in E$):

$$d_E(y) := \quad \min_{y_e \in E} \left(\max_{1 \leq i \leq m} y^{(i)} - y_e^{(i)} \right). \tag{2}$$

set-oriented method with two diversity measures

An algorithm that iteratively minimizes $d_E(y)$ already exists (DMA)[7] and is proven to obtain Pareto optimality.[8]
But how do we overcome the syndrome?

Given a metric $d_{\mathcal{X}}(\cdot,\cdot)$ over the decision space $\mathcal{X}$, for any subset $V \subseteq \mathcal{X}$ we define $d_{\mathcal{X}}(x, V) := \min\{\ d_{\mathcal{X}}(x, v \mid v \in V)\}\ \equiv d_V(x)$.

We then formulate a **paired optimization problem**:

$$\boxed{\textbf{Primal:}\quad \text{maximize}_x\ d_V(x),\ \text{subject to}\ x \in \mathcal{X},\ y = f(x),\ d_E(y) \leq \delta_E} \tag{3}$$

$$\boxed{\textbf{Dual:}\quad \text{minimize}_x\ d_E(y),\ \text{subject to}\ x \in \mathcal{X},\ y = f(x),\ d_V(x) \geq \delta_V} \tag{4}$$

greedy over the primal-dual is a proven 2-approximation

In the Primal (3), the decision-space diversity is enhanced by compromising the Pareto optimality by δ_E.
In the Dual (4), the aim is to add Pareto optimal points that are far enough w.r.t. $d_V(x)$, at least δ_V.

Dual Greedy Algorithm for solving (4):

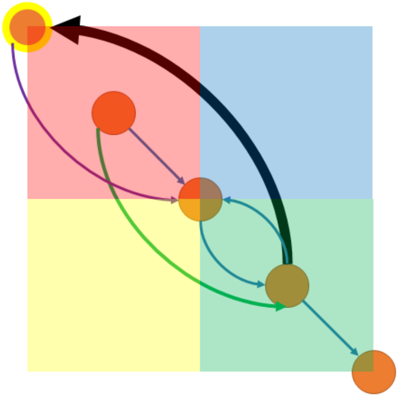
1. **Initialize:**
   (a) Find $V_1 = \{x_1\} \subseteq X_{Par}, E \subseteq E_{eff}$
   (b) Set $V = V_1, j = 2, J =$ (maximal number of iterations).
2. **Solve the Dual optimization problem (4).** If there is a feasible solution then $y^* = f(x^*)$ is the optimal solution, else go to Step 4.
3. **Set** $V_j = V_{j-1} \cup \{x^*\}$ and $V = V_j, j = j + 1$. If $j > J$, go to Step 4, otherwise go to Step 2.
4. **Return** V.

2-approximation proof + aircraft engine design

The obtained V is proved to be a 2-approximation solution.

We solved an **aircraft engine design** problem (UTRC: United Technologies Research Center) with highly-satisfactory results:
Objectives: minimize Design Cost + minimize Fuel Consumption
Search-space: 6 design variables, 70 state variables, a few hundreds of constraints; a tailored d_V: weighted normalized Euclidean, greater weight on integer.
Implementation: AMPL on top of Bonmin solver (COIN-OR).

Zadorojniy, A., Masin, M., Greenberg, L., Shir, O.M., Zeidner, L.: *Algorithms for Finding Maximum Diversity of Design Variables in Multi-Objective Optimization*. Volume 8 of Procedia Computer Science, Elsevier (2012) 171-176.



Discussion

danke

Bibliography

[1] J. J. Cristiano, C. C. White, and J. K. Liker, "Application of Multiattribute Decision Analysis to Quality Function Deployment for Target Setting," *IEEE Transactions on Systems, Man, and Cybernetics: Part C*, vol. 31, no. 3, pp. 366–382, 2001.

[2] O. M. Shir, *Handbook of Natural Computing: Theory, Experiments, and Applications*, ch. Niching in Evolutionary Algorithms, pp. 1035–1069.
Berlin-Heidelberg, Germany: Springer-Verlag, 2012.

[3] A. Solow and S. Polasky, "Measuring biological diversity," *Environmental and Ecological Statistics*, vol. 1, pp. 95–103, 1994.

[4] T. Ulrich, J. Bader, and L. Thiele, "Defining and Optimizing Indicator-Based Diversity Measures in Multiobjective Search," in *PPSN-XI* (R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, eds.), vol. 6238 of *Lecture Notes in Computer Science*, pp. 707–717, Springer, 2010.

[5] H. R. Lourenço, O. C. Martin, and T. Stützle, "Iterated local search: Framework and applications," *Handbook of metaheuristics*, pp. 129–168, 2019.

[6] O. M. Shir, "Niching In Evolutionary Algorithms," in *Handbook of Natural Computing* (G. Rozenberg, T. Baeck, and J. N. Kok, eds.), Springer Verlag, 2012.

[7] O. M. Shir and M. Emmerich, "Multi-objective mixed-integer quadratic models: A study on mathematical programming and evolutionary computation," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2024.

[8] M. Masin and Y. Bukchin, "Diversity maximization approach for multiobjective optimization," *Operations Research*, vol. 56, no. 2, pp. 411–424, 2008.

Backup

region of convergence

In the context of an iterative optimization algorithm (e.g., gradient descent), the **region of convergence** refers to the subset of the search space such that starting from any point within this region, the algorithm converges to a particular local minimum x^*. Formally, if $\{x_k\}_{k=0}^{\infty}$ is the sequence of iterates produced by the algorithm starting from $x_0 = x$, then the region of convergence to x^* is:

$$\mathcal{R}(x^*) = \left\{ x_0 \in \mathbb{R}^n \mid \lim_{k \to \infty} x_k = x^* \right\}$$

Here, the algorithm iteratively updates x_k based on some rule, and the region $\mathcal{R}(x^*)$ contains all points that eventually lead to x^*.

attraction basins

This definition formalizes the basin of attraction as a **set of points** in the search space. For a local minimum x^* of a continuously differentiable function $f(x)$, the **basin of attraction** is the set of points $x_0 \in \mathbb{R}^n$ such that if an algorithm starts at x_0, the sequence of points generated by the algorithm converges to x^*:

$$B(x^*) = \left\{ x_0 \in \mathbb{R}^n \mid \lim_{k \to \infty} x_k = x^* \text{ where } x_{k+1} = \mathcal{A}(x_k) \right\}$$

Here, $\mathcal{A}(x_k)$ represents the update rule of the optimization algorithm (e.g., gradient descent, Newton's method), and the set $B(x^*)$ contains all points that converge to x^*.

downhill region

The **downhill region** provides a geometric interpretation of the basin of attraction. It refers to the region in the search space where the objective function $f(x)$ is decreasing as you move toward a local minimum x^*. Mathematically, for a descent-based optimization method (such as gradient descent), a point x_0 is in the downhill region if:

$$\nabla f(x_0) \cdot (x_0 - x^*) > 0 \quad \text{for all} \quad x_0 \in \mathcal{D}(x^*)$$

This indicates that the gradient of $f(x)$ at x_0 points in the direction of the local minimum x^*, meaning that the function value decreases as the algorithm moves toward x^*.