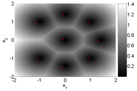
On Quality Indicators for Finite Level-Set Representations

Michael T. M. Emmerich, André H. Deutz, Johannes Kruisselbrink

LIACS, Natural Computing Group, Faculty of Science, Leiden Universiteit
Niels Bohrweg 1, 2333-CA Leiden, NL
http://natcomp.liacs.nl

EVOLVE 2011, Bourglinster Castle, Luxembourg, 26-May-2011



On Quality Indicators for Finite Level-Set Representations

Michael T. M. Emmerich, André H. Deutz, Johannes Kruisselbrink

LIACS, Natural Computing Group, Faculty of Science, Leiden Universiteit
Niels Bohrweg 1, 2333-CA Leiden, NL
http://natcomp.liacs.nl

EVOLVE 2011, Bourglinster Castle, Luxembourg, 26-May-2011

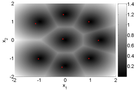
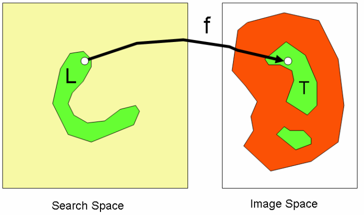


Table of Contents

- Consider the task of representing/approximating an implicitly defined compact set

$$L = \{x \in X | f(x) \in T\},$$

for instance level-sets, where T is a singleton.
- Consider continuous set that needs to be approximated by a finite set of feasible points (*representation set*).
- The computation of the indicator should be possible without explicit knowledge of the solution set L.



Example problems

Consider a black box system model $f(x) = y$ and a target set T:

- Find alternative molecules x with chemical properties within a certain user-defined range $T = [a, b]$.
- Find alternative solutions of an engineering problem that score y above a certain threshold τ, i.e. in a target set $T = [\tau, \infty)$.
- Find alternative causes x for a given effect $y = T$ using a computer model of the system (T is a singleton here).
- Classical: Find a level set of a function, e.g. $f : \mathbf{x} \mapsto x_1^2 + x_2^2 + sin(x_1 x_2)$, $L = \{\mathbf{x} | f(\mathbf{x}) = T\}$.

Overarching goal:

- We consider the level-set problem as a set-oriented optimization problem;
- In this paper we study unary indicator functions that assign a performance value to a (candidate) set of points.
- We are interested in indicators that do not require a-priori knowledge of the solution set, such as the Hausdorff distance.
- In particular, we envision an indicator that can be used for bounded archiving or selection in metaheuristics.

Desirable properties of a set indicator for finite level set representations

The following properties of quality indicators for representation sets we consider as desirable:

1. Representation sets that contain a large number of 'essentially' different points are more desirable.
2. Representation sets which are more 'evenly' spread are more desirable.

These two desired properties find their counterpart in the *counting* and *spread indicator*, introduced in the following.

Counting and Spread Indicator[1]

Definition

A set R is ϵ-disjoint, iff
$\forall x_1, x_2 \in R : x_1 \neq x_2 \Rightarrow B_\epsilon(x_1) \cap B_\epsilon(x_2) = \emptyset$, where $B_\epsilon(x)$ denotes the open ϵ-ball around x.

Definition

IC_ϵ (counting indicator): $\mathrm{IC}_\epsilon(R) = \max\{|C| \mid C \subseteq R \text{ and } C \text{ is } \epsilon\text{-disjoint}\}$

Definition

IS_N (spread indicator): Let N denote a fixed natural number and $|R| = N$. Then $\mathrm{IS}_N(R) = \sup\{\epsilon | \epsilon \in \mathbb{R} \text{ and } R \text{ is } \epsilon\text{-disjoint}\}$.

[1] We suspect that these indicators are already used in similar contexts and we are more interested in their conceptual comparison.

Non-Incremental Property and Example

Lemma

Let $q \in \mathbb{N}$ be such that $IC_N(R) = q$. Then it can occur that for some representation R_1 with $IC_N(R_1) < q$, there does not exist any representation set R_2 such that $R_1 \subset R_2$ and $IC_N(R_2) = q$.

Proof: Here is an example to support the statement: $L = [0, 1]$, $\epsilon = \frac{1}{2}$, $R = \{0, \frac{1}{2}, 1\}$, and $R_1 = \{\frac{1}{2}\}$. (Here $\mathrm{IC}_N(R) = |\{0, 1\}| = 2$.)

- $\Rightarrow$ No straightforward incremental algorithm for the computation of IC_N.
- IC_ϵ can be computed efficiently (Minimal distance between any two points (closest pair)[2]).

[2] time complexity: $\mathcal{O}(n^2)$ and in the plane $\Omega(n \log n)$ in the algebraic decision tree model of computation.

Average distance indicator

An attempt to integrate both of the desirable properties into one indicator gives rise to average distance oriented measures:

Definition

ID_X (Average distance indicator) Let $d(x, R)$ denote the distance of x to the nearest point in R and X denote a compact *reference space* that must include L. Then

$$\mathrm{ID}_X(R) = (1/\mathrm{Vol}(X)) \int_X d(x, R) \mathrm{d}\, x$$

.

Remark: This indicator is not the *average distance of points in the representation set*, which intuitively measures diversity. We are looking for another name of this, e.g. *Integrated Distance*.

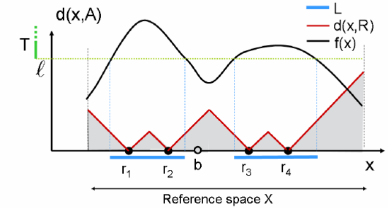


Fig. 1. ID_X for $X \subset \mathbb{R}$, $T = \{y \in \mathbb{R} | y > \ell\}$

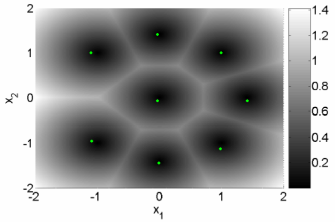


Fig. 2. The integrand $d(x, R)$ for ID_X as described in the 2-D example.

3-D Example for $f(x) = -x_1^2 - x_2^2 + 2\sqrt{x_1^2 + x_2^2}$, $T = \{0\}$ is plotted, where $R = \{(0,0), (-1,1), (0,\sqrt{2}), (1,1), (\sqrt{2},0), (1,-1), (0,-\sqrt{2}), (-1,-1)\}$, and $X = [-2,2]^2$.

Lemma

Given a reference set $X \supseteq L$ and d being a distance:

$$\arg\min_{R \subseteq L} \int_{x \in L} d(x, R) dx = \arg\min_{R \subseteq L} \int_{x \in X} d(x, R) dx$$

Remark 1 Lemma 6 shows that minimizing ID_X yields L. The knowledge of L is, however, not required for computing ID_X.
Remark 2 In general, for bounded size sets, the reference set X will influence the result. There exists X and L and $k > 1$ where

$$\arg\min_{\{R \subset L \mid |R| = k\}} \int_{x \in L} d(x, R) \mathrm{d}x \neq \arg\min_{\{R \subset L \mid |R| = k\}} \int_{x \in X} d(x, R) \mathrm{d}x$$

The resulting set can still be spread out.

Augmented Average Distance Indicator

- To guide the search to the feasible subspace we can use the augmented average distance indicator[3]:

$$I_X^+(R) = I_X(R \cap L) + \sum_{x \in R \setminus L} (d(f(x), T))$$

- We get the following property for $R' \subseteq L$:

$$I_X^+(R) \leq I_X^+(R') \Rightarrow I_X(R \cap L) \leq I_X(R')$$

- and whenever an infeasible solution is replaced by a feasible solution, the augmented indicator is improved.

$I_X(R) = 1/\mathrm{Vol}(X) \int_{x \in X} d(x, R) \mathrm{d}\, x$

[3] Note that $R \cap L$ as well as $R \setminus L$ can be computed using f, i.e. without knowing L.

Augmented Average Distance Indicator

We may ask for a stricter indicator with

$$R \subseteq L \text{ and } R' \not\subseteq L, then I_X^A(R) \leq I^A(R').$$

Lemma (Upper bound for average distance)

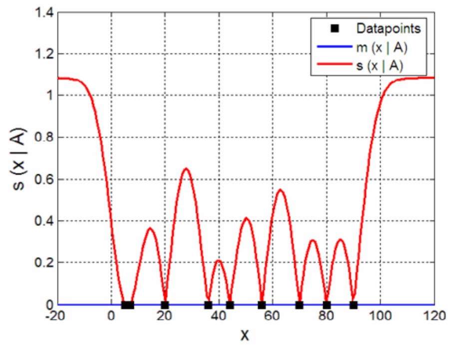
$$1/Vol(X) \int_{x \in X} d(x, R) dx \leq Diameter(X)$$

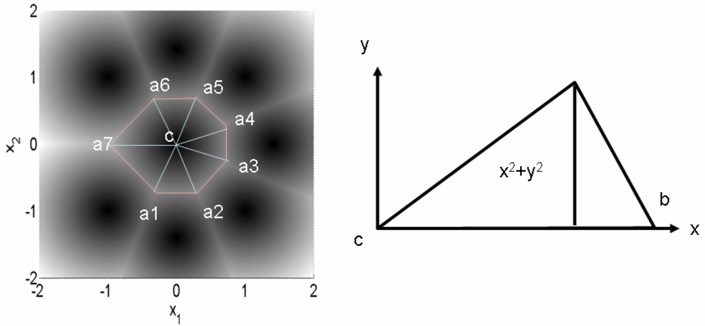
Remark A *penalized indicator function* can be constructed as follows: Let R denote a representation set containing infeasible solutions.

$$I_X^A(R) = \begin{cases} ID_X(R \cap L) + \sum_{x \in R \setminus L}(d(f(x), T)) \dots & \\ \qquad \dots + \text{Diameter}(X) & \text{if } R \cap L \neq \emptyset. \\ ID_X(R) & \text{otherwise.} \end{cases}$$

Average-uncertainty indicator

An indicator with similar properties is given by $\mathrm{IG}_X = \int_{x \in X} s(x|R) \mathrm{d}\, x$ where $s(x|R)$ denotes the local, conditional standard deviation of a zero mean, stationary Gaussian random field with fixed, positive definite covariance.






```
For each cell c in R:
  Triangulate(c) → (a1, a2, …,a7)
  For each triangle:
    (c,a,b)=RotateTriangle(c,a(i2),a(i3))
    v=v+IntegrateVolume(0,a,b)
```

Integration over a single triangle cell

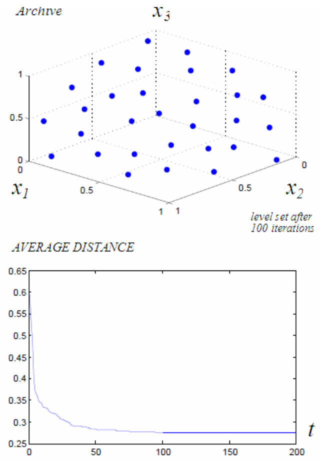
$$
\begin{aligned}
& \int_{\mathbf{x} \in \triangle(\mathbf{0},\mathbf{a},\mathbf{b})} x^2 + y^2 \qquad (1) \\
= \;& \int_{x=0}^{a_1} \int_{y=0}^{a2/a1x} x^2 + y^2 dxdy + \int_{x=0}^{a_2 - \frac{a_2}{(b_1 - a_1)} x} x^2 + y^2 dxdy \\
= \;& 1/12 a_1 a_2 (3(a_1)^2 + (a_2)^2) + \frac{1}{3}(-(a_1)^3 + (b_1)^3)(a_2 - \frac{a_2}{-a_1 + b_1}) \\
+ \;& \frac{1}{3}(-a_1 + b_1)(a_2 - a_2/(-a_1 + b_1))^3.
\end{aligned}
$$

Some remarks about Voronoi cells

- The number of Voronoi cell vertices in the plane is at most $2n - 5$
- For more than two dimensions (d) the number of vertices is bounded by $\mathcal{O}(n^{\lceil d/2 \rceil})$.
- The vertices can be computed in $\mathcal{O}(n \log n)$ optimal time in the plane, and $\mathcal{O}(n^{\lceil d/2 \rceil})$ time for $d > 2$ (Algorithm by Klee).

references: M. de Berg, M. van Kreveld, M. Overmars, O. Schwarzkopf: Computational Geometry: Algorithms and Applications, Second Edition, Springer, 2000

Algorithm



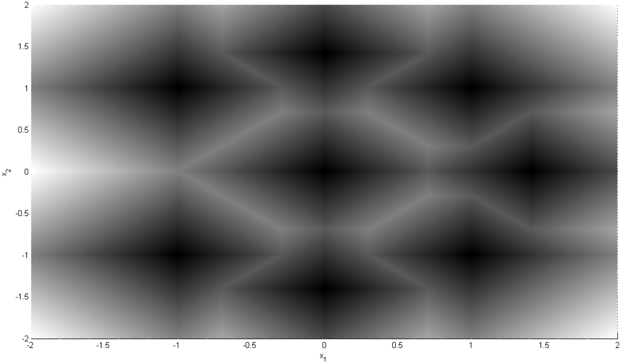
- Function

$$(x_1)^2 + (x_2)^2 + (x_3)^2 = 1,$$

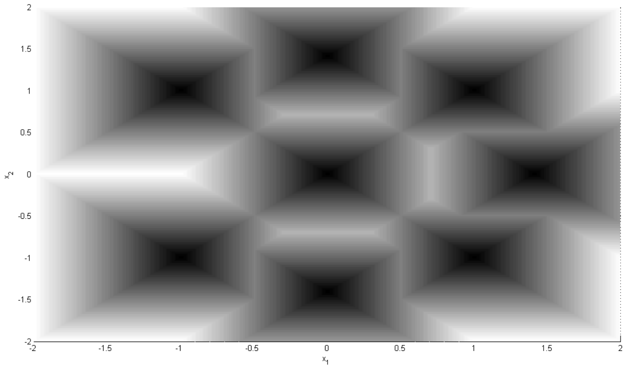
$$x \in [0, 1]^3$$

- Points are generated one by one and archived if they improve average distance indicator.
- Only feasible points are archived.
- Archive size is bounded.
- SUPPORT: `www.liacs.nl/~emmerich`

Voronoi Diagram for Manhattan distance



Voronoi Diagram for Tchebycheff distance



Summary

- Progress indicators with no a-priori knownledge of target set are searched for.
- Properties of counting and spread indicators have been contrasted to each other.
- Average-distance indicators combines favorable properties.
- Augmented average distance can be used to guide search in infeasible domain.
- Efficient computation need to be worked out.
- Future work items: Efficient computation, influence of reference set, Integration in metaheuristics/archivers.

Outlook

- Comparison to related work:
  - Tamara Ulrich, Johannes Bader, Eckart Zitzler: Integrating decision space diversity into hypervolume-based multiobjective search. GECCO 2010: 455-462
  - Tamara Ulrich, Johannes Bader, Lothar Thiele: Defining and Optimizing Indicator-Based Diversity Measures in Multiobjective Search. PPSN (1) 2010: 707-717
  - O. Schütze, X. Esquivel, A. Lara, and C. Coello Coello. Some Comments on GD and IGD and Relations to the Hausdorff Distance. GECCO 2010 Workshop on Theoretical Aspects of Evolutionary Multiobjective Optimization.
  - Oliver Schütze and Günter Rudolph: Average Hausdorff Distance (GECCO 2010, ...), Bounded Archiving for decision space diversity (PPSN2008)
- Implementation of tools; MATLAB demo in support material: `www.liacs.nl/~emmerich`