

# Mesure de performances

Instance d'un problème

$$p^5 = p(n, f_{\theta}, \theta_i, I, I_f^{\text{target}, \theta_i})$$

N: dimension

$f_{\theta}$ : fonction

$\theta_i$ : instance de la fonction où le domaine est translatée

I: Quality Indicator Mapping: comment choisir notre meilleure solution

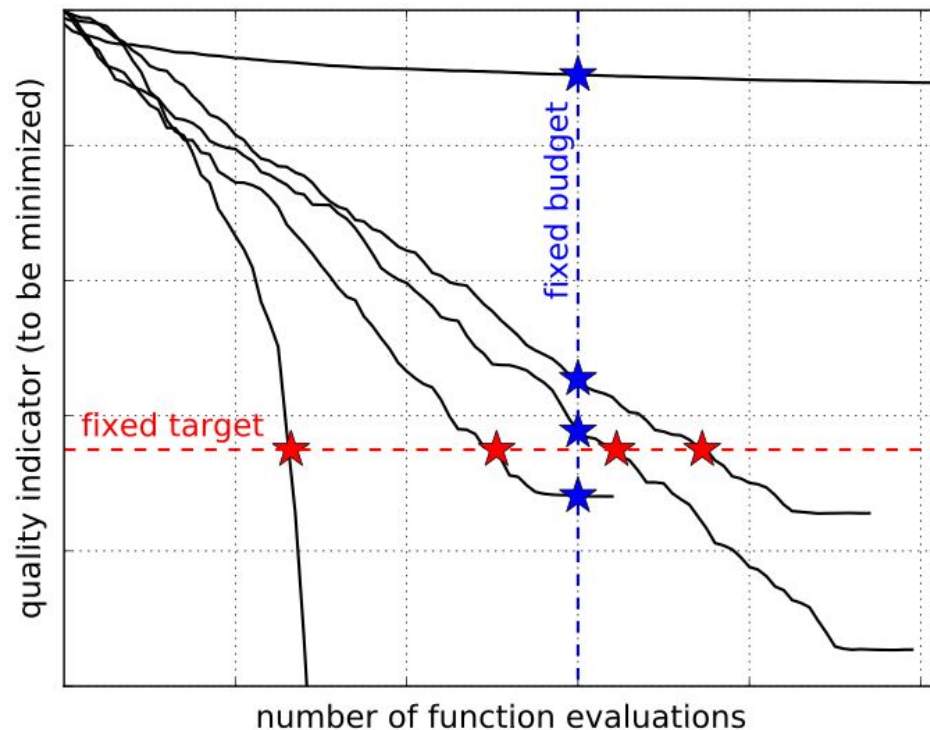
If target : valeur cible à atteindre pour l'instance du problème

# Mesure de performance

Algorithm runtime = Nombre d'évaluations pour arriver à un "target" prédéfini pendant l'exécution de l'algorithme.

Si nombre d'évaluations maximum atteint = runtime indéfini (unsuccessful)

# Deux méthodes de mesure de performance



# Runtime computation

Pour chaque triplet  $p^3 = p(n, f_\theta, \theta_i)$

Correspond plusieurs 5-uplets  $p^5 = p(n, f_\theta, \theta_i, I, I_f^{\text{target}, \theta_i})$

Dans un single run de l'algorithme sur  $p^3$  nous obtenons une mesure pour chaque target value ( $p^5$ ) qui a été atteinte durant cette exécution

# Runtime and problem instance

Les instances différentes du même problème sont à vrai dire des translations du domaine de recherche

→ K instances équivaut à K exécutions différentes sur le même problème

→ Simulation d'un restart en cas d'échec

$$\mathbf{RT}(n, f_{\theta}, \Delta I) = \sum_{j=1}^J \mathbf{RT}_j^{\text{us}}(n, f_{\theta}, \Delta I) + \mathbf{RT}^{\text{s}}(n, f_{\theta}, \Delta I) \ ,$$

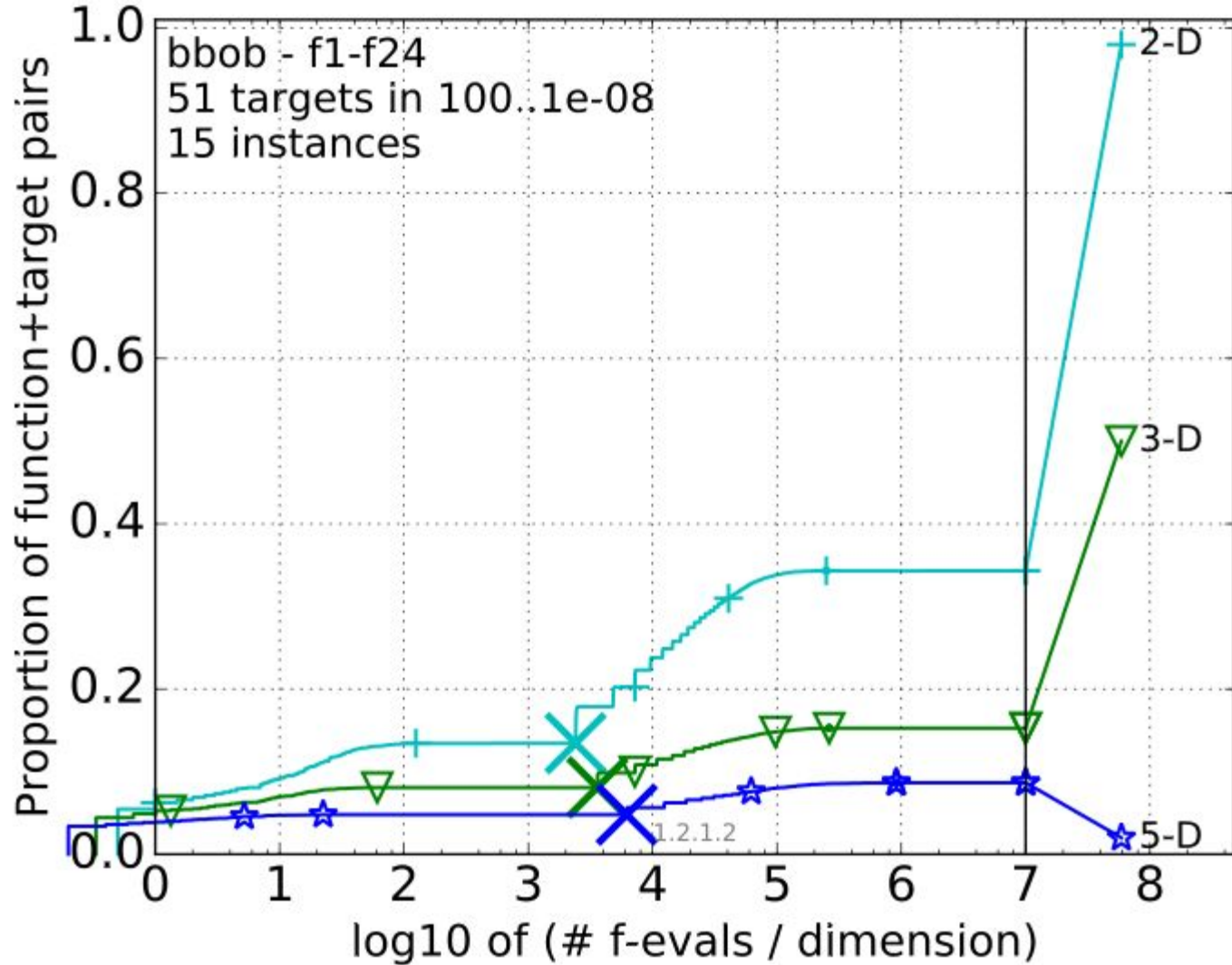
# ECDF

empirical cumulative distribution function: Proportion de problèmes résolus pour un budget donné

En abscisse: budget (pour chaque valeur de  $x$ ,  $y$  représente la fraction de problème résolu avec ce budget)

En ordonnée: Proportion des paires (fonction, budget) (pour chaque  $y$ ,  $x$  représente le runtime maximal observé sur ces problèmes)

L'aire sous la valeur  $y$  à gauche reflète la moyenne des runtime sur cette proportion: plus petit, mieux c'est.





Croix: médiane de la longueur maximale (en f-évaluations) d'un échec d'exécution (unsuccessful run)

Le point: la proportion des paires (fonction,target) pour lesquelles au moins une exécution avec succès