Empirical Performance Assessment

Black Box Optimization Benchmarking

Black-Box Optimization Benchmarking

$\longrightarrow$ assert / compare quantitatively performances

- $\longrightarrow$ gain some knowledge on algorithms
- $\longrightarrow$ select algorithms
- $\longrightarrow$ help to design new algorithms

Overview

→ Test functions

→ Experimental setup

→ How to measure performance

Test functions

- Many real world problems share common difficulties:

  → non separability (correlations between variables)

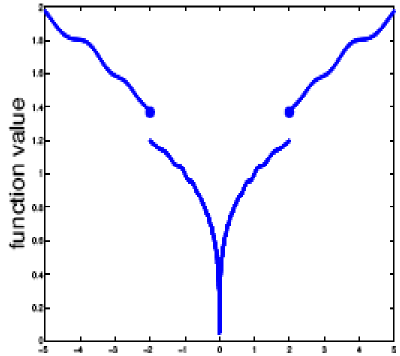
  → ill-conditioned (certain direction steeper than others),

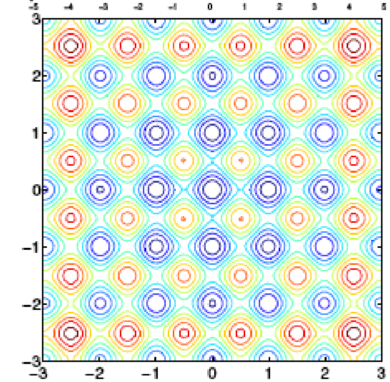
  → ruggedness (noise, ...),

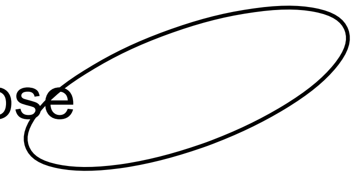
  → multi-modality

  → non-convexity

  Ideally an optimizer should cope with all of them



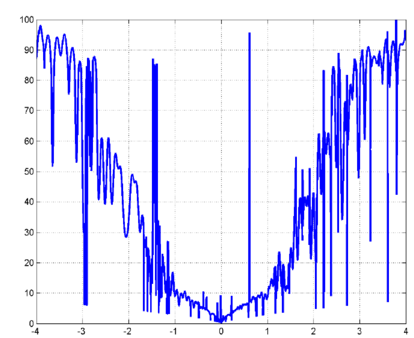




- Test functions chosen to assert on the performances w.r.t. those difficulties

  → non-noisy testbed (24 noise-free functions)

  → noisy testbed (30 noisy functions)



- Scalable with search space dimension
- Not too easy to solve but yet comprehensible

Noiseless Functions

24 functions within five sub-groups

→ Separable functions

→ Low or moderate conditioned functions

→ Ill-conditioned functions

→ Multimodal structured functions

→ Multimodal function with weak or without structure

Optimum of each function sampled uniformly in [-4 4]^d

For non separable functions, rotation sampled randomly for the different instances

The Noisy Functions

three noise-”models”, so-called:

- Gauss, Uniform (severe), Cauchy (outliers)

30 functions with three sub-groups

- 2x3 functions with weak noise
- 5x3 unimodal functions severe noise
- 3x3 multimodal functions severe noise

Experimental Setup

→ For each function and each dimension d(=2;3;5;10;20;40), 15 trials are carried out on 15 different function instances

Input to the algorithms:

→ search space dimension d

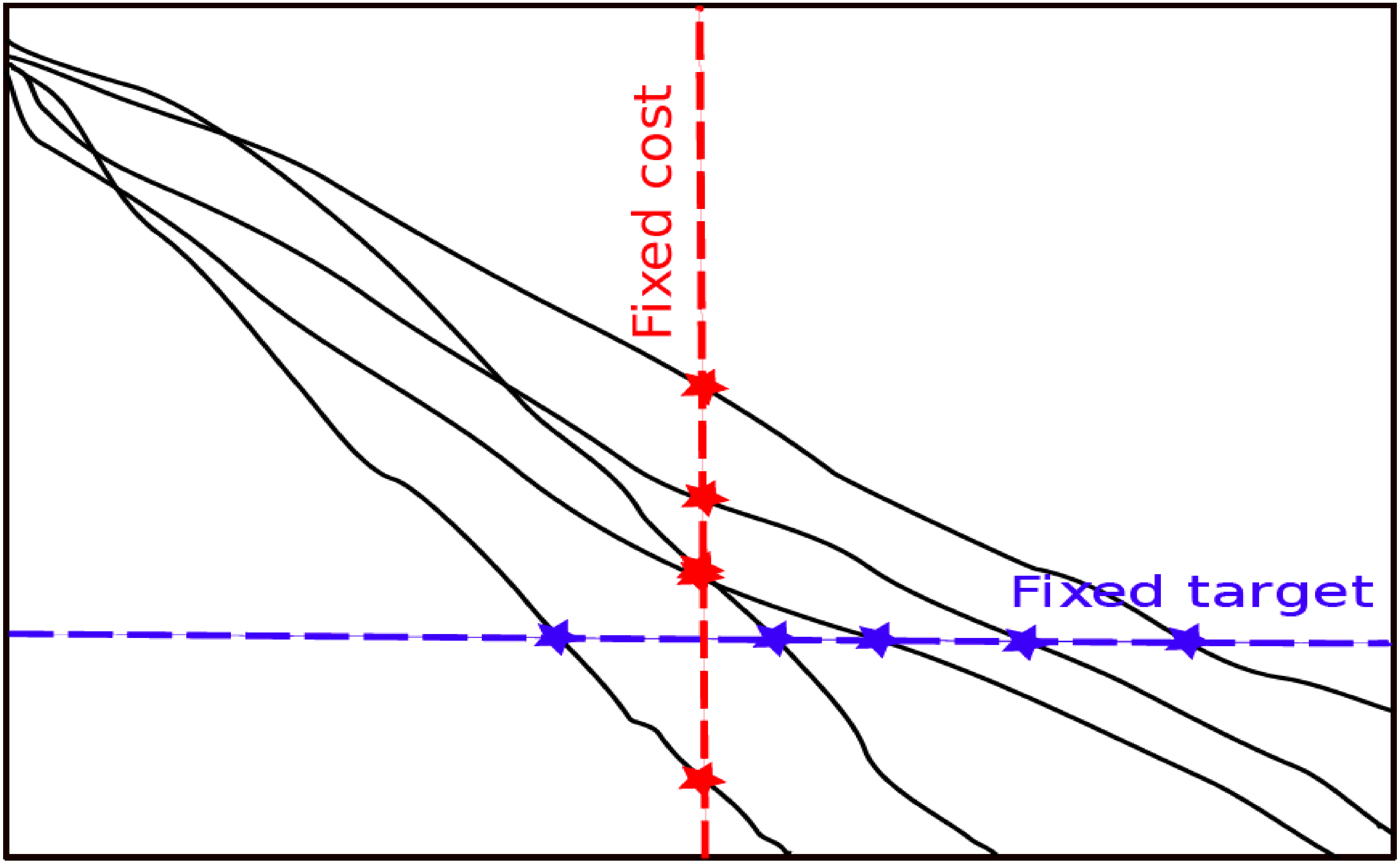
→ search domain [-5, 5]^d

→ testbed under consideration (noise-free / noisy)

[→ target function value, only for termination criteria]

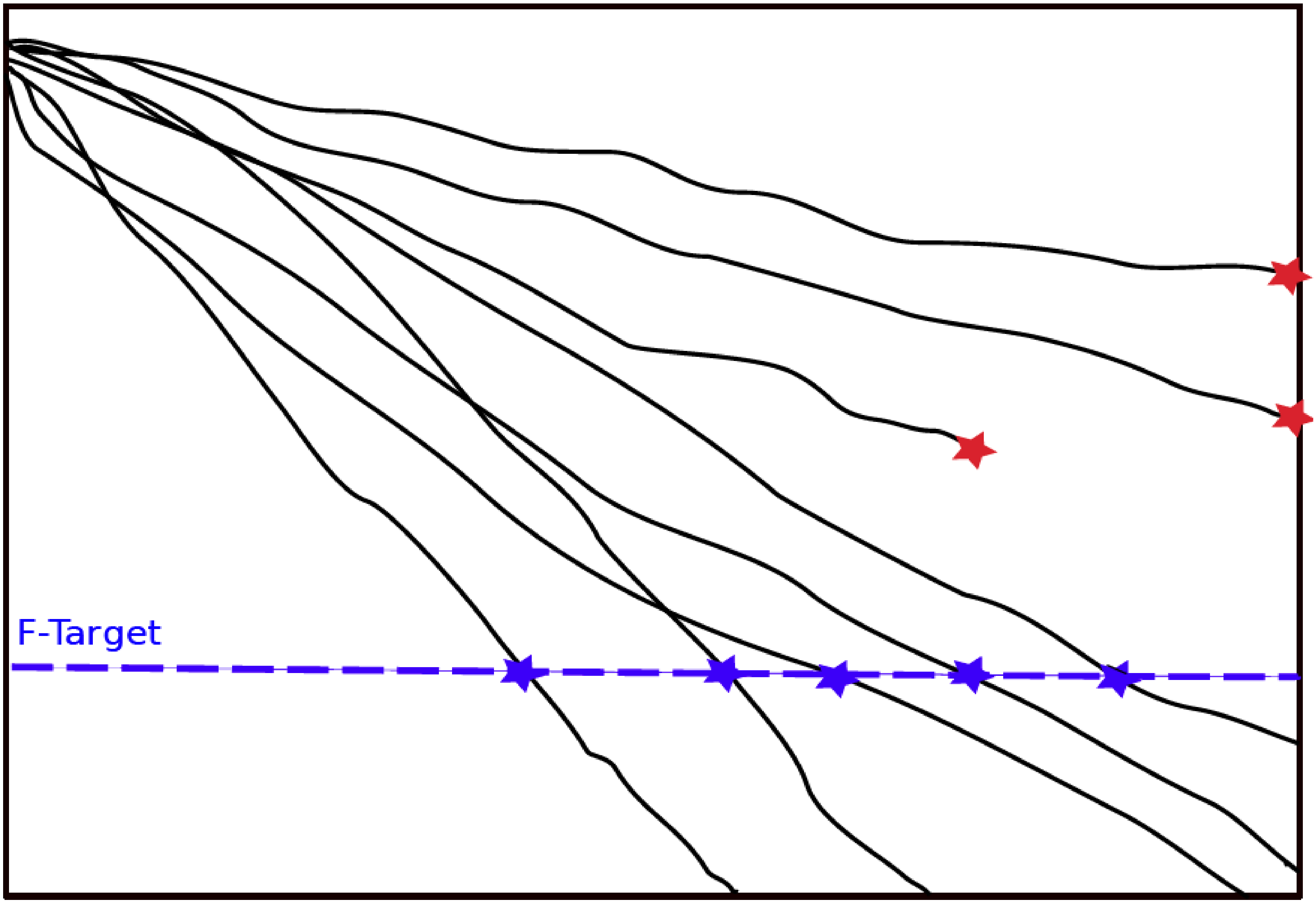
How to measure performance?

fixed-cost **fixed-target**

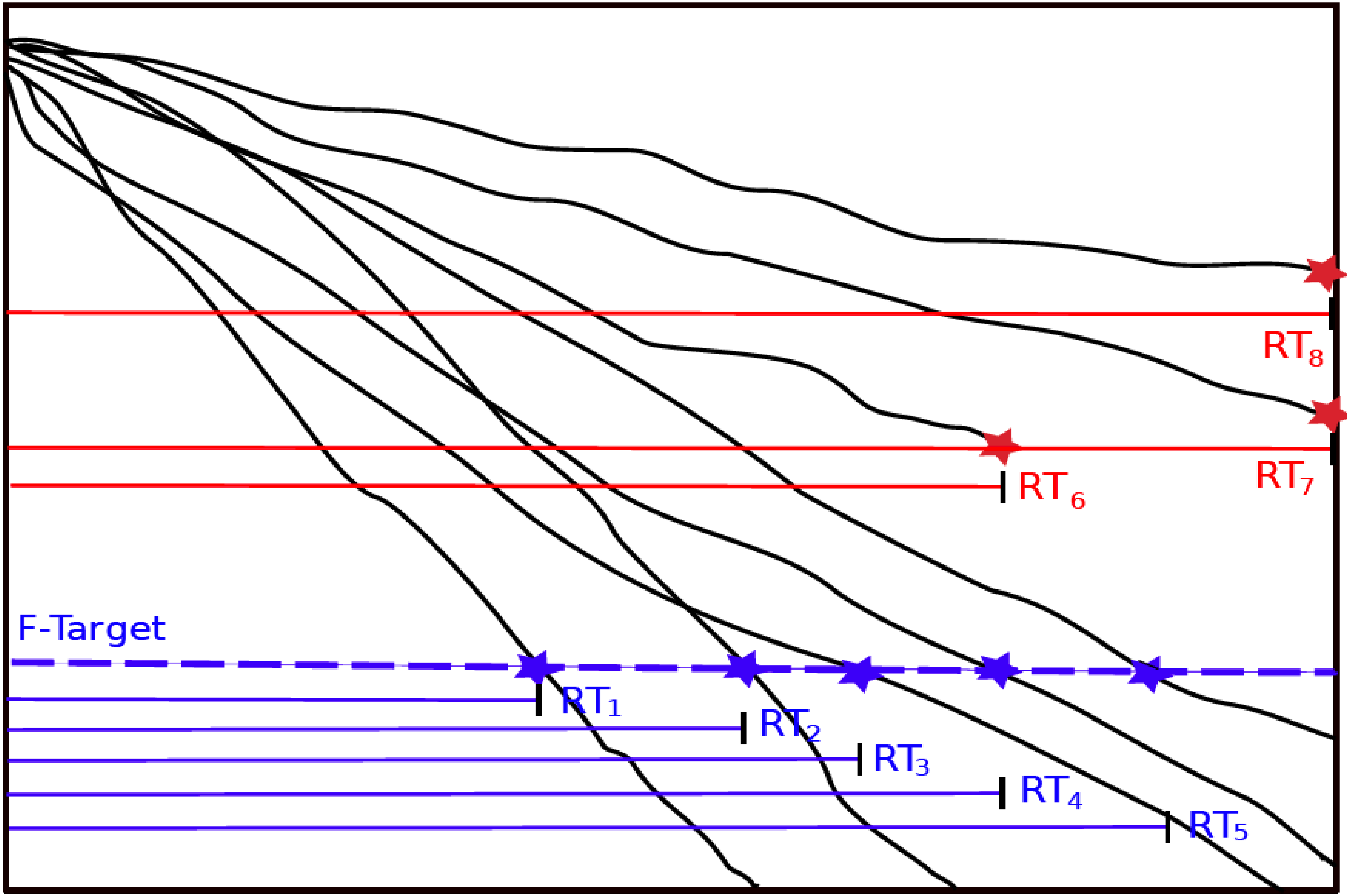


Black curves: best achieved function value versus number of function evaluations (time)

Collect Running Time (RT) to reach F-target

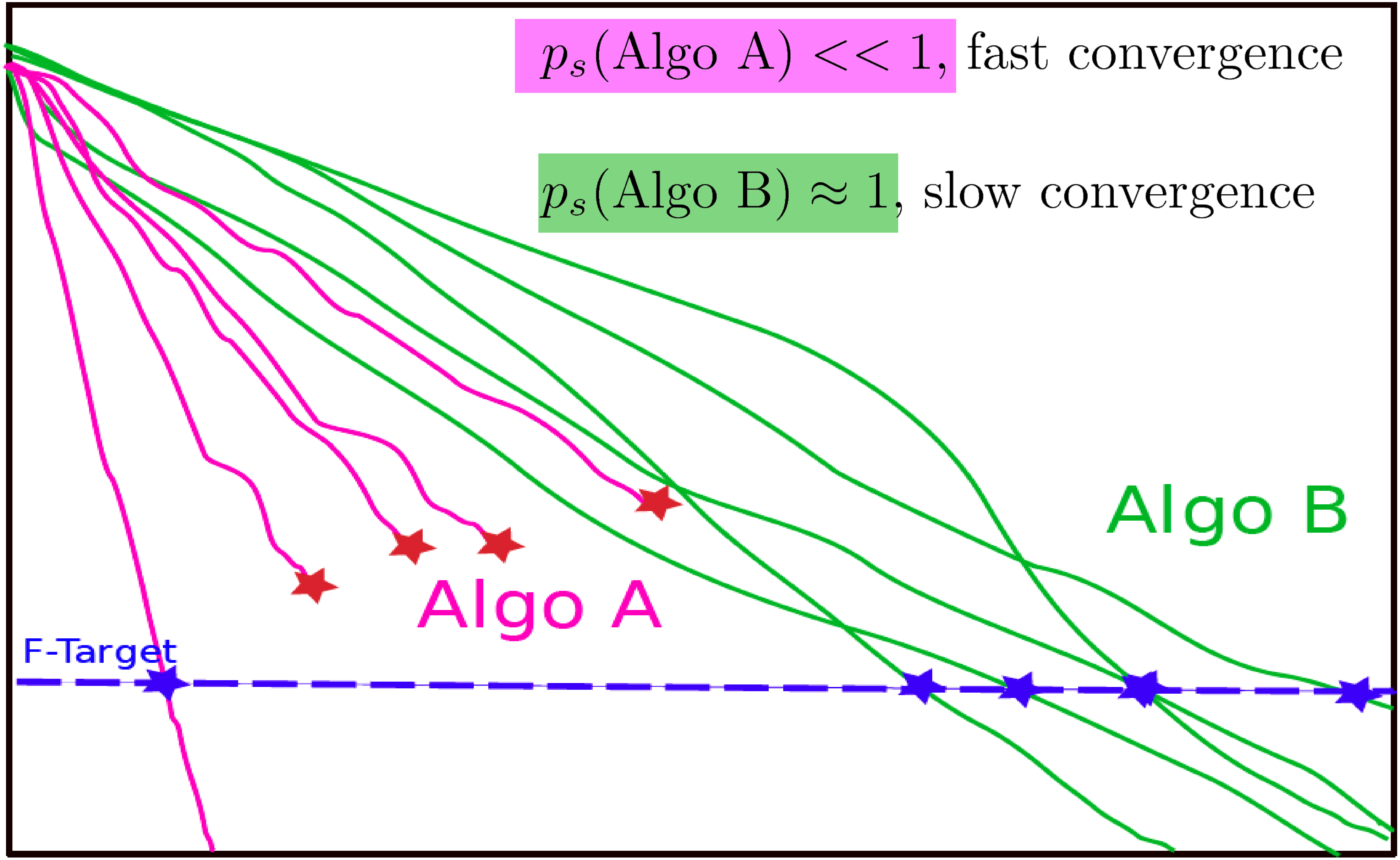


Collect Running Time (RT) to reach F-Target



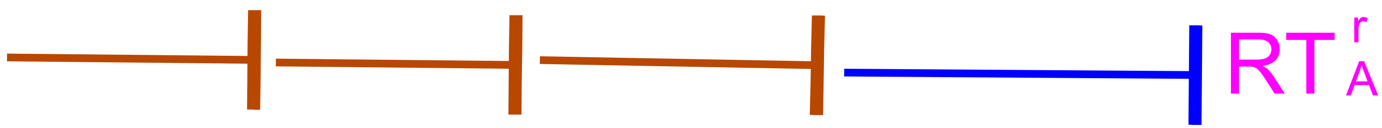
Which performance measure?

to compare the two following scenario?



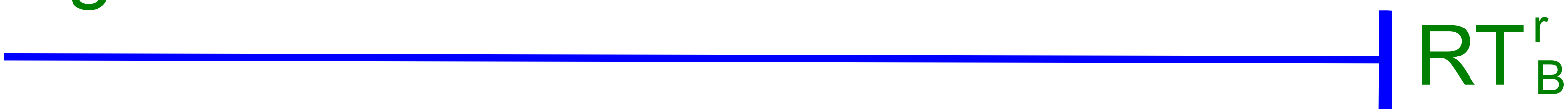
Compare Running Time of algorithms restarted independently till success:

Algo Restart A:



$$p_s(\text{Algo Restart A}) = 1$$

Algo Restart B:



$$p_s(\text{Algo Restart B}) = 1$$

Expected Running Time (ERT)

Running Time of Restarted Algorithm:

$$RT^r = \sum_{i=1}^{N-1} RT^i_{\text{unsuccesful}} + RT_{\text{successful}}$$

$$\text{N=Geometric RV with parameter } p_s \; [E(N) = \frac{1}{p_s}]$$

Expected Running Time (ERT) of Restarted Algo:

$$E[RT^r] = \frac{1-p_s}{p_s} E[RT_{\text{unsuccessful}}] + E[RT_{\text{successful}}]$$

Estimator for ERT

$$\text{ERT} = E[RT^r] = \frac{1-p_s}{p_s} E[RT_{\text{unsuccessful}}] + E[RT_{\text{successful}}]$$

Estimator for ERT:

$$\widehat{p_s} = \frac{\#\,\text{succ}}{\#\,\text{Runs}}$$

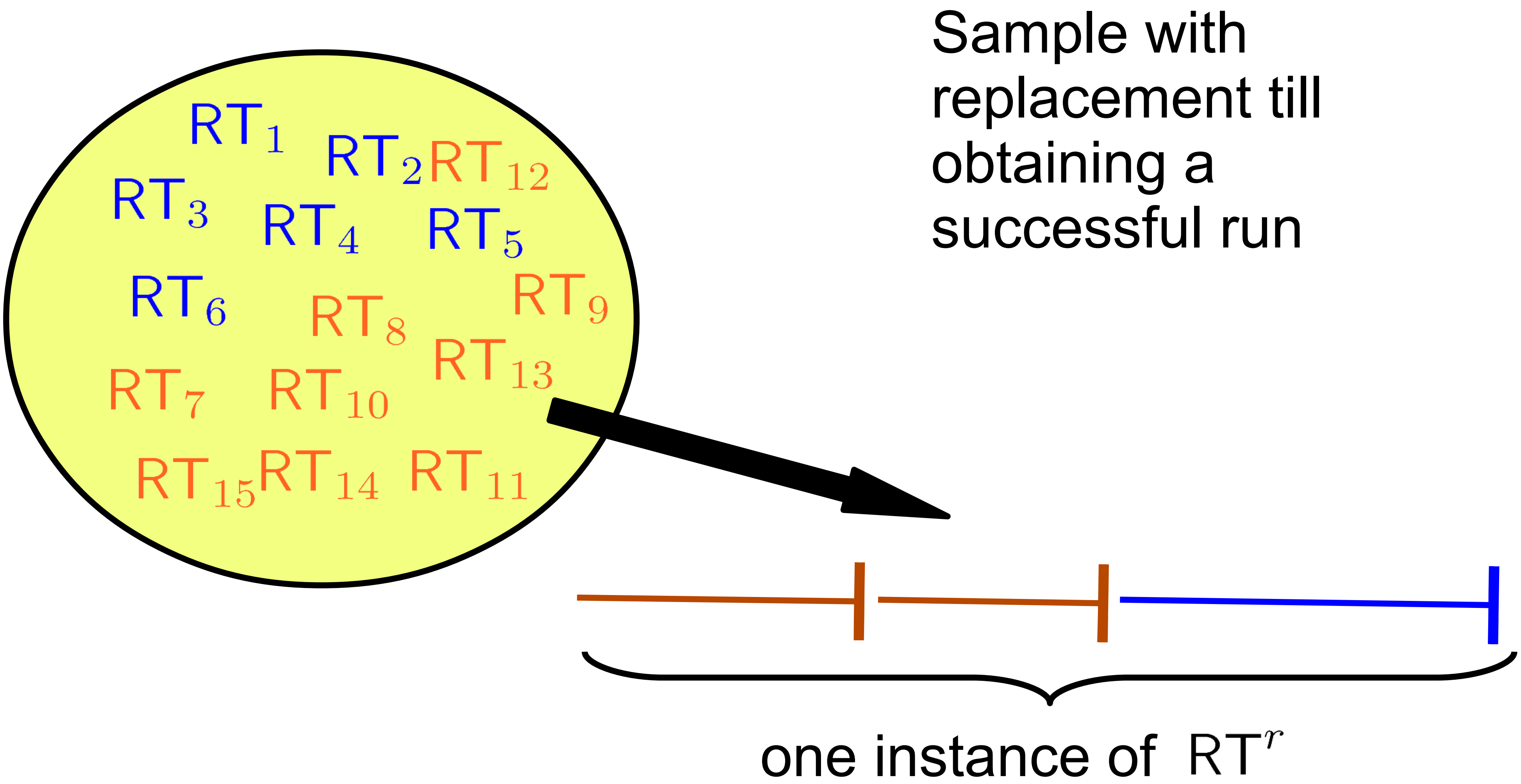
$$\widehat{RT_{\text{unsucc}}} = \text{Average Evals of unsuccessful runs}$$

$$\widehat{RT_{\text{succ}}} = \text{Average Evals of successful runs}$$

$$\widehat{\text{ERT}} = \frac{\#\text{Evals}}{\#\text{success}}$$

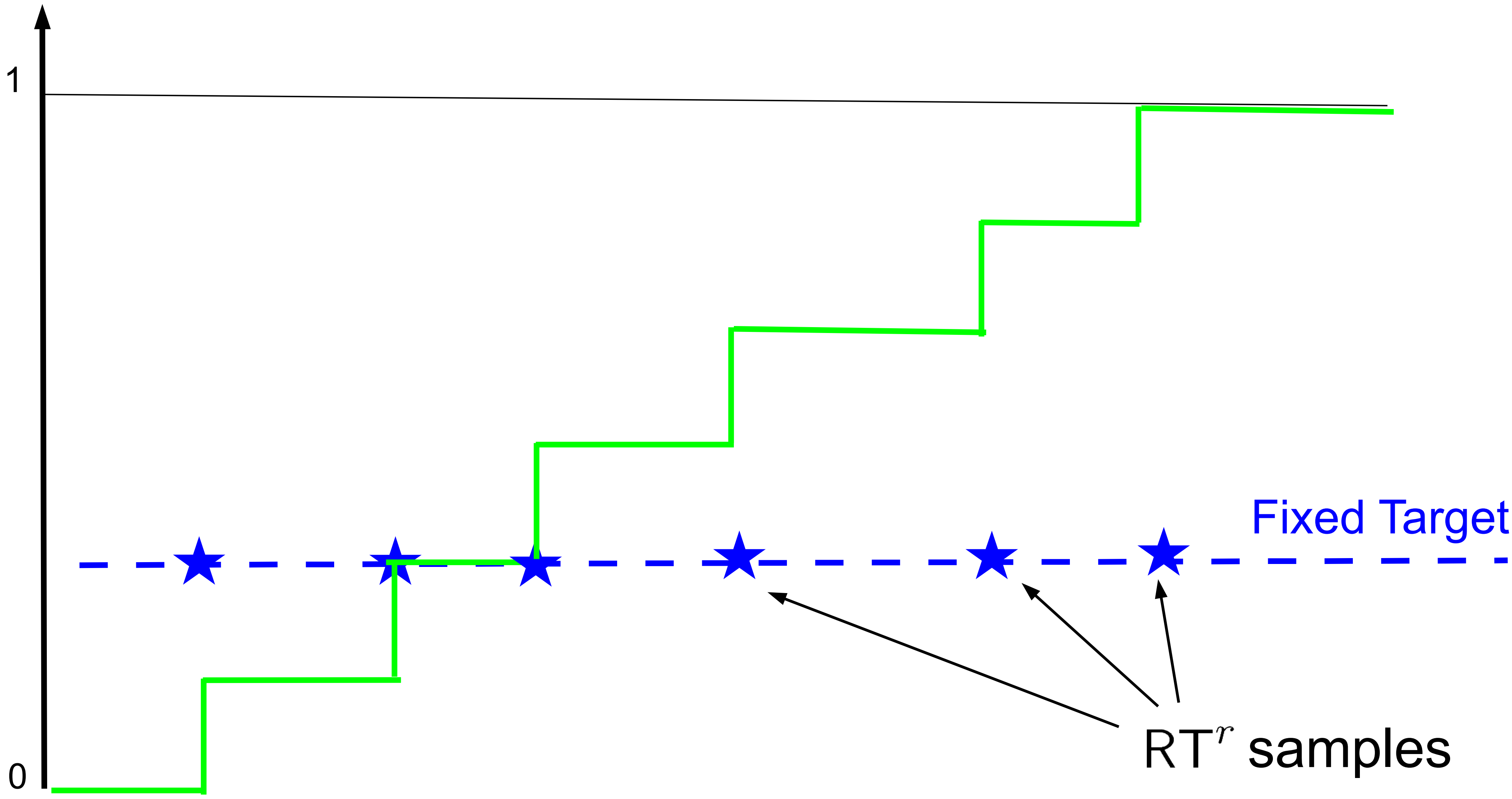
Obtaining single instances of RT^r

Generate a distribution of RT^r from collected data:



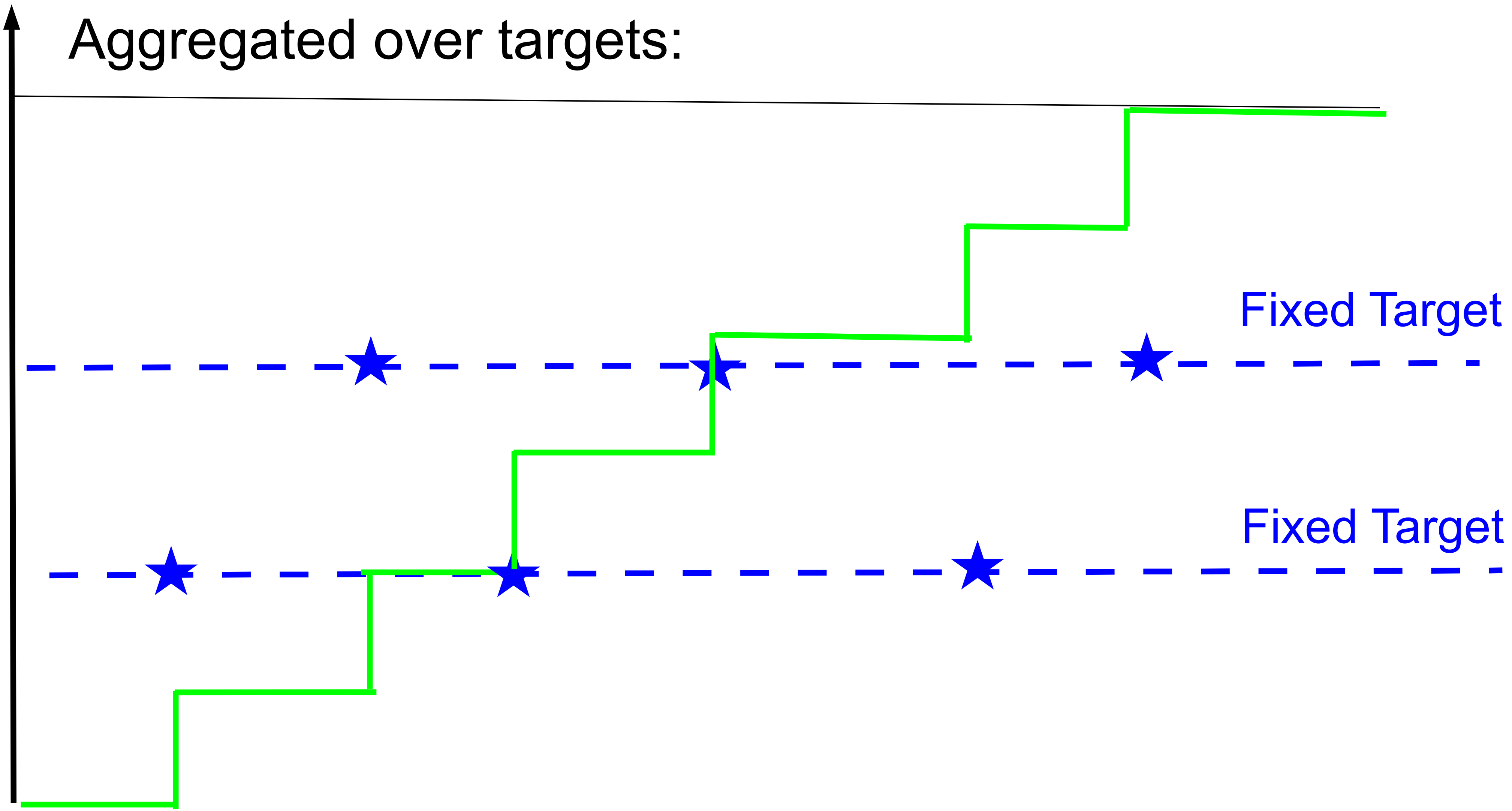
REPEAT 100 times

Empirical Cumulative distribution of RT^r

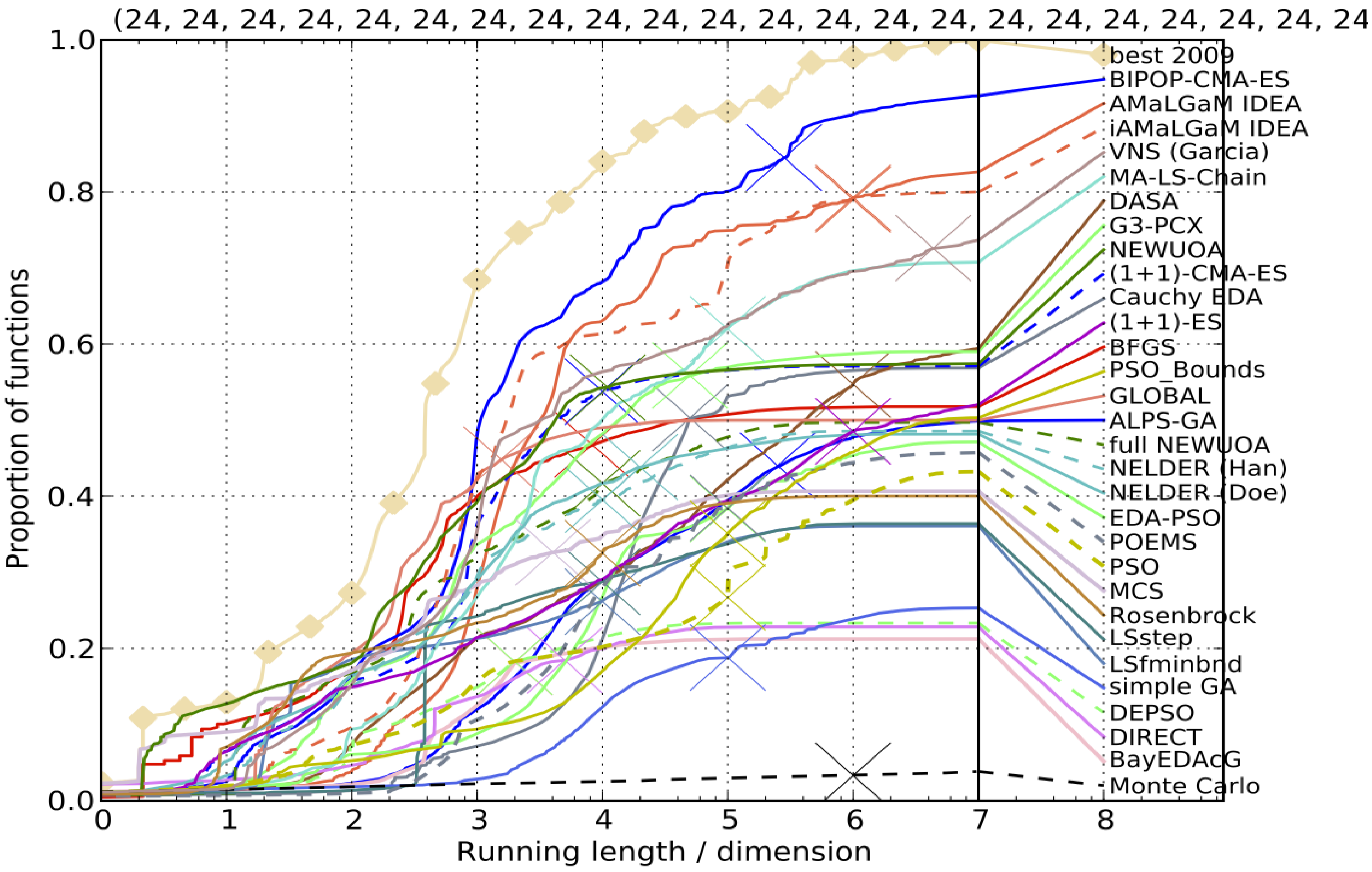


Empirical Cumulative distribution of RT^r

Aggregated over targets:



Empirical cumulative distributions of RT^r



% SEPARABLE
1 Sphere
2 Ellipsoid separable with monotone x-transformation, condition 1e6
3 Rastrigin separable with both x-transformations "condition" 10
4 Skew Rastrigin-Bueche separable, "condition" 10, skew-"condition" 100
5 Linear slope, neutral extension outside the domain (not flat)

% LOW OR MODERATE CONDITION
6 Attractive sector function
7 Step-ellipsoid, condition 100
8 Rosenbrock, original
9 Rosenbrock, rotated

% HIGH CONDITION
10 Ellipsoid with monotone x-transformation, condition 1e6
11 Discus with monotone x-transformation, condition 1e6
12 Bent cigar with asymmetric x-transformation, condition 1e6
13 Sharp ridge, slope 1:100, condition 10
14 Sum of different powers

% MULTI-MODAL
15 Rastrigin with both x-transformations, condition 10
16 Weierstrass with monotone x-transformation, condition 100
17 Schaffer F7 with asymmetric x-transformation, condition 10
18 Schaffer F7 with asymmetric x-transformation, condition 1000
19 F8F2 composition of 2-D Griewank-Rosenbrock

% MULTI-MODAL WITH WEAK GLOBAL STRUCTURE
20 Schwefel x*sin(x) with tridiagonal transformation, condition 10
21 Gallagher 101 Gaussian peaks, condition up to 1000
22 Gallagher 21 Gaussian peaks, condition up to 1000, 1000 for global opt
23 Katsuuras repetitive rugged function
24 Lunacek bi-Rastrigin, condition 100