Adaptive Operator Selection with Rank-based Multi-Armed Bandits

Alvaro Fialho, Marc Schoenauer & Michèle Sebag

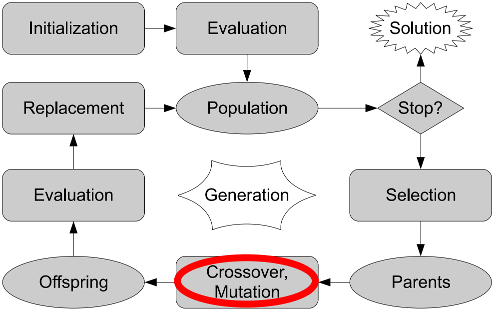
CENTRE DE RECHERCHE COMMUN INRIA MICROSOFT RESEARCH

26th COW, April 22., 2013

Outline

Context & Motivation

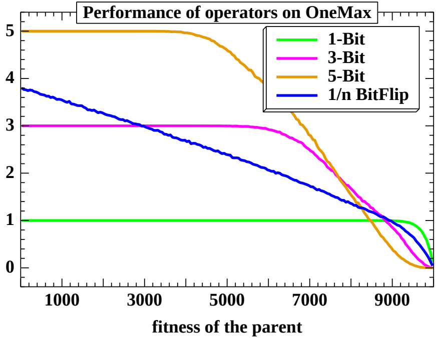
Evolutionary Algorithms



- Stochastic optimization algorithms (Darwinian paradigm)
  - Bottleneck: parameter setting
    - Population size and number of offspring
    - Selection and replacement methods (and their parameters)
    - Variation Operators (application rate, internal parameters)
- Goal: Automatic setting (Crossing the Chasm) [Moore, 1991]

Parameter Setting of Variation Operators

- Difficult to predict the performance
- Problem-dependent and inter-dependent choices
  - Off-line tuning → best static strategy (expensive)



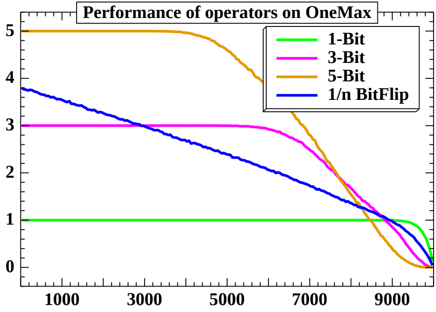
Also depends on ...

- Fitness of the parents
- Pop. fitness distribution

(sample fig. with a (1+50)-EA)

Parameter Setting of Variation Operators

- Difficult to predict the performance
- Problem-dependent and inter-dependent choices
  - Off-line tuning $\rightarrow$ best static strategy (expensive)



Also depends on . . .

- Fitness of the parents
- Pop. fitness distribution

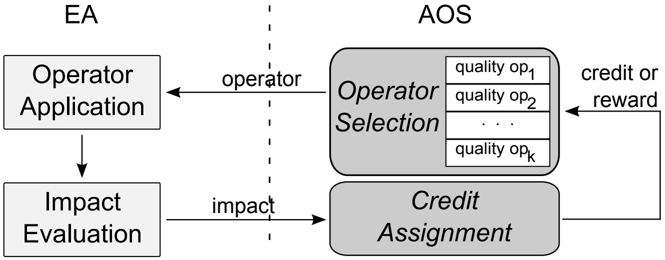
(sample fig. with a (1+50)-EA)

$\Rightarrow$ Should be adapted <u>on-line</u>, while solving the problem

Adaptive Operator Selection

Position of the Problem

- Given a set of K variation operators
- Select on-line the operator to be applied next
- Based on their recent effects



Operator Selection

A (kind of) Multi-Armed Bandit problem

The Basic Multi-Armed Bandit Problem

- Given K arms ($\equiv$ operators)
- At time t, *gambler* plays arm j and gets
  - $r_{j,t} = 1$ with (unknown) prob. p_j
  - $r_{j,t} = 0$ with prob. $1 - p_j$

Goal: maximize cumulative reward $\equiv$ minimize regret

$$\mathcal{L}(T) = \sum_{t=1}^{T} (r_t^* - r_t)$$

The Upper Confidence Bound MAB algorithm

- Assymptotic optimality guarantees (static context) [Auer et al., 2002]

$$\text{Optimal } \mathcal{L}(T) = \mathcal{O}(\log T)$$

- At time t, choose arm i maximizing:

$$score_{i,t} = \underbrace{\hat{q}_{i,t}}_{exploitation} + \underbrace{\sqrt{\frac{2\log \sum_k n_{k,t}}{n_{i,t}}}}_{exploration}$$

with $n_{i,t+1} = n_{i,t} + 1$ # times

and $\hat{q}_{i,t+1} = \left(1 - \frac{1}{n_{i,t+1}}\right) \cdot \hat{q}_{i,t} + \frac{1}{n_{i,t+1}} \cdot r_{i,t}$ emp. qual.

- Efficiency comes from optimal EvE balance
  - Interval between exploration trials increases exponentially w.r.t. # time steps

Operator Selection with UCB: shortcomings

Exploration vs. Exploitation (EvE) balance

- In UCB theory, rewards $\in \{0, 1\}$; fitness-based rewards $\in [a, b]$
- UCB's EvE balance is broken, Scaling is needed:

$$score_{i,t} = \hat{q}_{i,t} + \mathcal{C}\sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{i,t}}}$$

Dynamical setting (best arm/op changes along evolution)

- Adjusting $\hat{q}$'s after a change takes a long time
- Use change detection test (e.g. Page-Hinkley) [Hinkley, 1969]

$\Rightarrow$ Upon the detection of a change, restart the MAB.

DMAB = UCB + Scaling + Page-Hinkley

Operator Selection: Discussion

MAB = UCB + Scaling

- Optimal EvE, but in static setting. . . AOS is dynamic

DMAB = MAB + Page-Hinkley change-detection

- Won Pascal challenge on On-line EvE trade-off [Hartland et al., 2007]
  - Utilization in the AOS context [GECCO'08]
- 2 hyper-parameters: scaling C and Page-Hinkley threshold γ
- Very efficient, but very sensitive to hyper-parameter setting
- Change-detection works only when changes are abrupt

An alternative: 'More Dynamic' Reward

Credit Assignment

1. Context & Motivation
2. Operator Selection
3. Credit Assignment
   - Fitness-based Rewards
   - Area-Under-the-Curve (AUC)
   - Rank-based AUC with MAB
4. Empirical Validation
5. Conclusions & Further Work

Fitness-based Rewards

Impact of an operator application?

- Most common: Fitness Improvement $\Delta\mathcal{F}$
- For multi-modal problems: diversity also important [CEC'09]

From Impact to Credit (or reward)

- Instantaneous ($\Delta\mathcal{F}$ last application) likely to be unstable
- Average of the last $\mathcal{W}$ applications
- Extreme value over the last $\mathcal{W}$ applications [PPSN'08]
  - Rare extreme events are more important than average e.g. rogue waves, epidemic propagation

Issues: High sensitivity to scaling parameters

- ...likely to be dynamic, too

Higher robustness: Credit Assignment based on Ranks

Area-Under-the-Curve (AUC)

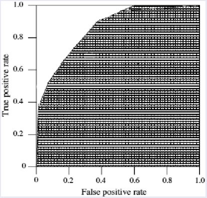
Area Under ROC Curve in ML

- Evaluation of binary classifiers [Fawcett, 2006]

$$[+ + - - + + + - - - - \ldots]$$

  - Performance: % of misclassification
  - Equivalent to MannWhitneyWilcoxon test

$$Pr\left(rank(n^{+}) > rank(n^{-})\right)$$

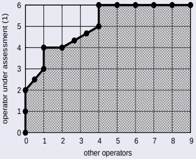


Area Under ROC Curve in AOS

- One operator versus others [GECCO'10]

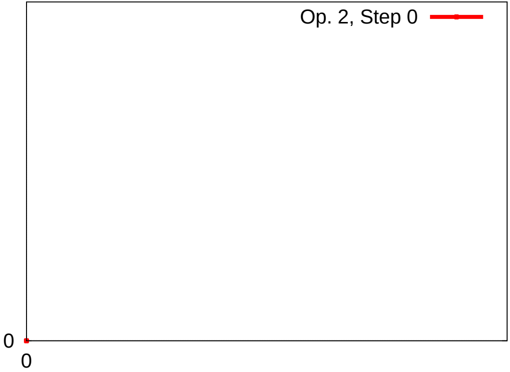
$$[\ \mathbf{op}_1, op_2, \mathbf{op}_1, \mathbf{op}_1, \mathbf{op}_1, op_2, op_2, \ldots]$$

- Fitness improvements are ranked
- Size of the segment = assigned rank-value



Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



Rank-Based AUC

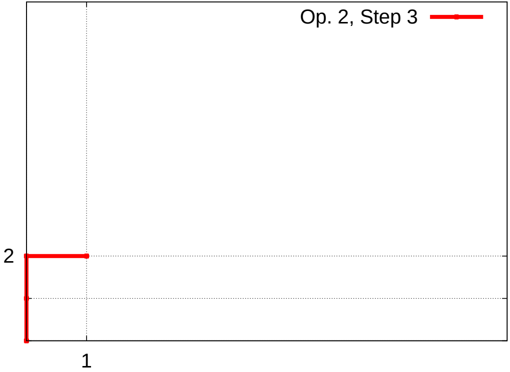
R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0

Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0

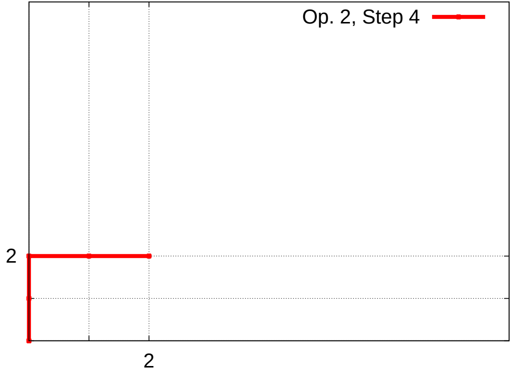
Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



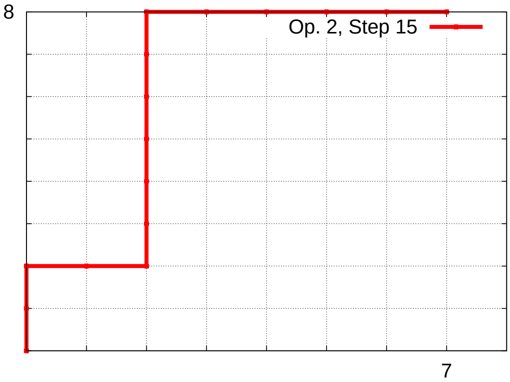
Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



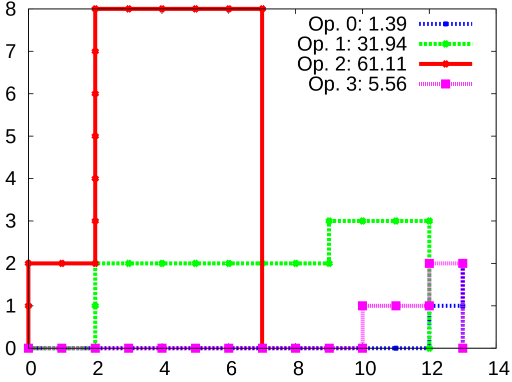
Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



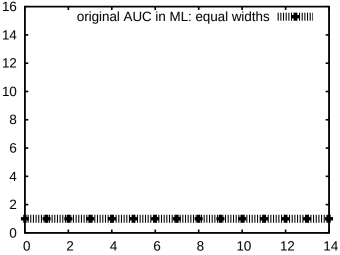
Rank-Based AUC

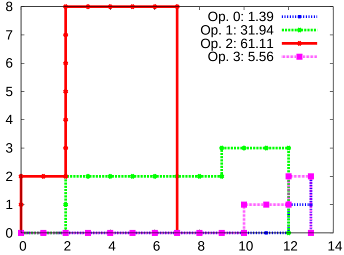
R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



Rank-Based AUC

R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0

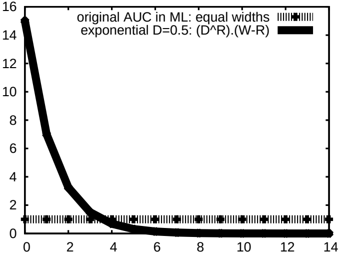




segments with same size

Rank-Based AUC

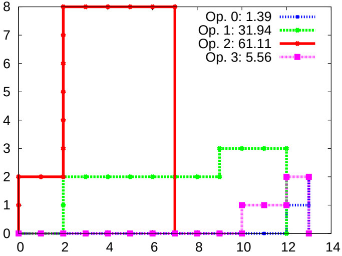
R	$\Delta\mathcal{F}$	Op
1	5.0	2
2	4.7	2
3	4.2	1
4	3.5	1
5	3.4	2
6	3.3	2
7	3.1	2
8	3.0	2
9	2.9	2
10	2.8	2
11	2.5	3
12	2.0	1
13	1.5	0
14	1.0	3
15	0.8	0



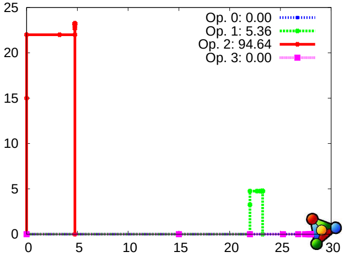
exponential decay

$$D^R(\mathcal{W} - R)$$

(example with $D = 0.5$) $\Rightarrow$



segments with same size



Rank-based AUC with MAB

Rationale

- AUC: behavior of all ops.: **dynamic by construction**
- AUC is already an aggregation: $\Rightarrow$ directly use AUC in UCB:

$$score_{j,t} = AUC_{j,t} + \mathcal{C} \cdot \sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{j,t}}}$$

Area-Under-Curve (AUC)

- Ranks over fitness improvements ($\Delta\mathcal{F}$)
- Invariant w.r.t. linear scaling of $\mathcal{F}$

Fitness-based AUC (FAUC)

- Ranks over fitness values ($\mathcal{F}$), rather than ranks over $\Delta\mathcal{F}$
- Invariant w.r.t monotonous transformations of $\mathcal{F}$

$\rightarrow$ Comparison-based AOS

Empirical Validation

1. Context & Motivation
2. Operator Selection
3. Credit Assignment
4. Empirical Validation
   - Goals of Experiments
   - (1+50)-EA on the OneMax Problem
   - DE on BBOB continuous Benchmarks
5. Conclusions & Further Work

Goals of Experiments

Given a set of K operators . . .

Performance ?

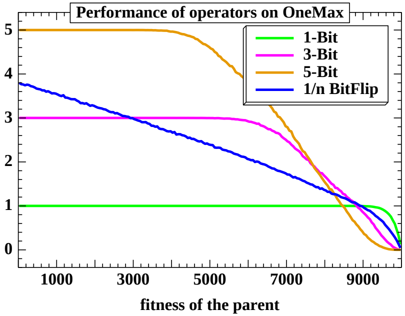
- Baseline methods
  1. Each operator being applied alone
  2. Naive uniform selection between operators
  3. Static off-line tuning of application rates (cost $\gg$)
  4. Optimal behavior (available only on simple benchmarks)
  5. State-of-the-art OS method: Adaptive Pursuit [Thierens, 2005]

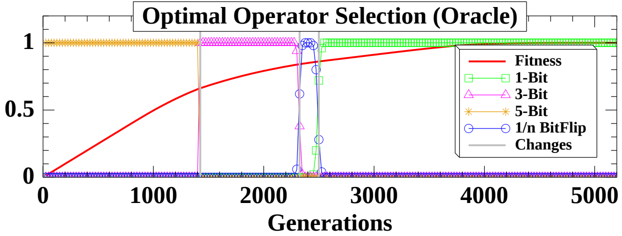
Robustness/Generality ?

- AOS methods have *hyper*-parameters
  Tuned off-line by F-RACE [Birattari et al., 2002]
  - Robustness w.r.t. hyper-parameter setting
  - Generality w.r.t. different problems/landscapes
    - Invariance properties

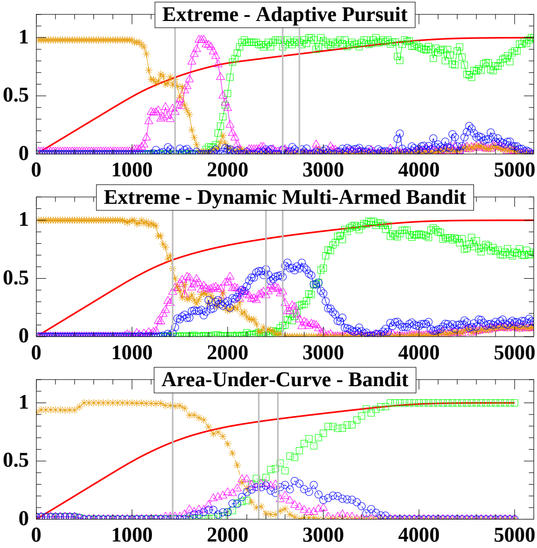
The OneMax Problem

- 10^4 bits
- Fitness: # of "1"s
- (1+50)-GA
- 4 mutation operators





Comparative Results



Monotonous Transformations of the Fitness

- Original OneMax: $\mathcal{F} = \sum_{i=1}^{n} b_i$
- 3 monotonous transformations: $\log(\mathcal{F})$, $\exp(\mathcal{F})$ and $\mathcal{F}^2$

(h-l)	$\mathcal{F} = \sum b_i$	$\log(\mathcal{F})$	$\exp(\mathcal{F})$	$\mathcal{F}^2$	AOS tech.
485	**5103/427**	**5195/430**	**5562/950**	**5588/950**	AUC-MAB
807	**5123/218**	5431/223	5930/334	5792/382	Ext-AP
0	5726/399	5726/399	5726/399	5726/399	FAUC-MAB
2591	5376/285	7967/718	7722/2151	6138/516	Ext-DMAB
6971	6059/667	8863/694	13030/3053	12136/949	Ext-SLMAB
7052	9044/840	7947/1267	14999/0	14999/0	Ext-MAB

Monotonous Transformations of the Fitness

- Original OneMax: $\mathcal{F} = \sum_{i=1}^{n} b_i$
- 3 monotonous transformations: $\log(\mathcal{F})$, $\exp(\mathcal{F})$ and $\mathcal{F}^2$

(h-l)	$\mathcal{F} = \sum b_i$	$\log(\mathcal{F})$	$\exp(\mathcal{F})$	$\mathcal{F}^2$	AOS tech.
485	**5103/427**	**5195/430**	**5562/950**	**5588/950**	AUC-MAB
807	**5123/218**	5431/223	5930/334	5792/382	Ext-AP
0	5726/399	5726/399	5726/399	5726/399	FAUC-MAB
2591	5376/285	7967/718	7722/2151	6138/516	Ext-DMAB
6971	6059/667	8863/694	13030/3053	12136/949	Ext-SLMAB
7052	9044/840	7947/1267	14999/0	14999/0	Ext-MAB

Other (artificial) scenarios

- Binary: Long K-Path, Royal Road, ...
- Combinatorial: SAT
- Continuous: BBOB

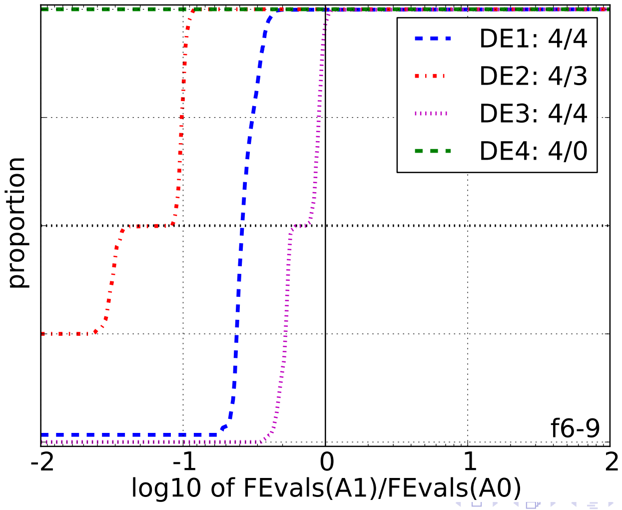
DE on BBOB continuous Benchmarks

- Exp. framework for rigorous benchmarking [Hansen et al., 2010]
- 24 continuous functions, 15 instances per function
- Several problem dimensions (2, 3, 5, 10, 20, 40)

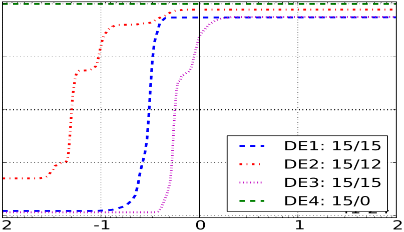
Adaptive Operator Selection in Differential Evolution

- A completely different evolutionary algorithm [Storn and Price, 1995]
- $NP = 100 \cdot DIM$; $CR = 1.0$; $F = 0.5$
- With 4 possible mutation strategies
  - rand/1, rand/2, rand-to-best/2, current-to-rand/1

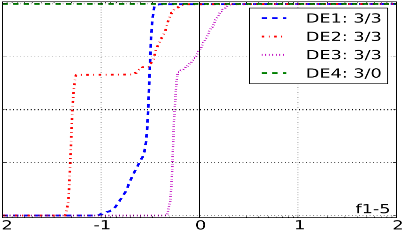
Parwise comparisons of FAUC-Bandit with . . . (sample fig)



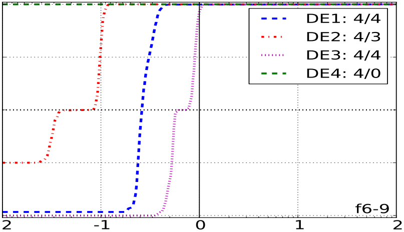
Parwise comparisons of FAUC-Bandit with ...



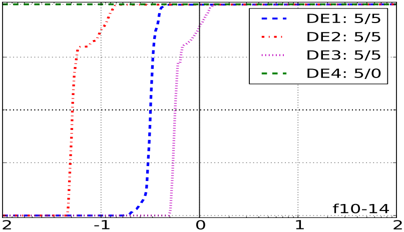
(a) all functions



(b) separable functions

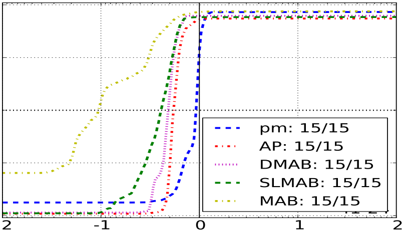


(c) moderate functions

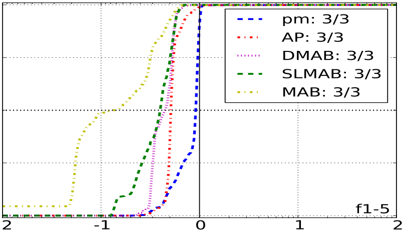


(d) ill-conditioned functions

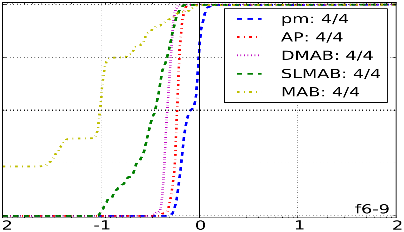
Parwise comparisons of FAUC-Bandit with ...



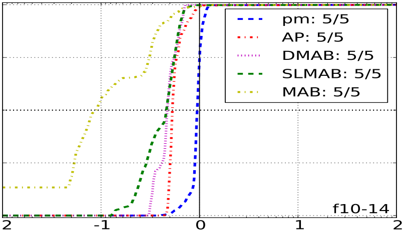
(e) all functions



(f) separable functions

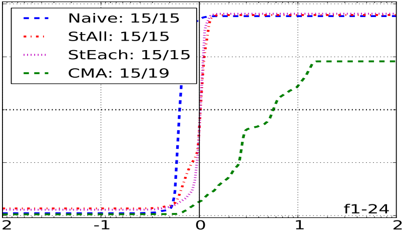


(g) moderate functions

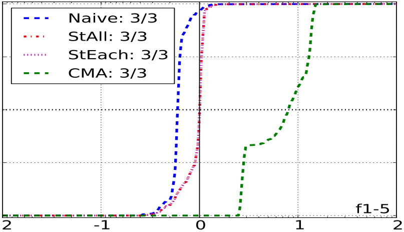


(h) ill-conditioned functions

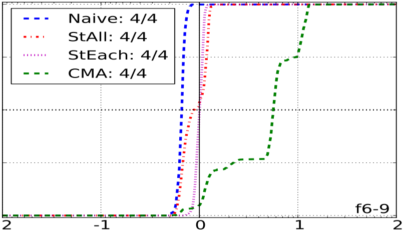
Parwise comparisons of FAUC-Bandit with . . .



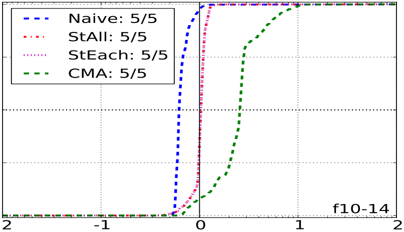
(i) all functions



(j) separable functions



(k) moderate functions



(l) ill-conditioned functions

Conclusions & Further Work

Conclusions

Algorithmic Contributions

- Operator Selection
  - MAB = UCB + Scaling
  - DMAB = MAB + Page-Hinkley test [GECCO'08]
- Credit Assignment
  - Extreme value-based ($\Delta\mathcal{F}$) [PPSN'08]
  - Rank-based methods [GECCO'10]
- AOS Combinations
  - Extreme-xMAB: efficient, but sensitive w.r.t. hyper-parameters
  - (F)AUC-MAB: efficient and robust w.r.t. hyper-parameters
    - FAUC: comparison-based

⇒ **Combining** concepts from ML: MABs and AUC

⇒ **Extending** them to a dynamic context

Conclusions (2)

Empirical Validation (performance, robustness and generality)

- Genetic Algorithms
  - Artificial scenarios [GECCO'08, AMAI'10, GECCO'10]
  - Boolean problems [PPSN'08, LION'09, GECCO'09, AMAI'10, GECCO'10]
    - OneMax, Long K-Path and Royal Road problems
- Memetic Algorithms
  - SAT problems, with the Compass Credit Assign. [CEC'09, Chapter'10]
    A highly multimodal context
- Differential Evolution
  - Continuous problems [BBOB'10, PPSN'10]

Some Perspectives for Further Work (from 12/2010!)

- Application extensions: AOS paradigm is very general
  - Use within other meta-heuristics
  - Use at the level of hyper-heuristics
    - Cross-domain Heuristic Search Challenge (CHeSC)

- Algorithmic extensions: towards real-world problems
  - Extend to multi-modal (diversity, pop.size, . . .)
  - Extend to multi-objective (Pareto, hyper-volume, . . .)

- First trial in real-world: sustainable development
  - Optimization of designs of buildings for energy efficiency

Our Publications I

Da Costa, L., Fialho, A., Schoenauer, M., and Sebag, M. (2008).
Adaptive operator selection with dynamic multi-armed bandits.
In *Proc. Genetic and Evolutionary Computation Conference (GECCO)*. ACM.

Fialho, A., Da Costa, L., Schoenauer, M., and Sebag, M. (2008).
Extreme value based adaptive operator selection.
In *Proc. Intl. Conf. on Parallel Problem Solving from Nature (PPSN)*. Springer.

Fialho, A., Da Costa, L., Schoenauer, M., and Sebag, M. (2009).
Dynamic multi-armed bandits and extreme value-based rewards for AOS in evolutionary algorithms.
In *Proc. Intl. Conf. on Learning and Intelligent Optimization (LION)*. Springer.

Maturana, J., Fialho, A., Saubion, F., Schoenauer, M., and Sebag, M. (2009).
Extreme compass and dynamic multi-armed bandits for adaptive operator selection.
In *Proc. IEEE Congress on Evolutionary Computation (CEC)*. IEEE.

Fialho, A., Schoenauer, M., and Sebag, M. (2009).
Analysis of adaptive operator selection techniques on the royal road and long k-path problems.
In *Proc. Genetic and Evolutionary Computation Conference (GECCO)*. ACM.

Maturana, J., Fialho, A., Saubion, F., Schoenauer, M., Lardeux, F., and Sebag, M. (2010).
Adaptive operator selection and management in evolutionary algorithms.
In Y. Hamadi et al, editor, *Autonomous Search*. Springer. (to appear)

Fialho, A., Da Costa, L., Schoenauer, M., and Sebag, M. (2010).
Analyzing bandit-based adaptive operator selection mechanisms.
Annals of Mathematics and A. I. – Special Issue on Learning and Intelligent Optimization. Springer.

Our Publications II

Fialho, A., Schoenauer, M., and Sebag, M. (2010).
Toward comparison-based adaptive operator selection.
In *Proc. Genetic and Evolutionary Computation Conference (GECCO)*. ACM.

Gong, W., Fialho, A., and Cai, Z. (2010).
Adaptive strategy selection in differential evolution.
In *Proc. Genetic and Evolutionary Computation Conference (GECCO)*. ACM.

Fialho, A., Schoenauer, M., and Sebag, M. (2010).
Fitness-AUC bandit adaptive strategy selection vs. the probability matching one within DE.
In *Black-Box Optimization Benchmarking Workshop (BBOB-GECCO)*. ACM.

Fialho, A., Gong, W., and Cai, Z. (2010).
Probability matching-based adaptive strategy selection vs. uniform strategy selection within DE.
In *Black-Box Optimization Benchmarking Workshop (BBOB-GECCO)*. ACM.

Fialho, A. and Ros, R. (2010).
Analysis of adaptive strategy selection within differential evolution on the BBOB-2010 noiseless benchmark.
Research Report RR-7259, INRIA.

Fialho, A., Ros, R., Schoenauer, M., and Sebag, M. (2010).
Comparison-based adaptive strategy selection in differential evolution.
In *Proc. Intl. Conf. on Parallel Problem Solving from Nature (PPSN)*. Springer.

Li, K., Fialho, A., and Kwong, S. (2011).
Multi-objective differential evolution with adaptive control of parameters and operators.
In *Proc. Intl. Conf. on Learning and Intelligent Optimization (LION)*. Springer. (to appear)

Adaptive Operator Selection with Rank-based Multi-Armed Bandits

Alvaro Fialho, Marc Schoenauer & Michèle Sebag

CENTRE DE RECHERCHE COMMUN
INRIA
MICROSOFT RESEARCH

26th COW, April 22., 2013

Other References I

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002).
Finite-time analysis of the multi-armed bandit problem.
Machine Learning, 47(2-3):235–256.

Birattari, M., Stützle, T., Paquete, L., and Varrentrapp, K. (2002).
A racing algorithm for configuring metaheuristics.
In W.B. Langdon et al., editor, *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, pages 11–18. Morgan Kaufmann.

Fawcett, T. (2006).
An introduction to ROC analysis.
Pattern Recogn. Lett., 27(8):861–874.

Hansen, N., Auger, A., Finck, S., and Ros, R. (2010).
Real-parameter black-box optimization benchmarking 2010: Experimental setup.
Technical Report RR-7215, INRIA.

Hartland, C., Baskiotis, N., Gelly, S., Teytaud, O., and Sebag, M. (2007).
Change point detection and meta-bandits for online learning in dynamic environments.
In *Proc. Conférence Francophone sur l'Apprentissage Automatique (CAPS)*.

Hinkley, D. (1969).
Inference about the change point in a sequence of random variables.
Biometrika, 57(1):1–17.

Other References II

Moore, G. (1991).
Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customer.
Collins Business Essentials.

Storn, R. and Price, K. (1995).
Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces.
Technical Report TR-95-012, Intl. Computer Science Institute.

Thierens, D. (2005).
An adaptive pursuit strategy for allocating operator probabilities.
In H.-G. Beyer et al., editor, *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, pages 1539–1546. ACM.