Applying Stopping Criteria in Evolutionary Multi-Objective Optimization

Nayat Sanchez-Pi¹    Luis Martí²

Group of Research on Intelligence and Optimization (RIO Group)

¹ Institute of Mathematics and Statistics, State University of Rio de Janeiro
² Institute of Computing, Universidade Federal Fluminense

2017 IEEE Congress on Evolutionary Computation
Talk outline

- Intro
- Multi objective optimization
- Evolutionary algorithms
- Stopping MOEAs
- Approaches
- Use cases
Do you know them?

Rudolph Diesel

Fred Duesenberg

Internal combustion ‘diesel’ engine

Hydraulic brakes

It seems that stopping things does not make you famous.
What does “until end condition is met” really mean?
Most *if not all* optimization problems involve more than one objective function to be optimized simultaneously.

- For example: optimize a given feature of an object while keeping under control the resources needed to elaborate that object.
- Sometimes those other objectives are converted to constraints or fixed to default values, but they are still there.
- Multi-objective optimization is also known as *multi-objective programming, vector optimization, multicriteria optimization, multiattribute optimization* or *Pareto optimization* (and probably by other names, depending on the field).
The multi-objective ‘fruit selection problem’

We do it all the time!

¹from http://xkcd.com/388
Two-dimensional example

minimize \( F(x) = \langle f_1(x), f_2(x) \rangle \),
with \( x \in \mathcal{D} \subseteq \mathbb{R}^2 \).
Multi-objective optimization problem

minimize $F(x) = \langle f_1(x), \ldots, f_M(x) \rangle$, with $x \in D$.

- $D$: feasible set — can be defined as constraints;
- $O$: objective set;
- optimality — *Pareto dominance*;
- $D^*$: Pareto-optimal set;
- $O^*$: Pareto-optimal front, and;
- $P^*$: optimizer solution.

A decision maker selects elements of $P^*_t$. 
Pareto dominance relation

Usually, there is not a unique solution that minimizes all objective functions simultaneously, but, instead, a set of equally good *trade-off* solutions.

- Optimality can be defined in terms of the Pareto dominance relation.
- That is, having \( x, y \in \mathcal{D} \), **\( x \) is said to dominate \( y \) (expressed as \( x \prec y \))** iff \( \forall f_j, f_j(x) \leq f_j(y) \) and \( \exists f_i \) such that \( f_i(x) < f_i(y) \).
- Having the set \( \mathcal{A} \). \( \mathcal{A}^* \), the *non-dominated subset* of \( \mathcal{A} \), is defined as

\[
\mathcal{A}^* = \{ x \in \mathcal{A} | \exists y \in \mathcal{A} : y \prec x \}.
\]

- The *Pareto-optimal set*, \( \mathcal{D}^* \), is the solution of the MOP. It is the subset of non-dominated elements of \( \mathcal{D} \).
- Its image in objective set is called the *Pareto-optimal front*, \( \mathcal{O}^* \).
Example: Dent problem

minimize $f_1(x), f_2(x)$

such that

$$f_1(x) = \frac{1}{2} \left( \sqrt{1 + (x_1 + x_2)^2} \sqrt{1 + (x_1 - x_2)^2} + x_1 - x_2 \right) + d,$$

$$f_2(x) = \frac{1}{2} \left( \sqrt{1 + (x_1 + x_2)^2} \sqrt{1 + (x_1 - x_2)^2} - x_1 - x_2 \right) + d,$$

with

$$d = \lambda e^{-(x_1 - x_2)^2}.$$

generally $\lambda = 0.85$ and $x \in [-1.5, 1.5]^2$. 
Example: Dent problem — Plots

**Decision set**

- $x_1$ and $x_2$ axes ranging from -1.5 to 1.5.

**Objective set**

- $f_1(x)$ and $f_2(x)$ axes ranging from 0.5 to 3.5.
Example: Dent problem — the dominance relation
Example: Dent problem — non-dominated front

Decision set

Objective set

- $f_1(x)$
- $f_2(x)$
Performance indicators

How can we compare different (sets of) solutions?

- Hypervolume indicator;
- additive/multiplicative epsilon indicator;
- R1/R2 indicators;
- inverted generational distance, etc.
The hypervolume indicator

2 From Günter Rudolph’s site:
Formalization of the hypervolume

For a set of solutions $A$,

$$h_{hyp}(A) = \text{volume} \left( \bigcup_{\forall a \in A} \text{hypercube}(a, r) \right).$$

- We need a reference point, $r$.
- Hypervolume is Pareto compliant (Fleischer, 2003): for sets $A$ and $B$, $A \prec B \implies h_{hyp}(A) > h_{hyp}(B)$.
- Calculating hypervolume is $\textbf{NP}$-hard.
A population of individuals;
- individuals are ranked using a **fitness assignment function**;
- **evolution-inspired** operators are applied;
- fittest individuals have a more active role.
Multi-objective evolutionary algorithms (MOEAs)

- One of the hottest topics in EA research.
- MOPs put EAs “to the limit”.
- Succeeded in yielding relevant results.
- Do not make any assumptions about the problem.
- Parallel nature of the search process produces sets of solutions.

Cornerstone issue: fitness assignment.
Ranking a multi-objective population

- Objective aggregation.
- Pareto-based ranking.
- Performance indicator-based ranking.
Many-objective problems

Problems with four or more objectives.

Challenges
- Visualization.
- Poor understanding of convergence and progress → stopping criteria.
- Scalability.

Scalability
- Exponential relation between the number of objectives and the amount of resources.
- Large populations are needed.
Stopping a multi-objective optimization

Stopping criteria

- Detect when there is no sense in proceeding with the search;
- they are usually a heuristic.
- This matter have been overlooked in the EMO context, but;
- complex and real-world applications demand them.

3 from https://www.cartoonstock.com/cartoonview.asp?catref=jhin64.
A stopping criterion is invoked at the end of an iteration of the algorithm being controlled.

**Scenarios**

1. the solution obtained so far is satisfactory;
2. we have a feasible solution that is not satisfactory in terms of optimality, but a better one is unlikely to be produced;
3. the method is unable to converge to any solution, or;
4. the amount of computation performed so far is *sufficient*.

**In brief...**

A stopping criterion should detect “**success**” and “**failure**” scenarios.
A multi-objective optimization stopping criteria

- Judging the advance of the optimization can become as complex as the optimization itself;
- unlike other problems there is no “axis” to be used as reference;
- therefore any assessment must be carried out in a relative fashion, but;
- current performance indicators have a high computational cost.

Desirable properties

- An execution-wise criterion is required because of the nature of the problem;
- Resource requirements should be kept as low as possible, in particular;
- the criterion should be embedded in other processes.
- As few parameters as possible!
Karush-Kuhn-Tucker (KKT) optimality

If we encounter problems, where the feasible set is given implicitly by constraints,

\[ S = \{ \mathbf{x} \in \mathbb{R}^n; (c_1(\mathbf{x}), \ldots, c_C(\mathbf{x})) \leq \mathbf{0} \} , \]

and \( f_1, \ldots, f_M \) and constraint functions \( c_1, \ldots, c_C \) are continuously differentiable.

Definition (Karush-Kuhn-Tucker proper optimality condition\(^4\))

A solution \( \mathbf{x} \in S \) is said to be properly Pareto-optimal if it holds the Pareto-optimality condition and \( \exists \mathbf{b} \in \mathbb{R}^n \) such that

\( \forall i = 1, \ldots, M : \nabla f_i(\mathbf{x})^T \mathbf{b} \leq \mathbf{0} ; \)  \hspace{1cm} (1)

\( \exists j = 1, \ldots, M \) such that \( \nabla f_j(\mathbf{x})^T \mathbf{b} < \mathbf{0} ; \) \hspace{1cm} (2)

\( \forall c_c(\mathbf{x}) = 0 : \nabla c_c(\mathbf{x})^T \mathbf{b} \leq \mathbf{0} . \) \hspace{1cm} (3)

---

\(^4\) Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master’s thesis, Department of Mathematics, University of Chicago, Chicago, IL, USA

Initialize: calculate ideal and approximated nadir objective vectors and show them to the decision maker.

Generate a Pareto optimal starting point (by using e.g. some no-preference method or solution given by the decision maker).

Ask for preference information from the decision maker (e.g. aspiration levels or number of new solutions to be generated).

Generate new Pareto optimal solution(s) according to the preferences and show it/them and possibly some other information about the problem to the decision maker.

If several solutions were generated, ask the decision maker to select the best solution so far.

Stop, if the decision maker wants to; otherwise, go to step 3).

They even have a name for that: psychological convergence\(^6\)

---

Where can we look for ideas?

“Classical” multi-criteria decision making

- We have proper definitions of optimality, like the Kuhn–Tucker condition, but;
- perhaps they are more relevant from a theoretical point of view.
- Ideal solution.

Single-objective EAs

- Search space exploration\(^7\): all states visited with a certain probability;
- objective convergence\(^8\);
- population convergence — population diversity.

---


Guidelines

Desirable properties
- An execution-wise criterion is required because of the nature of the problem;
- resource requirements should be kept as low as possible, in particular;
- the criterion should be embedded in (and profit from) other processes, and;
- to have as few parameters as possible!

At least two components
- A local progress indicator, and;
- An evidence gathering process that combines the local measurements.
Situation

- There has been few theoretical works\(^9\) that deal with EMO convergence, and;
- there has been even more sparse attempts to deal with the stopping issue.

The importance of this matter has not been correctly underscored until recently.\(^{10}\)

---


Deb and Jain were the first authors who proposed the investigation of performance metrics over the run of MOEAs. They used two metrics, one for evaluating the convergence and one for measuring the diversity of the Pareto front.

Convergence metric (CM) calculates the average of the smallest normalized Euclidean distance from each individual in the Pareto front to a precomputed reference set.

For the diversity metric (DVM), all objective value vectors of the Pareto front are projected onto a $m - 1$-dimensional hyperplane which is then uniformly divided into discrete grid cells.

The DVM tracks the number of attained grid cells and also evaluates the distribution by assigning different scores for predefined neighborhood patterns.

---

Approaches in chronological order

Rudenko and Schoenauer\textsuperscript{12} — 2004

- A stopping criterion to be used in conjunction with the NSGA–II algorithm.
- Measure the mean of the spread of the non-dominated individuals.
- Compute an average across iterations of the measurements.

Stoooop! We found a leitmotiv!

To stop when the “indicator” is \textbf{zero} and has a \textbf{flat} tendency.

Approaches in chronological order (II)

Martí, García, Berlanga and Molina (MGBM) — 2007, 2009

- Iteration-wise measurements with the *mutual domination rate* (MDR) indicator, and;
- a simplified Kalman filter tracks the measurements in an *evidence gathering process*.
- It detects when MDR is close to zero and flat.
- The KF helps to pay less attention to measurements at beginning.
- MDR can be embedded in the EMO supervised (if it relies on Pareto dominance).
- Shown to detect “failure” situations.

---


Mutual domination rate

Having...

- Non-dominated individuals of consecutive iterations: $\mathcal{P}_t^*$ and $\mathcal{P}_{t-1}^*$;
- $\Delta (A, B)$: elements of $A$ that are dominated by at least one element of $B$;

Mutual domination rate indicator

- Contrasts how many non-dominated individuals of iteration $t$ dominate the non-dominated individuals of the previous one ($t-1$) and vice versa.
- Formally,

$$ I_{mdr} (\mathcal{P}_t^*, \mathcal{P}_{t-1}^*) = \frac{|\Delta (\mathcal{P}_{t-1}^*, \mathcal{P}_t^*)|}{|\mathcal{P}_{t-1}^*|} - \frac{|\Delta (\mathcal{P}_t^*, \mathcal{P}_{t-1}^*)|}{|\mathcal{P}_t^*|}; $$

It is easy to embed MDR in Pareto-based fitness assignment computations.
## Accumulating evidence via Kalman filters

### Kalman filter
- Estimates the state of a discrete-time controlled process that is ruled by a *linear* stochastic difference equation;
- an efficient computational means to estimate the state of a dynamic system from a series of incomplete and noisy measurements.

### For the stopping purpose...
- It is tracked the MDR indicator at iteration $t$, $l_{mdr}(t)$.
- The filter is set to predict that the indicator will remain the constant.
- The filter status is updated with the measurements of the indicator.

Stop when estimation and confidence interval bellow threshold.
Evolution of the MDR indicator, $I_{mdr} \left( \mathcal{P}_t^*, \mathcal{P}_{t-1}^* \right)$, and the a priori and a posteriori estimations $\hat{I}_t^-$ and $\hat{I}_t$ across iterations. Here the NSGA–II algorithm is supervised as it successfully solves the DTLZ3 problem.
Approaches in chronological order (III)

Offline convergence detection (Trautmann et al.\textsuperscript{15}) — 2008

- A series of statistical hypothesis tests performed on concurrent algorithm executions.
- Generational distance, hypervolume, and spread performance indicators are computed.
- This information is used to decide when the best solutions of the algorithm were obtained.
- Applies Kolmogorov–Smirnov statistical test;
- Sound and robust, but;
- Heavy parameterized and very resource demanding.

Approaches in (…) order (IV): The 2009 boom

Online convergence detection (Wagner et al.\textsuperscript{16}) — 2009

- Also tracks a number of progress indicators;
- stops when the variance is below a threshold ($\Xi^2$ test), and;
- passes a linear trend statistical test — $t$-test.
- Less resource demanding than the previous one.

Dominance–based stability measure — Bui et al., 2009\textsuperscript{17}

- Dominance–based quality indicator.
- Determines how many solutions in a given radius dominate current solutions.
- Only local measurements.
- The radius is hard to set a priori.


Approaches in (...) order (V): The 2009 boom

Guerrero et al., 2009

- Improved formulation of filters by adding adaptation;
- combines information from progress indicators, and;
- voting is used to decide when to stop.
- More robust than MGBM.

---

More recent developments

Goel and Stander, 2010

- An indicator that measures how stable are solutions in an archive that contains the best solutions obtained.
- Tied to a specific version of NSGA–II.
- Track the growth of an archive of non-dominated solutions over a few generations.
- Determine the consolidation ratio (CR): the ratio of the number of members in a preceding archive that are retained in the current archive, to the size of the current archive (archive inter-spacing is user defined).
- CR represents the proportion of converged solutions in the current archive.
- Terminates when the improvement of CR falls below a threshold.

---

More recent developments

Least square stopping criterion (LSSC) — Guerrero et al., 2010

- Test if a sample of the indicator value can be modelled by a linear regression.
- Instead of using the variance (as in OCD) the slope is used.
- The value of the slope is independent of the scale of the indicators.

---

More recent developments (II)

Approximate KKT — Deb, Abouhawwash and Dutta (2015)

- Approximate KKT points as a measure of proximity to the Pareto-optimal front.

Population entropy-based — Saxena et al. (2016)

- Entropy-based dissimilarity measure.
- Identifies when the algorithm stabilizes.
- Relies on the relative entropy:
  - the two distributions compared correspond to objective values of non-dominated subset of populations in two successive iterations.

---


Summary

Custom indicators:
- mutual domination rate;
- dominance–based quality indicator;
- indicator-based progress indicator;
- consolidation ratio;
- approximate KKT, and;
- population entropy-based.

Evidence gathering:
- online convergence detection;
- Kalman filters;
- combination of filters, and;
- least squares stopping criterion.
Salient issues: Local progress

Are performance indicators suitable?
- Applying them measure the improvement of the solutions, but;
- in “plateau” situations when evolution could be progressing at a slow pace?
- Dominance–based methods are not suitable for the recent indicator–based EMOs.

What are the characteristics of stagnated EMOs?

What if we analyze the “health” of the evolutionary process?
Fitness homogeneity

- It has been shown that a balance between dominated and non-dominated solutions is needed.
- If most of the population is non-dominated the search process becomes stagnant.

We have been working in this matter\textsuperscript{24}

- Simple approach that computes the deviation of the fitness values;
- needs a transformation scheme, but;
- can be used with indicator-based EMOs.

Using stopping criteria in other contexts:
Population Restarting
Detected stagnation and population restart

What if we detect non-progress situations to act upon them.

...for example for doing a population restart and improve diversity.

Population restart in multi-objective estimation of distribution algorithms (MEDAs)

- Diversity loss has been well documented in MEDAs.
- This hampers their results, in particular in many-objective problems.
- There are different approaches can be used with indicator-based EMOs.
The importance of restarting

- Most successful algorithms in major competitions were restart-enhanced variants of the evolutionary algorithms.
- A restarted version of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)\(^{25}\) with increasing population size outperforms the original versions of CMA-ES.
- In the Adapted Maximum-Likelihood Gaussian Model Iterated Density-Estimation Algorithm (AMaLGaM)\(^{26}\) a specific restart scheme increases the number of solutions upon each restart by alternating between two approaches: a single run with a larger population and more parallel runs.

---


EDA based on the multivariate estimation of copulas (EDA-MEC)

Composed of the following 4 steps:

1. Selection or construction of the copula;
2. Estimation of the copula parameter(s);
3. Estimation of the marginal distributions; and
4. Sampling the copula and generating the new individuals.

Complemented by two processes:

- Transforming the probabilistic distribution to generate novel individuals, and
- a strategy for population restarting.

---

The $S$-Metric Selection EDA-MEC (SMS-EDA-MEC)\textsuperscript{28}

Question:
How to do the restarting in a multi-objectively friendly way?

SMS-EDA-MEC restarting

- We need to detect stagnation → stopping criteria;
- We have to restart a limited part of the (dominated?) population.

- We apply online convergence detection (OCD) (Wagner et al., 2009) configured in a ‘loose’ way.
- Two restart groups: local and global.
SMS-EDA-MEC restart

$\mathcal{P}_{\text{local\ restarted}}$ is created by sampling a uniform distribution on the interval of the feature set currently occupied by the population $\mathcal{P}$ as,

$$\mathcal{P}_{\text{local\ restarted}} = \left\{ x \sim U(\ell_{\text{lower}}^{\text{lower}}, \ell_{\text{upper}}) \right\},$$

where $\ell_{i}^{\text{lower}} = \min_{x \in \mathcal{P}} (x_i)$ and $\ell_{i}^{\text{lower}} = \max_{x \in \mathcal{P}} (x_i)$ for $i = 1, \ldots, n$.

It should be noted that in practical applications it is convenient to disable restarting in the last part of the optimization.
Using stopping criteria in other contexts: Progressively Adding Objectives
Evolving Voronoi diagrams for anomaly detection

VorEAI\textsuperscript{29,30} evolves Voronoi diagrams that encode which areas correspond to ‘normal’ or ‘anomalous’ data.

- **Multi-objective:**
  - Classification metrics, like accuracy and recall.
  - Volume-based objectives for compact representation of data.

- An individual represents a Voronoi diagram as a set of sites, \( \mathcal{I} = \{ \mathbf{S}_i \} \), where each site has an associated label, \( \mathbf{S}, \ell \in \{ \text{OK}, \text{Anom} \} \).

\[
\text{clsfy} (\mathcal{I}, \mathbf{x}) = \mathbf{S}^* . \ell; \quad \mathbf{S}^* = \arg \min_{\mathbf{S}_i \in \mathcal{I}} \| \mathbf{x} - \mathbf{S}_i \|
\]


In the multi-objective Animal Farm

Some objectives are important but some objectives are more important than others!

- Some objectives are more important than others.
- Some objectives are easier to attain than others.
- Some objectives are ‘disruptive’ and cause that individuals become unable to deal with other objectives.

That is the case of the length minimization objectives.
In the VorEAl case

- Tried to improve VorEAl by adding a length minimization objective.
- Severe diversity loss.
- Ended up having a lot of small individuals (in terms of number of sites).
- This could also happen in multi-objective genetic programming when one objective is meant to reduce bloat.
Progressive Addition of Objectives (PAO)\textsuperscript{31}

PAO starts with a set of ‘primary’ objectives selected \textit{a priori} and progressively adds objectives that are the least disruptive.

- A VorEAI instance is evolved until convergence is detected by the OCD method.
- Different combinations of objectives are tested in a limited number of iterations.
- The least disruptive one is selected.

Intrigued? — More on this in GECCO 2017. (the paper is available online)

Outlook
Salient issues: Comparing criteria

- Testing stopping criteria is an awkward task, but;
- comparing them is even more.

We need to come up with a set of problem+algorithm pairs.

- Different convergence rates;
- incorrect parameter setup, in particular;
- population sizes.
Available software

- Matlab taxonomy of stopping criteria:

- Python implementation of current state of the art:
  - Can be used as in conjunction with DEAP and inspyred.

- Python implementation of entropy-based termination criteria:
  - Includes the implementation of the MOEAs involved in the paper.
Final remarks

- There is room for improvement;
- **indicators**: assume elitism, take into account diversity, etc.
- **evidence gathering**: many tools available: statistics, Markov, filters, etc.
- Ideas from anomaly detection, time-series processing and/or outlier detection.
- *A posteriori* analysis.
- The most recent review on the topic is still valid\(^{32}\).

Eskerrik asko!
Thank you!
Obrigado!
¡Gracias!

nayat@ime.uerj.br; lmarti@ic.uff.br

http://lmarti.com/stopping
One final comment

Welcome to CBIC 2017

On behalf of the organizing committee, it is our pleasure to invite you to the XIII Brazilian Congress on Computational Intelligence - XIII Congresso Brasileiro de Inteligência Computacional (CBIC 2017) that will take place at the Institute of Computing of the Universidade Federal Fluminense in Niterói, Rio de Janeiro, Brazil from 30 October to 1 November 2017.

CBIC 2017 aims to provide a high-level international forum for scientists, researchers, engineers, professionals and educators to disseminate their latest research results and exchange views of the future research directions in the area of computational intelligence.

Apart from the technical programs, participants are also cordially invited to attend various social events, such as welcome reception, gala dinner and banquet, students and young professionals reception, women in computational intelligence meeting, etc. In addition, participants are also encouraged to explore the beautiful city of Niterói and its surroundings which have an endless supply of attractions and things to see and do.

We look forward to welcoming you in Niterói!

Tweets by @cbic2017
Bibliography I

New stopping criterion for genetic algorithms.

Benchmarking parameter-free amalgam on functions with and without noise.
*Evolutionary Computation*, 21(3):445–469.

A dominance–based stability measure for multi–objective evolutionary algorithms.

Springer International Publishing, Cham.

Evolutionary algorithms and elliptical copulas applied to continuous optimization problems.

Approximate KKT points and a proximity measure for termination.

Working group on EMO for interactive multiobjective optimization (1st round).
A non-dominance-based online stopping criterion for multi-objective evolutionary algorithms.

A stopping criterion based on Kalman estimation techniques with several progress indicators.

Introducing a robust and efficient stopping criterion for MOEAs.
In 2010 IEEE Conference on Evolutionary Computation (CEC), part of 2010 IEEE World Congress on Computational Intelligence (WCCI 2010), Piscataway, New Jersey. IEEE Press.

On termination criteria of evolutionary algorithms.

Minima of functions of several variables with inequalities as side constraints.
Master’s thesis, Department of Mathematics, University of Chicago, Chicago, IL, USA.

Nonlinear programming.
CMA-ES with restarts for solving CEC 2013 benchmark problems.
In *IEEE Congress on Evolutionary Computation (CEC)*, pages 369–376. IEEE.

SMS-EDA-MEC: Extending copula-based EDAs to multi-objective optimization.

Anomaly detection with the Voronoi diagram evolutionary algorithm.

VorAIS: A multi-objective Voronoi diagram-based artificial immune system.

Progressively adding objectives: A case study in anomaly detection.
In *Proceedings of the 2017 Annual Conference on Genetic and Evolutionary Computation (GECCO’17)*, pages 1–8, New York, NY, USA. ACM Press.
in press.

A cumulative evidential stopping criterion for multiobjective optimization evolutionary algorithms.
An approach to stopping criteria for multi–objective optimization evolutionary algorithms: The MGBM criterion.

A progress indicator for detecting success and failure in evolutionary multi–objective optimization.
In 2010 IEEE Conference on Evolutionary Computation (CEC), part of 2010 IEEE World Congress on Computational Intelligence (WCCI 2010), Piscataway, New Jersey. IEEE Press.

A steady performance stopping criterion for pareto–based evolutionary algorithms.
In The 6th International Multi–Objective Programming and Goal Programming Conference, Hammamet, Tunisia.

Convergence properties of some multi–objective evolutionary algorithms.

Entropy-based termination criterion for multiobjective evolutionary algorithms.

Stewart, T. J. (1997).
Convergence and validation of interactive methods in mcmd: simulation studies.
In Essays in Decision Making, pages 7–18. Springer.

A convergence criterion for multiobjective evolutionary algorithms based on systematic statistical testing.