CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering
Department of Cybernetics

Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions

Petr Pošík and Petr Baudiš

Introduction

Background

Consider the following optimization task of bounded separable functions, i.e.

$$\begin{aligned}
&\text{minimize } f(\mathbf{x}) && (1)\\
&\text{subject to } L_i \leq x_i \leq U_i \text{ for } i = 1, \ldots, D,\\
&\text{where } f(\mathbf{x}) = a_1 f_1(x_1) + \ldots + a_D f_D(x_D).
\end{aligned}$$

Problem not studied very often:

- Real-world problems are only seldom separable.
- Common belief that problems like (1) can be easilly solved by decomposing them to D univariate problems and solving them one by one.

Yet, researchers find it useful to make such methods part of their hybrid algorithm [LSS13] or algorithmic portfolio [BMTP12]), as a safeguard against separable problems.

[BMTP12] Bernd Bischl, Olaf Mersmann, Heike Trautmann, and Mike Preuss. Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pages 313–320, New York, NY, USA, 2012. ACM.

[LSS13] Ilya Loshchilov, Marc Schoenauer, and Michele Sèbag. Bi-population CMA-ES algorithms with surrogate models and line searches. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, pages 1177–1184, New York, NY, USA, 2013. ACM.

Introduction
- Background
- Issues
- Contents

BS Review

Dim. Selection

Results

Summary

Issues solved (?)

Even if we *decompose* the D-dimensional problem *to D independent univariate problems*, we face the following issues:

1. Which univariate solver shall one choose?
   - Shall we choose *a quickly converging local search method*, or
   - *a slower global search method*?
2. What stopping conditions shall one choose for the individual univariate solvers?
   - We *cannot use stopping conditions based on the acceptable quality of the candidate solution* as a whole.
   - Fixed budgets or stagnation detection lead to wasted resources or missed optima.

Issues solved (?)

Even if we *decompose* the D-dimensional problem *to D independent univariate problems*, we face the following issues:

1. Which univariate solver shall one choose?
   - Shall we choose *a quickly converging local search method*, or
   - *a slower global search method*?
2. What stopping conditions shall one choose for the individual univariate solvers?
   - We *cannot use stopping conditions based on the acceptable quality of the candidate solution* as a whole.
   - Fixed budgets or stagnation detection lead to wasted resources or missed optima.

Recently proposed **Brent-STEP algorithm** [BP15] solves these issues to a great extent:

1. It *combines fast local search with slower global search*: *we do not have to choose* one of them.
2. It *interleaves the steps of the univariate solvers* such that all dimensions are optimized "concurrently", in a similar spirit it is done in Rosenbrock's algorithm [Ros60], or in Loshchilov's HCMA [LSS13]. *We do not need to specify the stopping conditions for each univariate solver*, only for the algorithm as a whole.

[BP15] Petr Baudiš and Petr Pošík. Global line search algorithm hybridized with quadratic interpolation and its extension to separable functions. In *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*, New York, NY, USA, 2015. ACM.

[LSS13] Ilya Loshchilov, Marc Schoenauer, and Michele Sèbag. Bi-population CMA-ES algorithms with surrogate models and line searches. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, pages 1177–1184, New York, NY, USA, 2013. ACM.

[Ros60] H. H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, March 1960.

Contents

How is the dimension interleaving done in multivariate Brent-STEP [BP15]?

- Dimensions (individual univariate solvers) are chosen uniformly, round-robin.
- However, some dimensions may be easier to optimize/bringing higher profit.

Contents

How is the dimension interleaving done in multivariate Brent-STEP [BP15]?

- Dimensions (individual univariate solvers) are chosen uniformly, round-robin.
- However, some dimensions may be easier to optimize/bringing higher profit.

Goals of this work:

- Is there a way how to choose the dimensions smarter?
- How large gain can we expect?

Contents

How is the dimension interleaving done in multivariate Brent-STEP [BP15]?

- Dimensions (individual univariate solvers) are chosen uniformly, round-robin.
- However, some dimensions may be easier to optimize/bringing higher profit.

Goals of this work:

- Is there a way how to choose the dimensions smarter?
- How large gain can we expect?

Contents:

1. Brent-STEP review
2. Dimension selection methods
3. Experimental comparison on BBOB testbed

[BP15] Petr Baudiš and Petr Pošík. Global line search algorithm hybridized with quadratic interpolation and its extension to separable functions. In *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*, New York, NY, USA, 2015. ACM.

Brent-STEP Algorithm Review

Introduction
BS Review
- Brent and STEP
- Brent-STEP, 1D
- Brent-STEP, ND

Dim. Selection
Results
Summary

Constituent algorithms

STEP [LSB94]:

- Global line search algorithm.
- Iteratively divides the domain into intervals, splitting one of them into halves.
- Chooses the interval for which it seems the easiest to improve the current best-so-far (BSF) solution by sampling from the respective interval.

[LSB94] S. Langerman, G. Seront, and H. Bersini. S.T.E.P.: The Easiest Way to Optimize a Function. In *IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 519–524 vol.1, 1994.

Constituent algorithms

STEP [LSB94]:

- Global line search algorithm.
- Iteratively divides the domain into intervals, splitting one of them into halves.
- Chooses the interval for which it seems the easiest to improve the current best-so-far (BSF) solution by sampling from the respective interval.

[LSB94] S. Langerman, G. Seront, and H. Bersini. S.T.E.P.: The Easiest Way to Optimize a Function. In *IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 519–524 vol.1, 1994.

Brent's method [Bre73]:

- Local line search algorithm.
- Combines golden section search with quadratic interpolation.
- Each iteartion:
  1. Having triple of points bracketing the optimum, estimate the position of the optimum using quadratic interpolation.
  2. If it satisfies certain criteria, sample that point,
  3. otherwise sample the point resulting from the golden section step.
  4. Update the bracketing triple of points.

[Bre73] Richard P. Brent. *Algorithms for Minimisation Without Derivatives*. Prentice Hall, 1973.

Brent-STEP Hybrid, Univariate

Univariate Brent-STEP method (simplified):

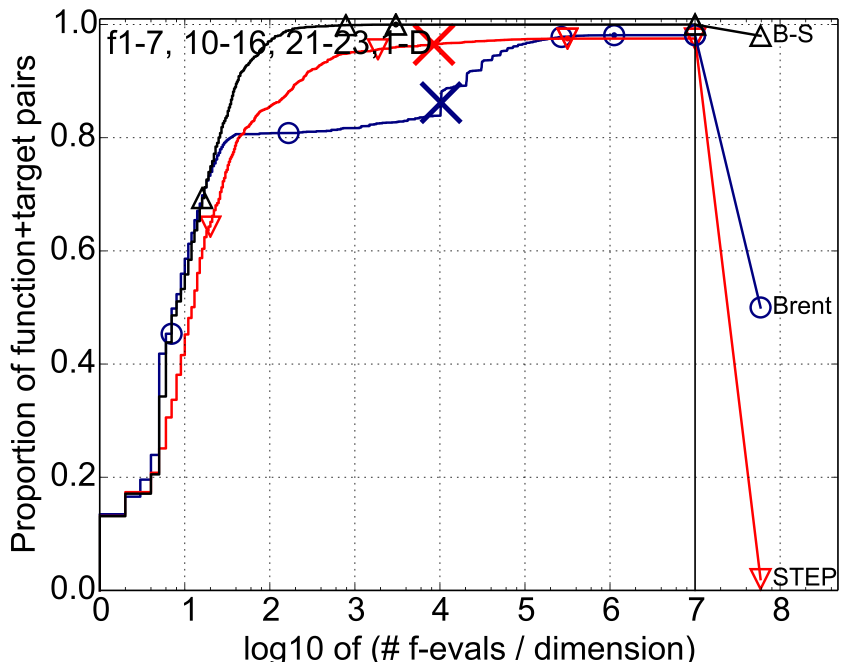
1. Among all triples of points bracketing any optimum, choose the most promising triple (quadratic interpolation).
2. If the estimated minimum on that part of function improves BSF solution *by a non-trivial amount*, use a single iteration of Brent to sample new point.
3. Otherwise, use a single iteration of STEP to split the easiest interval.
4. Update BSF and algorithm state using the new sampled point.

Brent-STEP Hybrid, Univariate

Univariate Brent-STEP method (simplified):

1. Among all triples of points bracketing any optimum, choose the most promising triple (quadratic interpolation).
2. If the estimated minimum on that part of function improves BSF solution *by a non-trivial amount*, use a single iteration of Brent to sample new point.
3. Otherwise, use a single iteration of STEP to split the easiest interval.
4. Update BSF and algorithm state using the new sampled point.

Brent-STEP on *univariate functions* (uni- and multimodal):



Properties:

- *Greedy behavior:* makes a local search step, if possible.
- Simple but effective combination: usually takes the best of both worlds.

Brent-STEP Hybrid, Multivariate

Multivariate generalization (simplified): Interleave the dimensions!

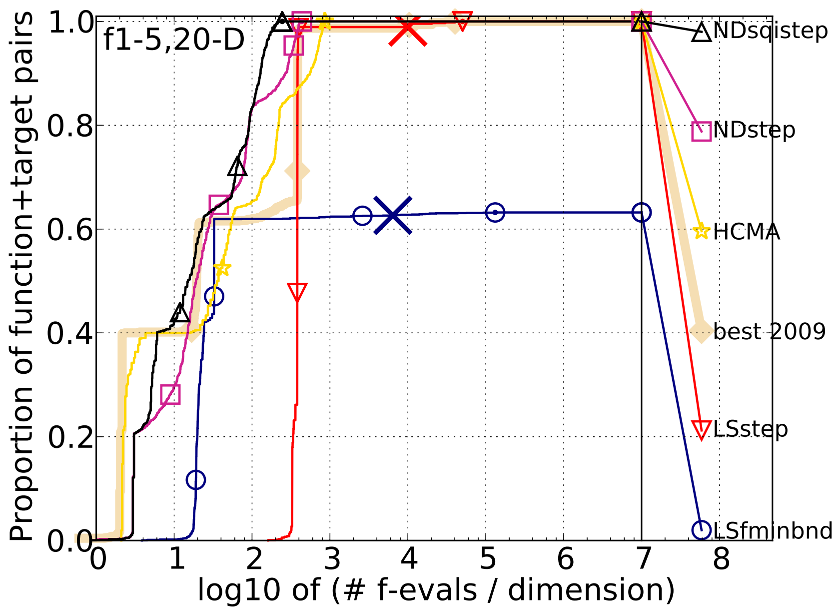
1. Choose a random point, make it the BSF solution.
2. While not happy:
   - Choose dimension (round-robin).
   - Perform a single step of univariate BS in the chosen dimension, taking the BSF solution as context.
   - If better than BSF solution found, update BSF and inform all univariate solvers.

Brent-STEP Hybrid, Multivariate

Multivariate generalization (simplified): Interleave the dimensions!

1. Choose a random point, make it the BSF solution.
2. While not happy:
   - Choose dimension (round-robin).
   - Perform a single step of univariate BS in the chosen dimension, taking the BSF solution as context.
   - If better than BSF solution found, update BSF and inform all univariate solvers.

Multivariate Brent-STEP on *multivariate separable functions*:



Properties:

- Works effectively for separable functions only.
- For non-separable. . . see the results later.

More details on the Brent-STEP algorithm: in [BP15] or in the presentation of the paper in the main CO track.

[BP15] Petr Baudiš and Petr Pošík. Global line search algorithm hybridized with quadratic interpolation and its extension to separable functions. In *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*, New York, NY, USA, 2015. ACM.

Dimension Selection Strategies

Motivation

Multivariate BS uses *round-robin strategy* for interleaving the dimensions:

- Budget distributed evenly among all dimensions.
- However, some dimensions may bring bigger profit than others.
- Could we use a method that would distribute the budget unevenly, concentrating on the dimensions with bigger profit?

Motivation

Multivariate BS uses *round-robin strategy* for interleaving the dimensions:

- Budget distributed evenly among all dimensions.
- However, some dimensions may bring bigger profit than others.
- Could we use a method that would distribute the budget unevenly, concentrating on the dimensions with bigger profit?

Warning! Easy to spoil the whole method! Example: consider mere STEP with dimension interleaving:

- In each dimension, the interval for splitting is chosen on the basis of the interval difficulty.
- Natural extension: choose the interval for splitting using all intervals in all dimensions.
- This does not work at all! The interval difficulties are not comparable across dimensions. The easiest intervals tend to be in the dimensions bringing the lowest profit. :-(
- Careful approach is needed.

Dimension Selection Methods

We compared the following dim. selection strategies:

- Round-robin [RR]: choose the dimensions one by one.
- Improvement frequency [IF]:
  - Track how often the BSF solution is improved using individual dimensions.
  - Each dimension is described by the relative frequency of improvement estimated using exponentially weighted moving average (EWMA).
  - Always choose the dimension with the best IF.
- Epsilon-greedy strategy with IF [IFEG]:
  - Each dimension evaluated by the relative IF.
  - Choose a random dimension in 100ϵ % of cases, otherwise choose the dimension with the best IF.
- Quadratic interpolation [QI]:
  - Used only in case of Brent-STEP.
  - The estimated minima arising from quadratic interpolation are comparable across dimensions.
  - Make Brent's iteration in the dimension with the most improving estimate of the minimum.
  - If no such dimension exists, use STEP with the round-robin strategy.

Introduction
BS Review
Dim. Selection
- Motivation
- Selection methods
- Compared Algorithms

Results
Summary

Dimension Selection Methods

We compared the following dim. selection strategies:

- Round-robin [RR]: choose the dimensions one by one.
- Improvement frequency [IF]:
  - Track how often the BSF solution is improved using individual dimensions.
  - Each dimension is described by the relative frequency of improvement estimated using exponentially weighted moving average (EWMA).
  - Always choose the dimension with the best IF.
- Epsilon-greedy strategy with IF [IFEG]:
  - Each dimension evaluated by the relative IF.
  - Choose a random dimension in 100ϵ % of cases, otherwise choose the dimension with the best IF.
- Quadratic interpolation [QI]:
  - Used only in case of Brent-STEP.
  - The estimated minima arising from quadratic interpolation are comparable across dimensions.
  - Make Brent's iteration in the dimension with the most improving estimate of the minimum.
  - If no such dimension exists, use STEP with the round-robin strategy.

Other portfolio strategies (MetaMax, Upper Confidence Bounds, . . .) for dimension selection were tested with disappointing results (not shown here).

Compared Algorithms

The capital letters denote the univariate solver (STEP or Brent-STEP), the lowercase letters denote the dimension selection strategy.

	STEP	Brent-STEP
Round-robin [RR]	Srr	BSrr
Improvement frequency [IF]	Sif	BSif
Epsilon-greedy with IF [IFEG]	Sifeg	BSifeg
Quadratic interpolation [QI]	—	BSqi

Compared Algorithms

The capital letters denote the univariate solver (STEP or Brent-STEP), the lowercase letters denote the dimension selection strategy.

	STEP	Brent-STEP
Round-robin [RR]	Srr	BSrr
Improvement frequency [IF]	Sif	BSif
Epsilon-greedy with IF [IFEG]	Sifeg	BSifeg
Quadratic interpolation [QI]	—	BSqi

Note: **Srr** is the same algorithm denoted as NDstep in [BP15], and (almost) the same as HCMA [LSS13] with NEWUOA and CMA-ES parts switched off.

[BP15] Petr Baudiš and Petr Pošík. Global line search algorithm hybridized with quadratic interpolation and its extension to separable functions. In *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*, New York, NY, USA, 2015. ACM.

[LSS13] Ilya Loshchilov, Marc Schoenauer, and Michele Sèbag. Bi-population CMA-ES algorithms with surrogate models and line searches. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, pages 1177–1184, New York, NY, USA, 2013. ACM.

Compared Algorithms

The capital letters denote the univariate solver (STEP or Brent-STEP), the lowercase letters denote the dimension selection strategy.

	STEP	Brent-STEP
Round-robin [RR]	Srr	BSrr
Improvement frequency [IF]	Sif	BSif
Epsilon-greedy with IF [IFEG]	Sifeg	BSifeg
Quadratic interpolation [QI]	—	BSqi

Note: **Srr** is the same algorithm denoted as NDstep in [BP15], and (almost) the same as HCMA [LSS13] with NEWUOA and CMA-ES parts switched off.

[BP15] Petr Baudiš and Petr Pošík. Global line search algorithm hybridized with quadratic interpolation and its extension to separable functions. In *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*, New York, NY, USA, 2015. ACM.

[LSS13] Ilya Loshchilov, Marc Schoenauer, and Michele Sèbag. Bi-population CMA-ES algorithms with surrogate models and line searches. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, pages 1177–1184, New York, NY, USA, 2013. ACM.

Parameters:

- The damping factor in EWMA is 0.9.
- The ϵ-greedy strategy uses $\epsilon = 0.5$.
- The size of a *non-trivial improvement* was set to 10^{-8}.
- *Burn-in phase:* the first $4D$ evaluations, round-robin was always applied.
- All methods are restarted if an improvement is not found for 2000 iterations.

Results

Results for unconstrained budget scenario

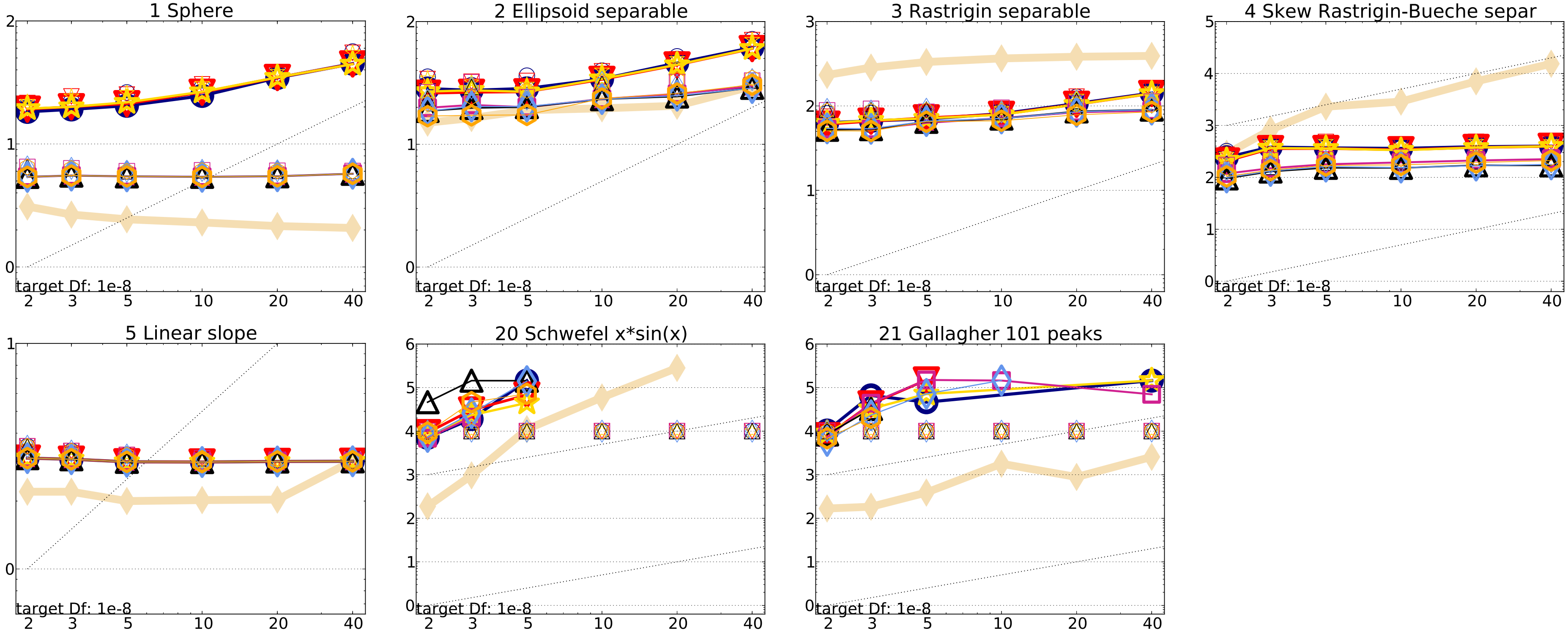
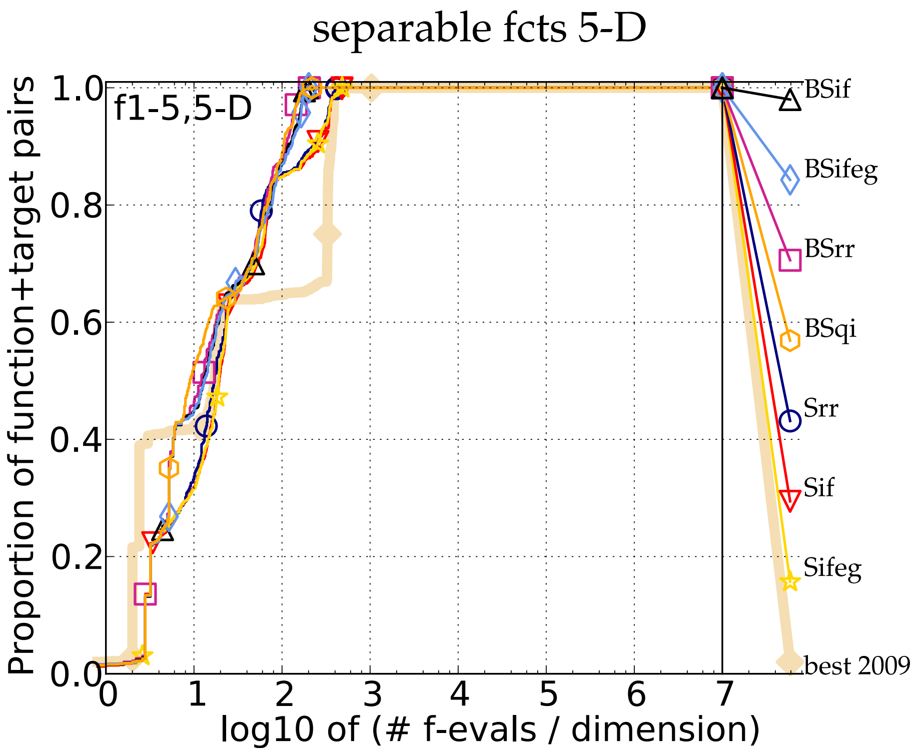
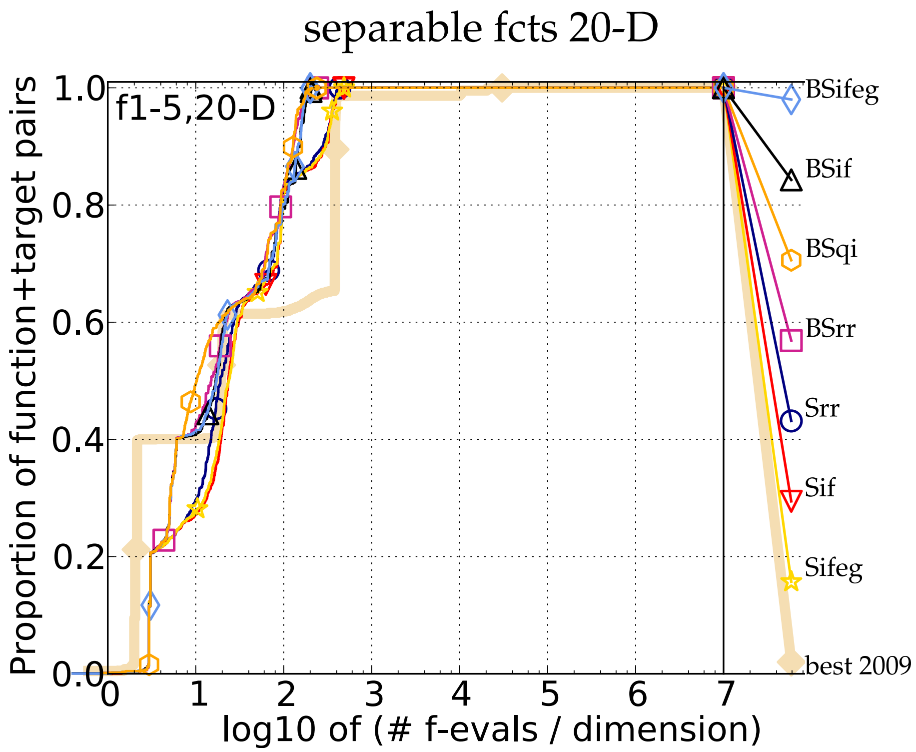


Figure 1: ○:Srr, ▽:Sif, ⋆:Sifeg, □:BSrr, △:BSif, ◇:BSifeg, ⬡:BSqi

- Target f-value is 10^{-8}.
- Graphs for non-separable problems show virtually nothing (with the exception of f20 and f21).
- Graphs for funcs 1, 2, and 4 (and 3 to a lesser extent) show 2 groups of methods: the better Brent-STEP familly and the worse STEP family.
- The best results obtained on funcs 3 and 4: separable, multimodal.

Results for unconstrained budget scenario (cont.)





Low budget, separable funcs

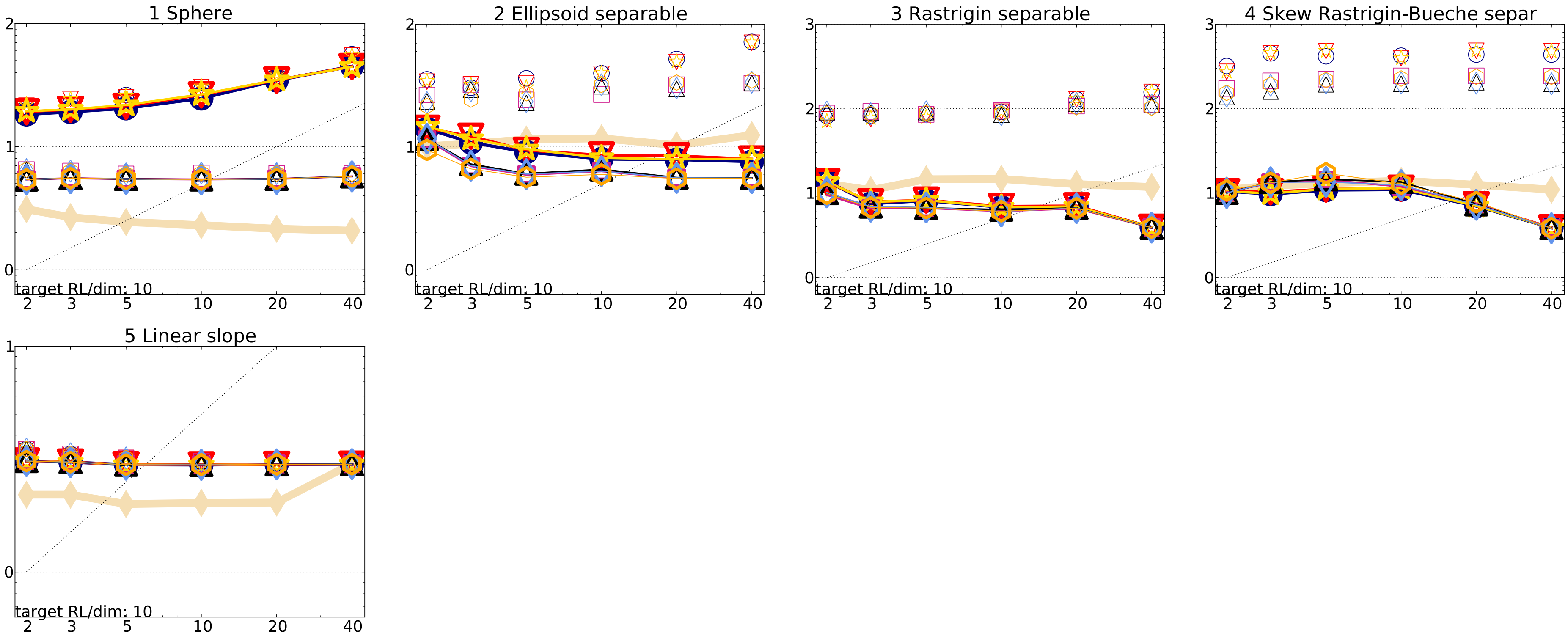


Figure 2: ○:Srr, ▽:Sif, ⋆:Sifeg, □:BSrr, △:BSif, ◇:BSifeg, ⬡:BSqi

- Target f-value determined relatively to performances observed in the past.
- For separable functions, results are similar to the unlimited budget case, maybe only less pronounced.

Low budget, non-separable funcs

Examples of funcs where STEP and BS are way behind other methods:

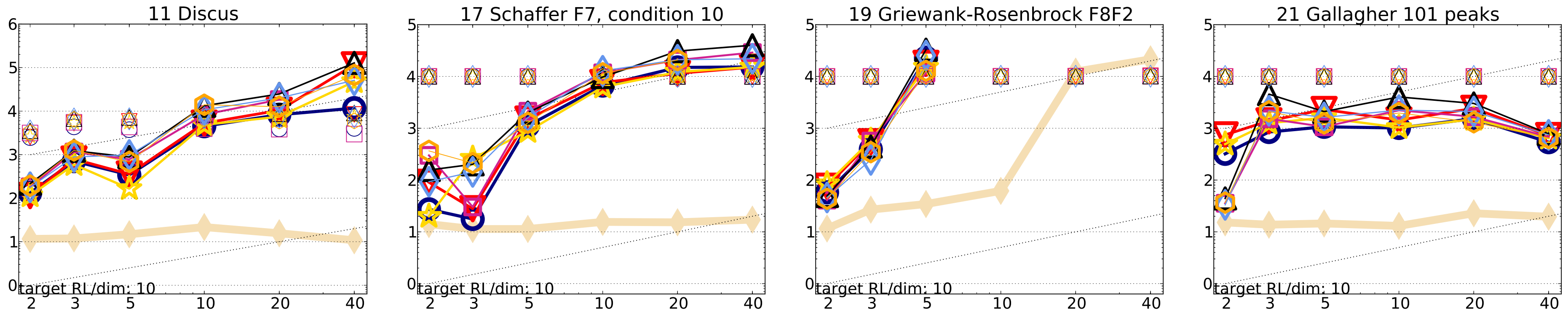


Figure 3: ○:Srr, ▽:Sif, ⋆:Sifeg, □:BSrr, △:BSif, ◇:BSifeg, ⬡:BSqi

Low budget, non-separable funcs

Examples of funcs where STEP and BS are way behind other methods:

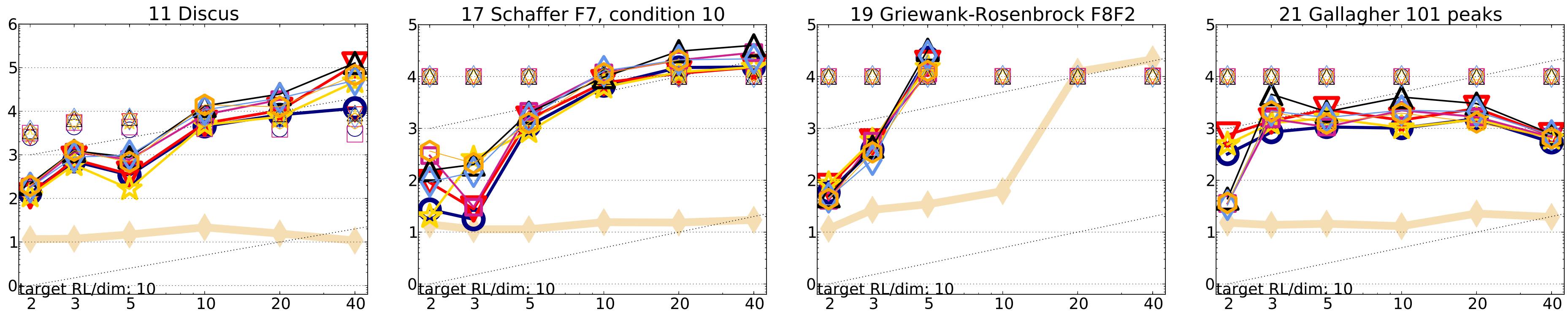


Figure 3: ○:Srr, ▽:Sif, ⋆:Sifeg, □:BSrr, △:BSif, ◇:BSifeg, ⬡:BSqi

But there are also non-separable funcs where STEP and BS are not that much worse:

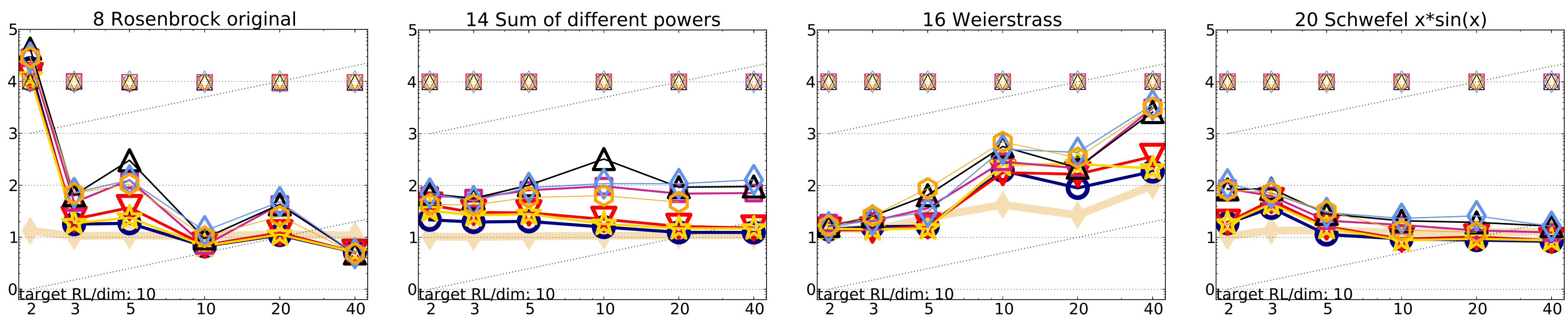
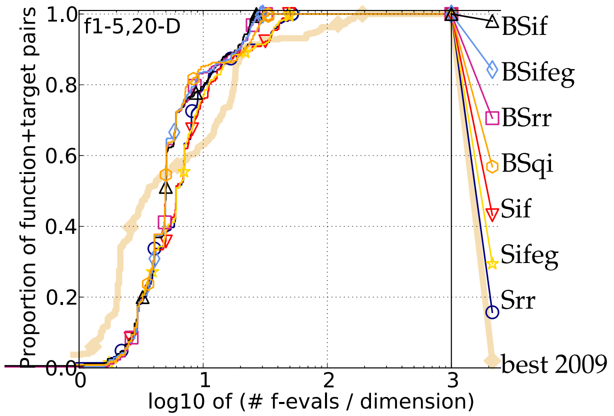


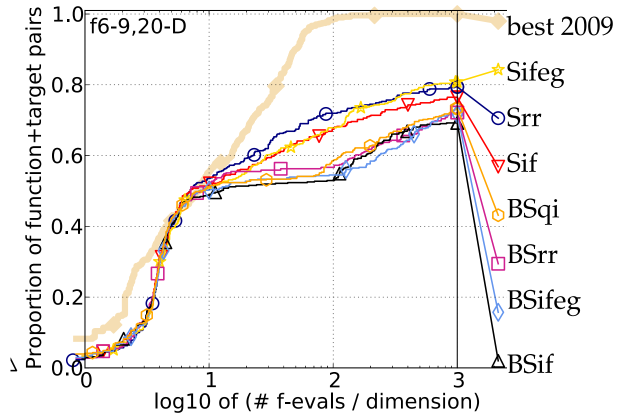
Figure 4: ○:Srr, ▽:Sif, ⋆:Sifeg, □:BSrr, △:BSif, ◇:BSifeg, ⬡:BSqi

Low budget, function groups

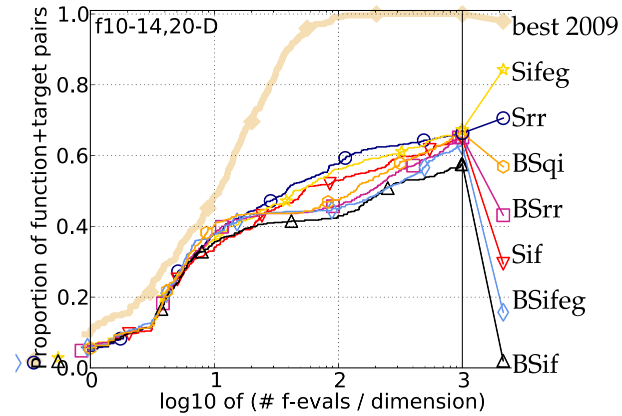
separable fcts



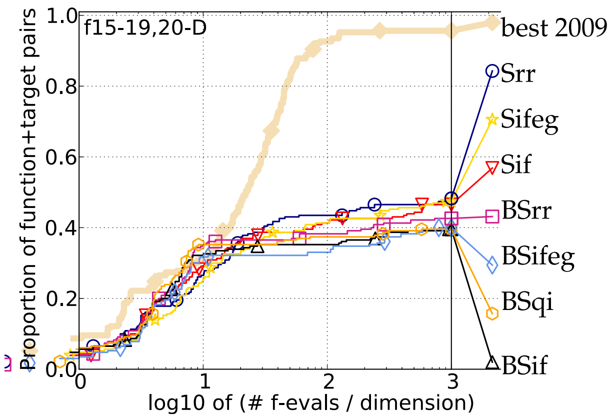
moderate fcts



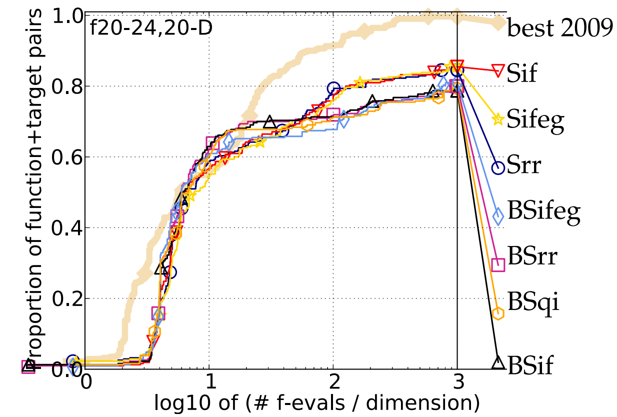
ill-conditioned fcts



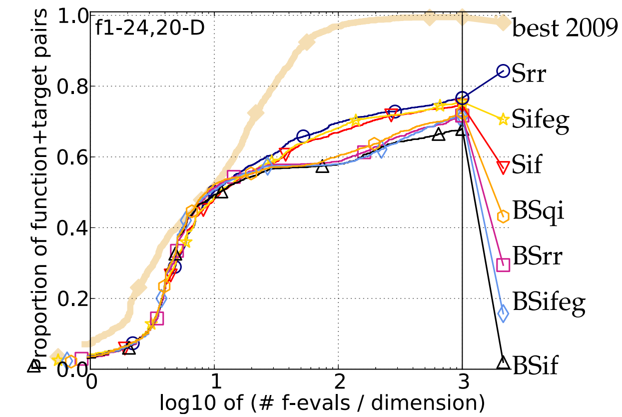
multi-modal fcts



weakly structured multi-modal fcts



all functions



- Sweet spot between, say $5D$ to $20D$ evaluations.
  - For loose target levels, even non-separable/multimodal functions may “look separably”.
  - In the beginning, more capable methods do not have enough info to show their potential.
- For non-separable functions, Brent’s component harms the algorithm.

Summary

Summary

Brent's component helps for searable functions.

- It harms for non-separable ones, but this algorithm is not primarily aimed at those.

Summary

Brent's component helps for searable functions.

- It harms for non-separable ones, but this algorithm is not primarily aimed at those.

Round-robin dimension selection was the default. Is there a better strategy?

- We do not know. We did not find any significantly better strategy. (This does not mean that it does not exist.)
- Observation: more complex strategies have higher chance to spoil the algorithm.

Summary

Brent's component helps for searable functions.

- It harms for non-separable ones, but this algorithm is not primarily aimed at those.

Round-robin dimension selection was the default. Is there a better strategy?

- We do not know. We did not find any significantly better strategy. (This does not mean that it does not exist.)
- Observation: more complex strategies have higher chance to spoil the algorithm.

Our recommendation: If you you want to safeguard your portfolio or hybrid algorithm against separable problems, *use the interleaved Brent-STEP algorithm with round-robin dimension selection strategy*.

Thank you!

Acknowledgements:

- This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/194/OHK3/3T/13.
- The authors also greatly appreciate the access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005).

Questions?