Adaptive Operator Selection for Optimization
Álvaro Fialho

To cite this version:

HAL Id: tel-00578431
https://tel.archives-ouvertes.fr/tel-00578431
Submitted on 20 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Adaptive Operator Selection for Optimization

Ph.D. Thesis
École Doctorale d’Informatique, ED 427
Université Paris-Sud 11

By Álvaro Roberto SILVESTRE FIALHO

Presented and publicly defended:
On December 22 2010
In Orsay, France

With the following jury:

Ágoston E. Eiben, Vrije Universiteit Amsterdam, The Netherlands (Reviewer)
Youssef Hamadi, Microsoft Research Cambridge, United Kingdom (Examiner)
Yannis Manoussakis, Université Paris-Sud XI, France (Examiner)
Marc Schoenaer, INRIA Saclay - Île-de-France, France (Thesis Advisor)
Michèle Sebag, LRI / CNRS, France (Thesis Advisor)
Thomas Stützle, IRIDIA / FNRS, Belgium (Reviewer)
Dirk Thierens, Universiteit Utrecht, The Netherlands (Examiner)

Microsoft Research – INRIA Joint Centre
28 rue Jean Rostand 91893 Orsay Cedex, France
SÉLECTION ADAPTATIVE D’OPÉRATEURS POUR L’OPTIMISATION

THÈSE DE DOCTORAT
ÉCOLE DOCTORALE D’INFORMATIQUE, ED 427
UNIVERSITÉ PARIS-SUD 11

Par Álvaro Roberto SILVESTRE FIALHO

Présentée et soutenue publiquement
Le 22 Décembre 2010
À Orsay, France
Devant le jury ci-dessous :

Ágoston E. Eiben, Vrije Universiteit Amsterdam, Pays-Bas (Rapporteur)
Youssef Hamadi, Microsoft Research Cambridge, Royaume-Uni (Examinateur)
Yannis Manoussakis, Université Paris-Sud XI, France (Examinateur)
Marc Schoenauer, INRIA Saclay - Île-de-France, France (Directeur de Thèse)
Michèle Sebag, LRI / CNRS, France (Directeur de Thèse)
Thomas Stützle, IRIDIA / FNRS, Belgique (Rapporteur)
Dirk Thierens, Universiteit Utrecht, Pays-Bas (Examinateur)

Centre de Recherche Commun INRIA – Microsoft Research
28 rue Jean Rostand 91893 Orsay Cedex, France
Abstract

Evolutionary Algorithms (EAs) are stochastic optimization algorithms which have already shown their efficiency on many application domains. This is achieved mainly due to the many parameters that can be defined by the user according to the problem at hand. However, the performance of EAs is very sensitive to the setting of these parameters, and there are no general guidelines for an efficient setting; as a consequence, EAs are rarely used by researchers from domains other than computer science. The methods proposed in this thesis contribute towards alleviating the user from the need of defining two very sensitive and problem-dependent choices: which variation operators should be used for the generation of new solutions, and at which rate each operator should be applied. The paradigm, referred to as Adaptive Operator Selection (AOS), provides the on-line autonomous control of the operator that should be applied at each instant of the search, i.e., while solving the problem. In order to do so, one needs to define a Credit Assignment scheme, which rewards the operators based on the impact of their recent applications on the current search process, and an Operator Selection mechanism, that decides which should be the next operator to be applied, based on the empirical quality estimates built by the rewards received. In this work, we have tackled the Operator Selection problem as an instance of the Exploration versus Exploitation dilemma: the best operator needs to be exploited as much as possible, while the others should also be minimally explored from time to time, as one of them might become the best in a further moment of the search. We have proposed different Operator Selection techniques to extend the Multi-Armed Bandit paradigm to the dynamic context of AOS. On the Credit Assignment side, we have proposed rewarding schemes based on extreme values and on ranks, in order to promote the use of outlier operators, while providing more robust operator assessments. The different AOS methods formed by the combinations of the proposed Operator Selection and Credit Assignment mechanisms have been validated on a very diverse set of benchmark problems. Based on empirical evidence gathered from this empirical analysis, the final recommended method, which uses the Rank-based Multi-Armed Bandit Operator Selection and the Area-Under-Curve Credit Assignment schemes, has been shown to achieve state-of-the-art performance while also being very robust with respect to different problems.
Acknowledgements

Firstly, I would like to deeply thank God, and my parents. Their continuous support in every aspect of my life were and will always be essential for my personal achievements.

To my beloved wife, Camila, I can not express in words all the gratitude I feel. She left her family and professional career in Brazil to follow and to support me into my personal dream of getting a Ph.D. from a top research institution in Europe. Thank you, with all my heart, for everything.

Needless to say, I’m also extremely grateful to my thesis advisors, Marc and Michèle. A big big thank you for having advised me in such a great (and always friendly) way. I hope our scientific and friendship relation will not end by the time I get back to the other side of the Atlantic ocean.

Thanks also to Youssef, and to Marc again, for having provided me this opportunity of making a Ph.D. with the financial support of the Microsoft Research – INRIA Joint-Centre. And a big thanks to Martine, for having greatly assisted me in every bureaucratic aspect, always in a very friendly way.

I also need to thank here my previous advisors, Marco António and Hemerson Pistori (UCDB), and Marco Túlio (USP). You were examples that have actively influenced my decision to become a researcher, thank you.

During this period as a Ph.D. student, I also had the opportunity to work in collaboration with different friends: Luis and Raymond (TAO), Jorge and Frédéric (Angers), Wenyin (Wuhan), Dave and Holger (Vancouver), and Ke (Hong Kong). My thesis work benefited a lot from these collaborations, thank you for having shared your time with me.

It is also important to thank all my friends from the TAO and ADAPT teams: Alejandro, Miguel, Luis, Raymond, Ibrahim, Cedric, Jacques, Romaric, Hassen, Mohammed, Fei, just to name a few, and also my friends from outside the lab. You made my stay here much nicer and funnier, thank you all, hope we will stay in touch.

Finally, I deeply thank the reviewers of my thesis, Gusz and Thomas, and the other members of the jury, Dirk, Youssef and Yannis. The detailed reviews and the insightful questions contributed a lot into further improving this manuscript, thank you.
To Camila, my beloved wife.
Résumé en Français

Les Algorithmes Évolutionnaires (AEs) sont des algorithmes stochastiques pour l’optimisation inspirés par le paradigme Darwinien de la “survie du plus adapté”. L’objectif est d’optimiser une fonction objective, aussi appelée fonction de fitness, définie sur l’espace de recherche $X$. Les éléments de $X$ sont appelés individus, et un ensemble d’individus est appelé population. L’AE fait évoluer une population d’individus en une succession d’itération (ou génération) contenant les étapes suivantes : (i) sélection de quelques individus (les parents), favorisant ceux avec une meilleure valeur de fitness; (ii) application des perturbations stochastiques sur les parents par le biais d’opérateurs de variation; (iii) évaluation de la valeur de fitness des enfants (i.e., des nouvelles solutions); et finalement, (iv) sélection de quelques individus parmi les parents et les enfants pour devenir les parents de la génération suivante, favorisant ici encore les individus avec une meilleure fitness. Cette boucle est répétée jusqu’à la découverte d’une solution satisfaisante, ou bien lorsqu’une autre condition d’arrêt est atteinte (e.g., ressources CPU épuisées).

Les AEs ont déjà démontré leur efficacité sur différents types de problèmes hors de portée de méthodes standards. Ce succès est dû principalement au fait que les AEs ne font pas de conjectures fortes sur le problème à optimiser : ils sont capables de traiter des espaces de recherche mixtes ou peu structurés, et des fonctions-objectif irrégulières, bruitées, rugueuses, fortement contraintes, etc. Malgré cela, les AEs sont rarement utilisés par des scientifiques d’autres domaines, car ils ne peuvent pas encore être considérés comme des outils “prêts à l’emploi”. Il y a beaucoup de raisons à cela, toutes se réduisant à un manque de support pratique lorsqu’il faut concevoir un AE pour un problème donné. Même si actuellement il existe quelques plate-formes logicielles (voir, e.g., EO [Keijzer et al., 2002] et GUIDE [Collet and Schoenauer, 2003; Da Costa and Schoenauer, 2009]) qui aident à modéliser les différents types d’approches évolutionnaires, le succès des AEs est très sensible à la définition des valeurs de leurs paramètres internes, par exemple la taille de la population, les types d’opérateur de variation et leur taux d’application, et les types de mécanismes de sélection.

Au début de leur existence, les AEs ont largement bénéficié de l’existence de ces paramètres, considérés comme une source de flexibilité qui a permis leur application à une gamme très vaste de domaines. La vision contemporaine reconnaît, cependant, que pour atteindre une performance satisfaisante, les paramètres des AEs doivent
être réglés spécialement pour chaque problème [Eiben et al., 2007]. A cause de cela, actuellement, le réglage de paramètres est vu comme le “talon d’Achille” des AE, au même titre que leur coût de calcul élevé. A partir de ces observations, la recherche et le développement de méthodes pour automatiser ce réglage est une tendance actuelle, très active à ce jour, ce dont témoigne la publication récente d’un livre [Lobo et al., 2007] et les nombreuses références récentes citées dans ce document [Eiben et al., 1999; Eiben et al., 2007]. Il est intéressant de noter que la recherche de méthodes pour automatiser le réglage de paramètres est aussi une priorité dans des domaines avoisinants, comme la Recherche Opérationnelle et la Programmation par Contraintes; de la même manière, ces domaines utilisent des algorithmes sophistiqués, qui requièrent une vaste expertise pour être utilisés au maximum de leur capacité [Hutter et al., 2006; Stützle, 2009].

Parmi les paramètres qui affectent le plus la performance de l’AE, il y a les choix qui concernent les opérateurs de variation : (i) lesquels parmi les opérateurs existants doivent être utilisés par l’AE pour la génération de nouvelles solutions, et (ii) à quel taux chacun des opérateurs choisis doit être appliqué pendant la résolution du problème. Il existe plusieurs types d’opérateurs, notamment les diverses façons de faire la mutation et le croisement. Chacun a ses propres caractéristiques, qui les amènent à affecter différemment le chemin effectué par l’algorithme dans l’espace de solutions : comment l’algorithme explore (globalement) l’espace de recherche, et comment il exploite (localement) les régions les plus prometteuses – le dilemme Exploration versus Exploitation (EvE) (voir [Eiben and Schippers, 1998] pour une discussion plus approfondie sur ce dilemme dans le contexte des AE). A cause de cela, les choix liés aux opérateurs de variation restent des décisions très sensibles et complexes, comme suit.

En premier lieu, la performance d’un opérateur donné dépend des caractéristiques du problème considéré. Comme il est très difficile de prédire la performance d’un opérateur sur un problème inconnu, le choix le plus naturel pour assister l’utilisateur dans ce sens est l’usage d’une technique de réglage hors-ligne : plusieurs expériences sont réalisées pour chaque configuration candidate, et le réglage est fait à partir de statistiques recueillies après ces expériences, e.g., la configuration qui obtient la meilleure performance en moyenne. Bien qu’étant une approche très coûteuse du point de vue du calcul, ce type de méthode parvient à trouver le meilleur jeu de paramètres … qui reste statique pendant toute la résolution du problème. Cependant, il est important de noter aussi que l’efficacité des opérateurs ne dépend pas que des caractéristiques globales du problème, mais aussi des caractéristiques locales de la région de l’espace de recherche qui est en train d’être explorée par la population. Enfin, leur performance dépend aussi de l’état du processus de recherche, i.e., si l’on s’approche de l’optimum, s’il y a beaucoup de diversité dans la population, etc. Par exemple, en suivant l’intuition sur le dilemme EvE, les opérateurs plus exploratoires peuvent être plus efficaces dans les premiers pas de la recherche; autrement, les opérateurs qui font un réglage plus fin des solutions (exploitation)
fonctionnent généralement mieux pendant les dernières étapes du processus.

Ces observations sont confirmées empiriquement par quelques expériences détaillées dans ce manuscrit : dans la plupart des problèmes considérés, il n’y a pas un opérateur unique qui reste le meilleur pendant toute la durée du processus de recherche. Basé sur ces résultats, et sur la nature stochastique des AEs, intuitivement, les meilleures configurations statiques trouvées par les méthodes de réglage hors-ligne ne permettent d’atteindre qu’une performance sous-optimale. Même si le problème d’optimisation est statique, le problème de sélection de l’opérateur est toujours dynamique : idéalement, le choix du meilleur opérateur doit être adapté en continu, pendant la résolution du problème, i.e., en-ligne.

Il existe différentes manières pour faire un réglage dynamique de paramètres en ligne (ce qui est couramment appelé Contrôle de Paramètres [Eiben et al., 2007]), notamment les approches dites auto-adaptatives et adaptatives. Les méthodes auto-adaptatives encodent les paramètres des opérateurs dans l’individu lui-même, et c’est le processus d’évolution qui est en charge d’optimiser, en même temps, les valeurs de ces paramètres et la solution pour le problème. Ces méthodes ont l’avantage d’avoir les paramètres réglés “gratuitement” par l’évolution elle-même, mais l’espace de recherche du problème d’optimisation est agrégé avec l’espace de recherche des configurations de paramètres, ce qui augmente considérablement la complexité de la recherche. De plus, ces méthodes sont intrinsèquement liées à la structure des AEs. Inversement, les méthodes adaptatives règlent les paramètres basés seulement sur l’histoire du processus d’optimisation courant, en ne modifiant pas la complexité originale du problème. Ils sont plus complexes à modéliser, mais comme l’adaptation est guidée par des mesures génériques sur l’avancement de la recherche, ils peuvent être facilement adaptés à d’autres types de meta-heuristiques et algorithmes de recherche, locale ou globale.

 Sélection Adaptative d’Opérateurs

En nous basant sur ces arguments, nous avons décidé d’attaquer le problème de sélection d’opérateurs avec des méthodes de contrôle de paramètre adaptatives, l’objectif étant de sélectionner en-ligne le meilleur opérateur à chaque étape de l’algorithme, pendant la résolution du problème. Nous nous référerons à ce paradigme sous la dénomination de Sélection Adaptative d’Opérateurs (SAO). La Figure 1 montre comment les méthodes de SAO peuvent être intégrées dans un AE, ce qui peut être décrit comme suit.

1. Pour la génération de chaque nouvelle solution (ou après n essais ou générations), l’AE demande à la SAO quel opérateur parmi ceux qui sont disponibles doit être appliqué.

2. La SAO donne la réponse guidée par son mécanisme de Sélection d’Opérateurs, qui sélectionne un opérateur basé sur les performances récentes des opérateurs.
disponibles. Ces performances sont le plus souvent représentées sous la forme d’estimations empiriques de leurs qualités.

3. L’opérateur choisi est appliqué par l’AE, une nouvelle solution candidate est générée, ce qui affecte le processus de recherche d’une certaine façon, e.g., en générant une solution qui est meilleure (ou pas) que ses parents (amélioration de la fitness), en variant la diversité moyenne de la population, etc.

4. Cette évaluation de l’impact est transformée en un crédit (ou récompense), selon le mécanisme d’Afféctation de Crédit employé.

5. Le crédit ou récompense est utilisé pour mettre à jour les estimations de qualité (ou performance) empiriques qui sont maintenues pour chaque opérateur, utilisées par le mécanisme de Sélection d’Opérateurs.

6. Cette boucle est utilisée en continu pendant la résolution du problème : elle peut donc être vue comme une méthode d’apprentissage par renforcement.

Suivant cette description, pour développer une méthode de SAO, il faut définir deux composants : (i) le schéma pour l’Affectation de Crédit, qui assigne un certain crédit à l’opérateur, à partir de l’impact de ses applications récentes sur le processus de recherche courant; et (ii) le mécanisme de Sélection d’Opérateurs, qui sélectionne le prochain opérateur à appliquer, guidé par la base de connaissance construite et maintenue par ces évaluations empiriques.

Contributions Principales

Dans cette thèse, différentes propositions ont été faites pour chacun des composants cités ci-dessus, ainsi que pour leur évaluation empirique. Voici un bref résumé sur ces contributions, par ordre chronologique.
La Sélection d’Opérateurs a été abordée en tant qu’instance du dilemme Exploitation/Exploitation : le meilleur opérateur courant doit être appliqué le plus souvent possible (exploitation), cependant que les autres opérateurs doivent aussi être appliqués de temps en temps (exploration). L’exploration est nécessaire pour deux raisons principales : tout d’abord, un opérateur peut être malchanceux à un moment donné, et à cause de cela recevoir de mauvaises récompenses ; d’autre part, dû à la dynamique des algorithmes évolutionnaires, i.e., un des opérateurs sous-optimaux peut éventuellement devenir le meilleur à une étape ultérieure de la résolution du problème.

Ce dilemme a été intensivement étudié dans le contexte de la Théorie des Jeux, plus spécifiquement dans le cadre des problèmes dits de “Bandits Manchots” (BM) [Lai and Robbins, 1985; Auer et al., 2002]. L’utilisation d’algorithmes de BM pour trouver un compromis optimal entre l’exploration et l’exploitation dans un cadre d’optimisation a été analysée pour la sélection au sein de portfolios d’algorithmes différents pour résoudre des problèmes de décision [Gagliolo and Schmidhuber, 2008], avant d’être étendu au contexte de la SAO dans le travail présenté ici. Notre première tentative a été d’utiliser directement un algorithme pour BM appelé “Upper Confidence Bound” (UCB) [Auer et al., 2002], qui a été choisi pour fournir des garanties d’optimalité asymptotique par rapport à la récompense totale cumulée. Cependant, ces garanties sont conservées seulement dans le cadre de problèmes statiques ; quelques modifications ont donc dû être proposées pour utiliser l’algorithme UCB d’une façon efficace dans le contexte dynamique de la SAO – c’est là que se concentrent la plupart des contributions développées dans cette thèse.

La première proposition, appelée “Dynamic Multi-Armed Bandit” (DMAB), combine l’algorithme UCB avec le test statistique de Page-Hinkley [Hinkley, 1970], qui est utilisé pour détecter de changements dans la distribution de récompenses. Une fois un changement détecté, le processus du BM est totalement réinitialisé, ce qui permet la découverte rapide du nouveau meilleur opérateur [Da Costa et al., 2008].

Concernant l’Affectation de Crédit, la plupart des combinaisons de SAO trouvées dans la littérature utilisent des statistiques simples basées sur les améliorations de fitness des solutions. Au lieu d’utiliser la “moyenne” des performances récentes, ou la valeur atteinte par la toute dernière application de cet opérateur (“instantanée”), nous avons proposé l’usage de valeurs “extrêmes”, i.e., la récompense que l’opérateur reçoit est égale au maximum des améliorations de fitness (ou d’autres évaluations d’impact) obtenues par ses applications récentes. L’hypothèse pour soutenir ce choix est que les améliorations plus rares mais plus élevées peuvent avoir des conséquences plus importantes pour le résultat final que les améliorations plus régulières mais plus modérées.

La combinaison entre l’Affectation de Crédit “extrême” et la Sélection d’Opérateurs “DMAB”, appelée technique de SAO Ex-DMAB, s’est montré très efficace, dépassant d’autres approches de base dans plusieurs scénarios de benchmark différents [Fialho et al., 2008; Fialho et al., 2009a; Maturana et al., 2009a]. Néan-
moins, l’usage direct de valeurs brutes d’amélioration de fitness pour mettre à jour les préférences de la SAO a rapidement montré ses limites : des problèmes différents ont des espaces de recherche de caractéristiques très diverses, ce qui affecte l’écart des récompenses reçues. Par conséquent, pour atteindre une bonne performance, cette méthode de SAO nécessite un réglage fin de ses propres paramètres pour chaque nouveau problème. Pour cette raison, nous avons ensuite proposé l’utilisation d’un simple schéma de normalisation [Fialho et al., 2009b].

Du coté de la Sélection d’Opérateurs, même avec des récompenses normalisées, l’hyper-paramètre de l’algorithme DMAB contrôlant la détection de changements a continué à dépendre très fortement du problème, car le mécanisme de réinitialisation est directement lié à la dynamique de l’espace de recherche du problème. Cela a été le facteur motivant pour la proposition d’une manière plus douce d’adapter l’algorithme UCB à l’environnement dynamique de la SAO, ce que nous avons appelé le “Sliding Multi-Armed Bandit” (SLMAB). SLMAB utilise une fenêtre glissante pour mettre à jour les estimations empiriques de qualité pour chaque opérateur, en rejetant les événements trop anciens et en ne gardant que l’information des applications d’opérateur plus récentes. Par rapport à DMAB, SLMAB n’a pas besoin d’un “observateur” externe pour contrôler les changements de situation. Grâce à cela, cette méthode réussit à adapter l’UCB au contexte dynamique sans augmenter le nombre d’hyper-paramètres.

Avec l’usage de la normalisation, l’effet de la dépendance du problème est lissé, mais pas éliminé. Cela nous a amené à proposer les deux dernières méthodes pour l’Affectation de Crédit, complètement basées sur les rangs, le “Area Under Curve” (AUC) et le “Sum-of-Ranks” (SR) [Fialho et al., 2010c]. En plus du gain de robustesse atteint par l’utilisation de mesures basées sur le rang, l’usage de rangs sur les valeurs de fitness (respectivement appelés FAUC et FSR dans ce cas), au lieu de rangs sur les valeurs d’amélioration de fitness, permet éventuellement de préserver une propriété d’invariance très importante, l’invariance par rapport à toute transformation monotone de la fonction-objectif, du fait qu’il n’utilise que des comparaisons entre solutions, sans tenir compte des valeurs exactes de fitness prises par ces solutions (en anglais, “comparison-based algorithms”). Ces schémas d’Affectation de Crédit basés sur le rang ont été combinés avec une version simplifiée de l’UCB, laquelle est appelée le “Rank-based Multi-Armed Bandit” (RMAB).

En outre, pendant la phase de développement de ces combinaisons de SAO, nous avons aussi proposé plusieurs scénarios artificiels pour leur analyse empirique. Les scénarios Boolean et Outlier [Da Costa et al., 2008], inspirés par le scénario Uniform [Thierens, 2005], ont été introduits pour évaluer les schémas de SAO dans de situations incluant 5 opérateurs artificiels avec de distributions de récompense différentes. Par ailleurs, une autre famille de scénarios artificiels, appelée “Two-Values” (TV) benchmark, a été proposée [Fialho et al., 2010a] pour simuler différentes situations par rapport à la moyenne et à la variance des récompenses données par deux opérateurs artificiels.
Analyse Empirique

La dernière contribution de cette thèse consiste en une comparaison empirique très complète des méthodes proposées. Leur performance a été analysée dans plusieurs scénarios de benchmark, avec de caractéristiques et niveaux de complexité différents, ce qui a rendu possible l'analyse de différents aspects comportementaux des méthodes de SAO, qui vont être rapidement décrits ci-dessous.

Les expériences sur les scénarios artificiels déjà cités ont permis d’analyser, par exemple, l’efficacité des méthodes de SAO pour s’adapter à de situations complètement différentes, dans divers contextes en ce qui concerne la distribution et le niveau d’informativité des récompenses reçues. De plus, quelques vrais problèmes d’optimisation, des problèmes test très connus dans la communauté d’AEs, ont été utilisés : “OneMax”, “Long K-Path” et “Royal Road”. Ces derniers représentent les premières expériences pour les méthodes de SAO dans un contexte plus réaliste que celui de scénarios artificiels, en choisissant parmi plusieurs opérateurs de mutation et croisement au sein d’un vrai Algorithme Génétique (AG), appliqué à des problèmes simples, mais avec des caractéristiques très diverses. Par ailleurs, de résultats ont aussi été obtenus dans le contexte d’une classe de problèmes combinatoires difficiles, les problèmes de satisfiabilité booléenne (SAT). Dans ce scénario, seule la méthode DMAB a été analysée, en combinaison avec un schéma d’Affectation de Crédit particulier proposé par des collègues de l’Université d’Angers, le Compass [Maturana and Saubion, 2008a]. Le Compass prend en compte les variations de fitness et diversité pour récompenser l’opérateur. Finalement, ces méthodes ont été aussi analysées dans le contexte d’un ensemble très complet de fonctions continues mono-objectifs, cette fois-ci en choisissant parmi différentes stratégies de mutation de l’algorithme Evolution Différentielle (DE, Differential Evolution). L’hétérogénéité de cet ensemble de tests a aussi rendu possible l’analyse de la robustesse et de la sensibilité des méthodes de SAO par rapport à leurs hyper-paramètres.

En effet, avant toute expérience, les hyper-paramètres des méthodes de SAO ont été réglés par le biais de F-Race [Birattari et al., 2002], une technique de réglage hors-ligne, afin de ne comparer que leurs meilleures performances. Étant donné un sous-ensemble d’opérateurs, les méthodes proposées ont été comparées entre elles, et avec d’autres approches utilisées comme référence :

1. Seul : le même AE, mais sans SAO, en utilisant seulement un opérateur (pour chaque opérateur considéré) ;
2. Statique : le même AE, sans SAO, mais avec le taux d’application de chaque opérateur réglé a priori par l’algorithme F-Race [Birattari et al., 2002] ;
3.Uniforme : le choix naïf, i.e., la sélection uniforme entre les opérateurs disponibles ;
4. Optimal : le choix optimal, i.e., la sélection du meilleur opérateur pour chaque moment de la recherche, disponible seulement pour quelques scénarios ;
5. Et la méthode SAO état de l’art, Adaptive Pursuit (AP) [Thierens, 2007].

Même dans les cas des scénarios artificiels et booléens, dans lesquels les paramètres des méthodes ont été réglés pour chaque problème, les différentes combinaisons basées sur les rangs (RMAB + AUC/SR/FAUC/FSR) se sont montrées capables de suivre de près la performance de la meilleure SAO dans la plupart des problèmes. Et, plus le scénario devenait hétérogène, i.e., plus la quantité de problèmes différents résolus en utilisant le même réglage de paramètres était grand, et plus la différence de performance entre les méthodes basées sur les rangs et les autres est devenue importante, comme on peut le voir sur les résultats obtenus dans le scénario mettant en jeu DE avec des fonctions continues, où RMAB est clairement le grand vainqueur.

Conclusions et Discussion

Pour résumer tous ces résultats empiriques, au moment où ce manuscrit est écrit, la technique de SAO constituée par le RMAB pour la Sélection d’Opérateurs, et l’AUC pour l’Affectation de Crédit, est la combinaison recommandée pour la mise en oeuvre du paradigme de SAO. Cette méthode atteint une performance égale, ou très proche, de celle des meilleures méthodes classiques connues, tout en étant très robuste par rapport à ses hyper-paramètres, comme l’ont confirmées les expériences réalisées [Fialho et al., 2010c; Fialho et al., 2010b].

Il est important de noter que la réalisation du réglage hors-line avant chaque comparaison empirique a été motivée par l’intention de comparer les différents méthodes de SAO à leur meilleur niveau. Car de fait, idéalement, à chaque fois qu’un nouveau problème doit être abordé, aucun réglage hors-ligne ne devrait être nécessaire – et c’est pour cela que tant d’effort a été fait pour tenter d’augmenter la robustesse des méthodes proposées. Dans le cas de la méthode de SAO recommandée, la même configuration pour les hyper-paramètres (C.5W50) s’est révélée la meilleure dans des conditions expérimentales très différentes, dans le contexte des fonctions continues, et des valeurs avoisinantes (étant donné l’ensemble de valeurs essayées) ont été trouvées pour les problèmes OneMax et Royal Road (C.1W100) et pour la plupart des problèmes artificiels (C ∈ {1, 5, 1}, W ∈ {50, 100}). Ceci dit, l’utilisation du réglage hors-ligne reste une étape optionnelle, car des performances acceptables peuvent être atteintes par la méthode de SAO recommandée une fois ses hyper-paramètres configurés avec de valeurs proches des configurations citées. Et même si l’utilisateur opte pour l’utilisation du réglage hors-ligne, la méthode de SAO recommandée n’a que deux hyper-paramètres, tandis que dans l’option standard (l’AE sans SAO) il faut choisir les opérateurs, puis définir le taux d’application pour chaque opérateur (i.e., le nombre de paramètres qui doivent être réglés est un multiple du nombre d’opérateurs considérés), en se rappelant que ce dernier réglage reste statique pendant toute la durée du processus de recherche, ce qui dans la plupart des cas ne peut pas être optimal.
Maintenant que nous disposons d’une méthode de SAO efficace et robuste, il faut l’étendre en prenant en compte d’autres types de complexité. Les prochaines étapes pour ce travail de recherche seront, naturellement, son extension pour atteindre une meilleure efficacité sur des problèmes fortement multi-modaux, problèmes pour lesquels il faut probablement également prendre en compte l’évolution de la diversité dans la récompense des opérateurs. De même, dans les extensions au contexte multi-objectif qui sont envisagées, il faudra imaginer une récompense qui prendra en compte à la fois la convergence vers le front de Pareto et la diversité des solutions non-dominées.
# Contents

## Part I  General Introduction

1  Introduction  
   1.1  Context/Motivation  
   1.2  Main Contributions  
      1.2.1  Operator Selection  
      1.2.2  Credit Assignment  
      1.2.3  Empirical Validation  
   1.3  Organization  

## Part II  Background Review

2  Evolutionary Algorithms  
   2.1  Introduction  
   2.2  Modus Operandi  
   2.3  Components  
      2.3.1  Problem-dependent Components  
      2.3.2  Representation-specific Components  
      2.3.3  General Components  
   2.4  Popular EA Variants  
      2.4.1  Evolution Strategies  
      2.4.2  Evolutionary Programming  
      2.4.3  Genetic Programming  
      2.4.4  Genetic Algorithms  
      2.4.5  Differential Evolution  
   2.5  Application Areas  
   2.6  Discussion  

xvii
### 3 Parameter Setting in EAs

3.1 Introduction ................................................. 30

3.2 Parameters Influence and Possible Settings .......................... 31
  3.2.1 Parent and Offspring Population Sizes .................. 31
  3.2.2 Selection Procedures .................................. 32
  3.2.3 Offspring Production .................................... 32
  3.2.4 Stopping Criterion ...................................... 33
  3.2.5 Representation .......................................... 34

3.3 Classification of Parameter Setting Techniques ..................... 34
  3.3.1 Which parameter is changed? ............................ 34
  3.3.2 How the changes are made? ............................. 35
  3.3.3 Which evidences guide the changes? ..................... 40
  3.3.4 Which is the scope of the change? ....................... 40

3.4 Discussion .................................................... 41

### 4 Adaptive Operator Selection

4.1 Introduction .................................................. 44

4.2 Adaptive Operator Selection .................................... 45

4.3 Credit Assignment ............................................. 46
  4.3.1 How to measure the Impact? ............................. 46
  4.3.2 How to assign Credit? .................................... 47
  4.3.3 Whom to assign Credit to? ............................... 48
  4.3.4 Compass: Aggregating Fitness and Diversity ............. 48

4.4 Operator Selection .............................................. 50
  4.4.1 Probability Matching .................................... 50
  4.4.2 Adaptive Pursuit ....................................... 52

4.5 Some Adaptive Operator Selection Combinations .................... 53
  4.5.1 Fitness-based Approaches ................................ 54
  4.5.2 Diversity-based Approaches ............................. 56
  4.5.3 Fuzzy-based Approaches ................................. 57
  4.5.4 Other Approaches ....................................... 58
  4.5.5 AOS within Other Evolutionary Algorithms ............... 59

4.6 Discussion ..................................................... 59

### Part III Contributions

5 Contributions to Adaptive Operator Selection .......................... 65
  5.1 Introduction .................................................. 66
  5.2 Contributions to Credit Assignment ............................ 68
    5.2.1 Basic Credit Assignment Scheme: Fitness Improvements ... 68
5.2.2 Extreme Fitness Improvement ........................................... 69
5.2.3 Normalized Fitness Improvement ..................................... 70
5.2.4 Rank-based Credit Assignment Schemes ............................... 70
5.2.5 Comparison-based Credit Assignment Schemes ..................... 77
5.3 Contributions to Operator Selection ..................................... 78
5.3.1 Basic Operator Selection Scheme: Multi-Armed Bandit ............. 78
5.3.2 Dynamic Multi-Armed Bandit ......................................... 81
5.3.3 Sliding Multi-Armed Bandit .................................. 84
5.3.4 Rank-based Multi-Armed Bandit ..................................... 86
5.4 Contributions to Empirical Assessment .................................. 88
5.4.1 Basic Artificial Scenario: Uniform .................................. 89
5.4.2 Boolean and Outlier Scenarios ....................................... 89
5.4.3 Two-Value Scenarios ........................................ 90
5.5 Discussion ......................................................................... 92

6 Experimental Results ............................................................. 95
6.1 Introduction ....................................................................... 97
6.2 General Experimental Settings .......................................... 98
6.2.1 AOS Combinations and Respective Hyper-Parameters ............ 98
6.2.2 Off-line Tuning of Hyper-Parameters ................................ 100
6.2.3 Performance Indicators and Results Presentation .................. 102
6.3 On Artificial Scenarios ....................................................... 103
6.3.1 Experimental Settings ................................................. 104
6.3.2 Results on Uniform, Boolean and Outlier Scenarios .............. 104
6.3.3 Results on ART Scenarios ........................................... 120
6.3.4 Discussion....................................................................... 130
6.4 On Boolean Benchmark Problems ...................................... 131
6.4.1 Experimental Settings .................................................. 131
6.4.2 The OneMax Problem .................................................. 132
6.4.3 The Long k-Path Problem ............................................ 138
6.4.4 The Royal Road Problem ............................................. 143
6.4.5 Discussion....................................................................... 147
6.5 Collaboration On Satisfiability Problems .............................. 148
6.5.1 Compass + Ex-DMAB = ExCoDyMAB ......................... 148
6.5.2 Boolean Satisfiability Problems ...................................... 149
6.5.3 Experimental Settings ............................................... 150
6.5.4 Architecture definition and tuning of hyper-parameters ........ 151
6.5.5 Empirical Results ...................................................... 152
6.5.6 Discussion ................................................................... 154
6.6 On Continuous Benchmark Problems .................................. 156
6.6.1 Black-Box Optimization Benchmarking ......................... 156
6.6.2 Experimental Settings .................................................. 157
6.6.3 The PM-AdapSS-DE Method .......................................... 159
6.6.4 Empirical Results ...................................................... 160
Part IV  General Conclusion

7  Final Considerations
   7.1  Summary of Contributions
   7.2  Further Work

Bibliography
List of Figures

1. Le schéma général de la Sélection Adaptative d’Opérateurs.  
   General cycle of Evolutionary Algorithms.  
2. Classification of parameter setting methods.  
5. Comparison between different decaying mechanisms.  
6. Comparison between AUC computations with different decay.  
7. Two samples drawn from different Two-Values distributions.  
8. TCR and \(p(\text{best})\) w.r.t. \(\Delta T\) on Uniform scenario.  
10. TCR and \(p(\text{best})\) w.r.t. \(\Delta T\) on Boolean scenario.  
12. TCR and \(p(\text{best})\) w.r.t. \(\Delta T\) on Outlier scenario.  
14. TCR and \(p(\text{best})\) w.r.t. \(\Delta T\) on \(\text{ART}(0.01, 101, 0.5, 10)\) scenario.  
15. Behavior of AOS combinations on the \(\text{ART}(0.01, 101, 0.5, 10)\) scenario.  
16. TCR and \(p(\text{best})\) w.r.t. \(\Delta T\) on \(\text{ART}(0.1, 39, 0.5, 3)\) scenario.  
17. Behavior of AOS combinations on the \(\text{ART}(0.1, 39, 0.5, 3)\) scenario.  
21. Average fitness gain of operators on the Long \(k\)-Path problem.  
22. Different levels of deceptiveness on the Royal Road problem.  
23. Best fitness curve for different values of \(\Theta\) on Compass.  
24. ECDF of all rank-based schemes on BBOB with \(d = 20\).  
25. ECDF of speed-up ratios: FAUC-B vs. DE using single strategy.  
26. ECDF of speed-up ratios: FAUC-B vs. other AOS schemes.  
27. ECDF of speed-up ratios: FAUC-B vs. Naive, Static and IPOP-CMA-ES.  
28. ECDF sensitivity plots for AbsExt-AP.
List of Figures

6.21 ECDF sensitivity plots for AbsExt-MAB and AbsExt-SLMAB . . . . . . . 173
6.22 ECDF sensitivity plots for AbsExt-DMAB and Decay/AUC-RMAB . . . . 175
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Two examples of static sets of parameters for GAs</td>
<td>30</td>
</tr>
<tr>
<td>5.1</td>
<td>Sample list of 15 operator applications ordered by $\Delta F$</td>
<td>74</td>
</tr>
<tr>
<td>6.1</td>
<td>Credit Assignment schemes and hyper-parameters</td>
<td>99</td>
</tr>
<tr>
<td>6.2</td>
<td>Operator Selection methods and hyper-parameters</td>
<td>100</td>
</tr>
<tr>
<td>6.3</td>
<td>List of Adaptive Operator Selection combinations</td>
<td>100</td>
</tr>
<tr>
<td>6.4</td>
<td>Ranges of values tried for the corresponding hyper-parameters</td>
<td>101</td>
</tr>
<tr>
<td>6.5</td>
<td>Results on the Uniform scenario for $\Delta T \in {50, 200}$</td>
<td>107</td>
</tr>
<tr>
<td>6.6</td>
<td>Results on the Uniform scenario for $\Delta T \in {500, 2000}$</td>
<td>108</td>
</tr>
<tr>
<td>6.7</td>
<td>Results on the Boolean scenario for $\Delta T \in {50, 200}$</td>
<td>112</td>
</tr>
<tr>
<td>6.8</td>
<td>Results on the Boolean scenario for $\Delta T \in {500, 2000}$</td>
<td>113</td>
</tr>
<tr>
<td>6.9</td>
<td>Results on the Outlier scenario for $\Delta T \in {50, 200}$</td>
<td>117</td>
</tr>
<tr>
<td>6.10</td>
<td>Results on the Outlier scenario for $\Delta T \in {500, 2000}$</td>
<td>118</td>
</tr>
<tr>
<td>6.11</td>
<td>Results on ART(0.01, 101, 0.5, 10), 10 epochs for $\Delta T \in {50, 200}$</td>
<td>122</td>
</tr>
<tr>
<td>6.12</td>
<td>Results ART(0.01, 101, 0.5, 10), 2 epochs for $\Delta T \in {500, 2000}$</td>
<td>123</td>
</tr>
<tr>
<td>6.13</td>
<td>Results ART(0.1, 39, 0.5, 3), 10 epochs for $\Delta T \in {50, 200}$</td>
<td>127</td>
</tr>
<tr>
<td>6.14</td>
<td>Results ART(0.1, 39, 0.5, 3), 2 epochs for $\Delta T \in {500, 2000}$</td>
<td>128</td>
</tr>
<tr>
<td>6.15</td>
<td>Results on the 10k-bits OneMax problem</td>
<td>136</td>
</tr>
<tr>
<td>6.16</td>
<td>Examples of Long 3-Paths of different length</td>
<td>139</td>
</tr>
<tr>
<td>6.17</td>
<td>Results on the Long 3-Path ($\ell = 49$) problem</td>
<td>142</td>
</tr>
<tr>
<td>6.18</td>
<td>Construction of schemata on the Royal Road problem</td>
<td>143</td>
</tr>
<tr>
<td>6.19</td>
<td>Results on the Royal Road ($m = 4$) problem</td>
<td>146</td>
</tr>
<tr>
<td>6.20</td>
<td>SAT instances used in the empirical assessment of ExCoDyMAB</td>
<td>150</td>
</tr>
<tr>
<td>6.21</td>
<td>Racing survivors after ExCoDyMAB hyper-parameters tuning</td>
<td>153</td>
</tr>
<tr>
<td>6.22</td>
<td>Comparison between configurations on SAT instances</td>
<td>154</td>
</tr>
<tr>
<td>6.23</td>
<td>Empirical results on the 22 SAT instances</td>
<td>155</td>
</tr>
<tr>
<td>6.24</td>
<td>Hyper-parameter configurations used on BBOB dimension 20</td>
<td>159</td>
</tr>
<tr>
<td>6.25</td>
<td>Off-line tuned application rates on BBOB dimension 20</td>
<td>159</td>
</tr>
<tr>
<td>6.26</td>
<td>Robustness analysis on transformations over the OneMax problem</td>
<td>177</td>
</tr>
<tr>
<td>6.27</td>
<td>Robustness analysis on the BBOB functions</td>
<td>179</td>
</tr>
</tbody>
</table>
# List of Algorithms

<table>
<thead>
<tr>
<th>Section</th>
<th>Algorithm Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>General pseudo-algorithm for a Genetic Algorithm</td>
<td>23</td>
</tr>
<tr>
<td>2.2</td>
<td>The Differential Evolution algorithm with DE/rand/1/bin strategy</td>
<td>24</td>
</tr>
<tr>
<td>4.1</td>
<td>Credit Assignment: Compass ($K, \Theta$)</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Operator Selection: Probability Matching ($K, p_{min}, \alpha$)</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>Operator Selection: Adaptive Pursuit ($K, p_{min}, \alpha, \beta$)</td>
<td>53</td>
</tr>
<tr>
<td>5.1</td>
<td>Credit Assignment Schemes over $\Delta F$ (op, type, norm, W, K)</td>
<td>71</td>
</tr>
<tr>
<td>5.2</td>
<td>Credit Assignment: Rank-based Area-Under-Curve (W, D, op)</td>
<td>77</td>
</tr>
<tr>
<td>5.3</td>
<td>Credit Assignment: Sum-of-Ranks (W, D, op)</td>
<td>78</td>
</tr>
<tr>
<td>5.4</td>
<td>Operator Selection: Multi-Armed Bandit ($K, C$)</td>
<td>81</td>
</tr>
<tr>
<td>5.5</td>
<td>Operator Selection: Dynamic MAB ($K, C, \gamma, \delta = 0.15$)</td>
<td>84</td>
</tr>
<tr>
<td>5.6</td>
<td>Operator Selection: Sliding Multi-Armed Bandit ($K, C, W$)</td>
<td>86</td>
</tr>
<tr>
<td>5.7</td>
<td>Operator Selection: Rank-based Multi-Armed Bandit ($K, C$)</td>
<td>88</td>
</tr>
</tbody>
</table>
List of Acronyms

**AbsIns** Absolute Instantaneous
**AbsAvg** Absolute Average
**AbsExt** Absolute Extreme
**AOM** Adaptive Operator Management
**AOS** Adaptive Operator Selection
**AP** Adaptive Pursuit
**AUC** Area-Under-Curve
**BBOB** Black-Box Optimization Benchmarking
**ExCoDyMAB** Extreme Compass - DMAB
**DE** Differential Evolution
**DMAB** Dynamic Multi-Armed Bandit
**EA** Evolutionary Algorithm
**EC** Evolutionary Computation
**ECDF** Empirical Cumulative Distribution Function
**EP** Evolutionary Programming
**ERT** Expected Running Time
**ES** Evolution Strategy
**EvE** Exploration versus Exploitation
**FAUC** Fitness-based Area-Under-Curve
**FSR** Fitness-based Sum-of-Ranks
**GA** Genetic Algorithm
List of Acronyms

GP  Genetic Programming
MAB  Multi-Armed Bandit
NDCG  Normalized Discounted Cumulative Gain
NormIns  Normalized Instantaneous
NormAvg  Normalized Average
NormExt  Normalized Extreme
PH  Page-Hinkley
PM  Probability Matching
RMAB  Rank-based Multi-Armed Bandit
ROC  Receiver Operating Characteristic
SAT  Boolean Satisfiability
SLMAB  Sliding Multi-Armed Bandit
SR  Sum-of-Ranks
TCR  Total Cumulative Reward
TSP  Traveling Salesman Problem
TV  Two-Values
UCB  Upper Confidence Bound
Part I

General Introduction
Chapter 1

Introduction

1.1 Context/Motivation

EAs are stochastic optimization algorithms remotely inspired by the Darwinian “survival of the fittest” paradigm. Let the goal be to optimize some objective function, referred to as fitness function, defined on search space $X$; elements of $X$ are called individuals, and a set of individuals is termed a population. EAs evolve a population of individuals by iteratively (i) selecting some individuals (the parents), favoring those with better fitness; (ii) applying stochastic perturbations on the parents using some variation operators, thus generating offspring; (iii) evaluating the offspring (i.e., computing their fitness values); and finally, (iv) selecting some individuals among the parents and the offspring to become the next parents, again favoring fitter individuals. This cycle is iterated until a satisfactory solution is found, or another stopping condition is attained. A more comprehensive description is presented in Chapter 2.

EAs have already shown to be efficient optimization methods on many different types of problems beyond the reach of standard methods: see, e.g., all applications described in [Yu et al., 2008], and the very diverse works that are presented every year in the “Real-World Applications” track\(^1\) of the ACM Genetic and Evolutionary Computation Conference (GECCO). This success is achieved specially because EAs do not make any strong assumption about the problem to be solved: they are able to handle structured and mixed search spaces, irregular, noisy, rugged, or highly constrained objective functions, etc. But, despite this, EAs are rarely used outside the circle of knowledgeable practitioners. They still miss reaching the status of off-the-shelf tools. There are several reasons for this, all boiling down to a lack of practical support when it comes to actually design an EA for a given application. On a conceptual level, despite Michalewicz’ seminal book [Michalewicz, 1996] and the two more recent books by [Eiben and Smith, 2003] and [De Jong, 2006], the terminology used by many authors still reflect the evolutionary trend they historically belong to. On a practical level, while some software packages provide a unifying framework for the various evolutionary approaches (see, e.g., the EO [Keijzer et al., 2002] and the GUIDE [Collet and Schoenauer, 2003;\(^3\)](http://sigevo.org/gecco-2010/organizers-tracks.html#rwa)

\(^1\)RWA track on GECCO’10: http://sigevo.org/gecco-2010/organizers-tracks.html#rwa
The success of EAs is still very sensitive to the setting of quite a few parameters. Examples are the population size, the types of variation operators and respective application rates, and the types of selection mechanisms.

In early days, Evolutionary Computation (EC) actually benefited from those numerous parameters, considering them as a source of flexibility that enabled the application of EAs to the mentioned wide spectrum of applications. The contemporary view of EAs, however, acknowledges that specific problems require specific setups for satisfactory performance: when it comes to solving a given problem, parameter setting is viewed as the Achilles’ heel of EAs, on par with their high computational cost. From these observations, a current trend in EC is to focus on the definition of more autonomous solving processes, which aim at enabling the basic user to benefit from a more efficient and easy-to-use algorithmic framework. Parameter setting in EAs appears thus as a major issue that has deserved much attention during recent years. Research on this topic is still very active nowadays, as witnessed by a complete edited book that has been recently published, and by the numerous recent references cited in this document. The current state-of-the-art of research in parameter setting of EAs is summarized in Chapter 3. Interestingly, the search for algorithmic technologies enabling the (naive) end-user to benefit from good performances through autonomous parameter setting is also considered as a priority in neighboring fields, such as operation research or constraint programming; in the same way as in EC, these fields involve sophisticated solver platforms, requiring an extensive expertise in order to be used to their fullest extent.

Some of the user choices that most affect the performance of EAs concern the variation operators: which operators should be used for the generation of new solutions, and at which rate should be applied each of the chosen operators. These choices affect the way in which the algorithm (globally) explores the search space, and how it (locally) exploits the most promising regions—the so-called Exploration versus Exploitation (EvE) dilemma (we refer the reader to [Eiben and Schippers, 1998] for a comprehensive overview devoted to the EvE balance in EAs). The setting of these parameters is usually done by following the user’s intuition, or by using an off-line tuning procedure aimed at identifying the best operator for the problem at hand. Besides being computationally expensive, off-line tuning, however, generally delivers sub-optimal performance. Intuitively, the EA should proceed from a global (early) exploration of the landscape to a more focused exploitation-like behavior, as already empirically and theoretically demonstrated (see, e.g., [Eiben et al., 2007] and references therein). Thus, its parameter values should be varied accordingly, while solving the problem (i.e., on-line): more exploratory operators should be preferred in the earlier stages of the search, and more priority should be given to the fine-tuning/exploitation operators when approaching the optimum.

One of the approaches for controlling on-line the application of the variation operators is the so-called Adaptive Operator Selection (AOS). This is the context of the contributions proposed in this work, which will be summarized in the following. A more detailed presentation will be done, respectively, in Chapters 4 and 5.
1.2 Main Contributions

In essence, the goal of AOS is to select on the fly the best operator at each stage of the search, i.e., the operator that is currently maximizing some measure of quality, usually, though not exclusively [Maturana et al., 2009a], reflecting the fitness improvement brought by its application. AOS requires two tasks to be solved, the Operator Selection and the Credit Assignment. These tasks will be explained in the following, together with the contributions proposed in this thesis to address each of them.

1.2.1 Operator Selection

The first task, Operator Selection, defines how the next operator to be applied should be selected, based on its known empirical quality. Indeed, it might be seen as yet another level of the EvE dilemma. While an operator that has performed well in the recent past should certainly be used again (exploitation), other operators that did not perform so well should also be tried (exploration). The rationale for exploration is rooted, firstly, in the stochastic nature of the evolutionary process (some seemingly poorly performing operators might just have been unlucky); and secondly, on its dynamics: the quality of an operator depends on the region of the fitness landscape being explored by the current population, i.e., good operators might become poor as evolution goes on, and vice-versa. These changes in operator qualities are empirically confirmed in most of the benchmarking scenarios tackled in the experimental section of this manuscript.

Notably, the EvE trade-off has been intensively studied in the context of Game Theory, in the so-called Multi-Armed Bandit (MAB) framework [Lai and Robbins, 1985]. The Upper Confidence Bound (UCB) [Auer et al., 2002] is a MAB algorithm that provides asymptotic optimality guarantees with respect to the total cumulative reward in a stationary context. However, as previously mentioned, the AOS context is dynamic. The main contribution of this thesis, in summary, lies in the proposal and analysis of schemes to solve the AOS problem based on the UCB algorithm; we have proposed different extensions to it, in order to enable it to efficiently cope with the dynamics of evolution and with the very different characteristics of the problems to be tackled.

Starting from the original UCB algorithm (referred to as the original or standard MAB algorithm in the following, for the sake of convenience), presented in Section 5.3.1, our first proposal to extend it to the dynamic context of AOS was the Dynamic Multi-Armed Bandit (DMAB) algorithm [Da Costa et al., 2008]. It proceeds by coupling the original MAB technique with a statistical change-point test, the Page-Hinkley test [Hinkley, 1970]: upon the detection of a change in the operator quality distribution, the MAB process is restarted from scratch.

Although showing to be very efficient, the DMAB required the tuning of a very sensitive and problem-dependent hyper-parameter, the threshold value for the change-detection test. This led to the proposal of a smoother way to account for dynamic environments in the MAB framework, referred to as Sliding Multi-Armed Bandit (SLMAB) [Fialho et al., 2010a]. It uses a sliding time window to gracefully update the operator quality estimates, discarding ancient events while preserving the information from the recent
Contrasting with DMAB, the SLMAB does not call upon an external monitoring of the evolution process, involving only 1 hyper-parameter, while DMAB has two.

The latest proposal concerning the Operator Selection part is what we refer to as Rank-based Multi-Armed Bandit (RMAB), in which the evaluations provided by a rank-based Credit Assignment scheme (that is part of the contributions that will be presented in the following sub-section) are used directly in the place of the UCB empirical estimation. In this way, as the rewarding of one operator affects the ranks, and consequently the quality assessments, of all the other operators, this technique is already dynamic by definition, while being very robust with respect to its hyper-parameters.

1.2.2 Credit Assignment

All the previously mentioned bandit-based Operator Selection methods (and other existing approaches for the same purpose) select the operator to be applied next based on some assessment of their respective qualities. Defining how to estimate their quality based on the impact brought by their most recent applications is what we refer to as Credit Assignment, the second task to be defined for an AOS algorithm.

The most common way of assigning credit is to account for the fitness improvements brought by the operators applications. The use of the instantaneous value, i.e., the latest fitness improvement achieved by the operator, is known to be an unstable measure, due to the stochastic nature of operators (one operator might have been unlucky on its latest trial, although being a very good option in the longer term). To alleviate such effect, the average of the latest rewards is more commonly used; however, by using such kind of assessment, operators that regularly achieve very small improvements are preferred over operators that get rare but highly beneficial improvements. Motivated by other complex systems (e.g., rogue waves, financial market, etc), and also by another empirical analysis within EAs [Whitacre et al., 2006], in this thesis we support that the latter operator should be preferred in the place of the former. This can be achieved by the use of Extreme values, i.e., the maximum reward recently received by the operator [Fialho et al., 2008]. Better results have been achieved when compared to the usual Instantaneous and Average schemes.

Nevertheless, with the use of the raw values of the fitness improvements, the AOS schemes implementing these Credit Assignment mechanisms need to have their behavior (which is controlled by a couple of hyper-parameters) tuned for each new problem. This effect was partially diminished with the use of a simple normalization scheme [Fialho et al., 2009b]. In order to have a controller robust to many different situations, two Credit Assignment schemes based on ranks were lately proposed, namely, the Area-Under-Curve (AUC) and the Sum-of-Ranks (SR) [Fialho et al., 2010c; Fialho et al., 2010b]. Besides being rank-based, both of them, when considering the fitness values instead of the fitness improvements for the ranking, are completely comparison-based, i.e., invariant with respect to monotonous transformations over the original fitness function. In this way, this very important property, which is guaranteed by construction in most of the recent EAs², is maintained when employing these AOS mechanisms.

²See, e.g., an extensive mathematical analysis of the advantages of comparison-based randomized heuristics, presented in [Gelly et al., 2007]; and the Covariance Matrix Adaptation - ES (CMA-ES)
1.3 Organization

1.2.3 Empirical Validation

The different combinations of these proposals to Operator Selection and Credit Assignment gave rise to novel AOS methods. A last contribution of this thesis concerns their empirical validation. In order to do so, some artificial scenarios were proposed, which enable a detailed analysis of the behavior of each AOS method with respect to different situations.

The proposed AOS combinations have been compared among each other and with other baseline techniques on many different scenarios, as presented in Chapter 6: they were tried (i) on the proposed artificially generated scenarios, (ii) on some boolean benchmark problems, (iii) on a comprehensive set of single-objective continuous problems, and (iv) on a set of Boolean Satisfiability (SAT) instances. In the latter case, an aggregation of fitness and diversity, named Compass, was used for as Credit Assignment [Maturana et al., 2009a; Maturana et al., 2010a], considering the fact that in multi-modal problems some diversity should be maintained in order to avoid premature convergence.

The Rank-based Multi-Armed Bandit techniques have shown to be very efficient, while being also very robust to the many different situations in which they were assessed. After all the empirical evidences gathered, they are thus the recommended choices, if one wants to implement AOS within a given algorithm/problem.

It is also important to note that the AOS paradigm is not exclusive to the EC framework. Indeed, any stochastic/local search algorithm that has different options for the exploration of the search space might profit from the proposed methods. Besides, the same paradigm can be directly used at the hyper-heuristics level, i.e., selecting between different heuristics, instead of selecting between different variations of a given heuristic.

1.3 Organization

The remainder of this thesis manuscript is organized as follows. In the first part, a review of the context of this work will be done, starting with a more comprehensive presentation of EAs in Chapter 2, in which more focus will be given to the EA variants used in the experimental section, namely, Genetic Algorithms and Differential Evolution. Then, an overview of the background concerning the research area of parameter setting in EAs and other meta-heuristics will be presented in Chapter 3. In Chapter 4, a more detailed description and bibliographic review of the issues arising in Adaptive Operator Selection will be presented.

In the second part, the contributions of this thesis work will be presented. Chapter 5 will present and describe in detail the proposed AOS techniques. Chapter 6 will describe and analyze the extensive empirical evaluation done on some very diverse scenarios. Finally, Chapter 7 concludes this thesis, summarizing the contributions, and pointing out possible directions for further work.

[Hansen and Ostermeier, 2001], a comparison-based state-of-the-art adaptive method for Evolution Strategies (ESs), that is also invariant with respect to rotations of the search space
Part II

Background Review
Chapter 2
Evolutionary Algorithms

Contents

2.1 Introduction .................................................. 12
2.2 Modus Operandi .................................................. 13
2.3 Components ...................................................... 14
   2.3.1 Problem-dependent Components ..................... 14
   2.3.2 Representation-specific Components .................. 16
   2.3.3 General Components ..................................... 18
2.4 Popular EA Variants ............................................ 21
   2.4.1 Evolution Strategies .................................. 21
   2.4.2 Evolutionary Programming .......................... 22
   2.4.3 Genetic Programming ............................... 22
   2.4.4 Genetic Algorithms ................................. 22
   2.4.5 Differential Evolution ............................. 23
2.5 Application Areas .............................................. 25
2.6 Discussion ..................................................... 26
In this Chapter, we present an overview of EAs, depicting their general behavior and parameters. Besides, some popular EA variants are described, and some examples of application domains are presented.

2.1 Introduction

An important source of inspiration for the development of computational methods to automate problem solving is the “intelligent” way in which biological processes solve complex problems found in nature. Some popular examples of these approaches, referred to as bio-inspired methods, are Neural Networks [Arbib, 2002], which are based on the structure of the biological brain; Fuzzy Logic [Klir and Yuan, 1995], which is inspired on the human way of reasoning; Swarm Intelligence algorithms, which are inspired by living examples of collective social behavior, e.g., the Particle Swarm Optimization (PSO) [Eberhart et al., 2001] and the Ant Colony Optimization (ACO) [Dorigo et al., 1996] methods; and finally, Evolutionary Algorithms (EAs), which are global optimization methods that mimic the Darwinian “survival of the fittest” paradigm in order to solve optimization and search problems.

Since the seminal works on EAs (we refer the reader to the edited book [Fogel, 1998] for a compilation of them), many variants have been independently developed around the world, originally for different domains of application, with the main difference being the representation and the variation operators used. Historically speaking, the pioneer methods were Genetic Algorithms (GAs) [Holland, 1975; Goldberg, 1989], Evolution Strategies (ESs) [Rechenberg, 1972; Schwefel, 1981], Evolutionary Programming (EP) [Fogel et al., 1966; Fogel, 1995] and, more recently, Genetic Programming (GP) [Koza, 1992; Koza, 1994]. Research has been very active on GAs, ES and GP, but EP has been gradually disappearing from the literature, as it can be considered a special case of ES. Besides, other popular techniques have more recently been created, such as the Differential Evolution (DE) [Storn and Price, 1997; Price et al., 2005], and the already mentioned PSO [Eberhart et al., 2001] and ACO [Dorigo et al., 1996] methods. Except for the latter two, which do not follow the evolution paradigm, each of the mentioned techniques will be separately described in Section 2.4, with a more extensive presentation being done for the methods used in the experimental section of this manuscript, namely GAs and DE. The contemporary view on EAs, however, is that, given their continuous development and frequent hybridizations, it is becoming more and more difficult to differ between the historically relevant techniques. This justifies the recent proposal of a “unified view” for EA methods [De Jong, 2006]. Indeed, they all share the same modus operandi, as described in Section 2.2.

EAs have already demonstrated their efficiency on a wide range of optimization problems beyond the reach of standard methods, such as problems involving structured and mixed search spaces, irregular, noisy, rugged or highly constrained fitness functions, etc. A few examples will be surveyed in Section 2.5. This flexibility is mainly due to the fact
that EAs are general meta-heuristics that can be specialized for each problem by means of several parameters. The main parameters will be introduced in Section 2.3.

2.2 Modus Operandi

The different variants of EAs follow the same general outlines, depicted in Figure 2.1. They differ only in a few technical details, as explained in Section 2.4. Generally speaking, the modus operandi of the EAs can be described as follows.

1. A list (population) of candidate solutions (individuals) is initialized, usually representing a random sampling of the search space.

2. Each individual is then evaluated, according to the fitness function, which defines the problem objective: the higher the degree of achievement of a given candidate solution with respect to the problem at hand, the “fitter” it is.

3. If none of the stopping criteria are satisfied (e.g., optimal solution found, or total computational budget spent), go on to the next steps.

4. The first Darwinian natural selection-based process takes place. Individuals are selected as parents to reproduce, usually based on the fitness evaluation, as in nature: stronger/fitter individuals, i.e., better candidate solutions, have higher chances of being selected for reproduction.

5. These selected individuals are then subject to blind variations (blind in the sense that no information about the problem or the consequences of the variation are considered), by the application of stochastic operators, namely crossover (recombination) and mutation operators, generating offspring.
6. The newly generated offspring are then evaluated, according to the same fitness function, which defines the problem.

7. Then comes the second Darwinian process, the replacement or survival selection, that defines which individuals, from both the parental population and the newly generated offspring population, will survive to the next iteration (generation) of the algorithm.

8. From this evolved population, a new generation can start, going back to Step 3.

From this general cycle, it can be seen that the evolution itself happens mainly due to two opposing forces. On one hand, there is the application of mutation and recombination operators, which introduce random variation in the population, consequently performing an exploration of the solutions search space; intuitively, their sole application would lead to a random search. On the other hand, as in the inspiring theory, better candidate solutions (i.e., fitter individuals) have higher chances to be used in the generation of new (hopefully also fitter) solutions, and to survive for the next generations. These Darwinian procedures are the responsible for giving a search direction, leading to the most promising regions of the search space. This process of blind variation + natural selection is then iterated until an optimal solution arises, or another stopping criterion is attained.

Besides being bio-inspired, EAs are thus stochastic algorithms that work by following a kind of generate-and-test (also known as trial-and-error) approach [Eiben and Schoenauer, 2002], in the same spirit as many other meta-heuristics, such as Simulated Annealing [Laarhoven and Aarts, 1987]. While describing the general cycle, several structures and procedures were mentioned; they will be described into more detail in the following.

## 2.3 Components

EAs have mainly three kinds of components: some are related to the problem to be solved (Section 2.3.1), others depend on the representation being used (Section 2.3.2), while others are totally general (Section 2.3.3). Each of these groups will now be briefly described in turn.

### 2.3.1 Problem-dependent Components

The components that need to be defined according to the optimization problem can be described as follows.

**Evaluation/Fitness Function**

The evaluation or fitness function plays the role of the environment in the Darwinian natural selection-like procedures, assigning a score to each individual according to its degree of “achievement” with respect to the optimization problem at hand. The fitness function is thus the core of the algorithm, which needs to be very carefully designed, as it
is often the only source of information about the problem that is available to the algorithm [Eiben and Schoenauer, 2002].

Although EAs are said to be robust with respect to very different situations (e.g., irregular, noisy, and highly rugged fitness landscapes), a minimal level of continuity and/or regularity needs to be provided to guide the search towards the most promising regions of the search space; otherwise it tends to act as a random search, with no direction to follow. In some cases, however, the definition of the fitness function itself is a very complex task; a recent paradigm, referred to as Interactive Evolutionary Computation (IEC), addresses this issue by “outsourcing” the fitness evaluation to humans, as in [Quiroz et al., 2007] for example. In the neighboring local search community, such paradigm is commonly referred to as “Human Guided Search” [Anderson et al., 2000].

The fitness evaluation of a candidate solution is undoubtedly the most computationally expensive step of the EA cycle, and its computational cost affects other user choices, mainly the size of the population and the number of offspring created at each generation, as each generated offspring will require a fitness evaluation. In case the fitness evaluation cost becomes prohibitive for evolution to take place (usually many generations, consequently fitness evaluations, are needed), some approximations between the fitness values found in the neighborhood of the candidate solution under assessment might also be used, as in [Martikainen and Ovaska, 2006]. Besides, in some application fields, the evaluation might also be very noisy, thus requiring the averaging of several independent assessments in order to have a reliable measure of quality.

In cases where there is more than one objective to be optimized, referred to as multi-objective in the literature, special fitness assessments need to be used to take into account all the objectives. The most popular criterion for this is the Pareto optimality (see, e.g., [Deb, 2001; Mueller-Gritschneder et al., 2009]).

**Representation**

From the structural point of view, in order to solve a given problem, the main issue that needs to be defined is how the candidate solutions are going to be represented. The solutions themselves, referred to as phenotypes, might be very complex structures; but their corresponding low-level representation, the genotypes, which are the structures manipulated by the algorithm, are usually much simpler. As in the inspiring theory, genotypes are constituted by genes, which store the values of the candidate solution for each variable of the problem at hand. The most common representation or encoding schemes can be listed as follows.

- **Binary encoding**: Vectors of binary values, or bit-strings, are commonly employed to represent problem solutions that have only two possible values for each variable. For example, in the SAT problem [Cook, 1971], which consists in assigning values to binary variables in order to satisfy a Boolean formula, each gene represents the boolean state of each variable of the problem [Lardeux et al., 2006].

- **Permutation encoding**: Vectors of integers are usually used for sequencing problems. The classical application example is the well-known Traveling Salesman Prob-
Chapter 2. Evolutionary Algorithms

Problem (TSP), in which there is a set of cities that need to be visited by a salesman, and the objective is to find the order of cities that minimizes the distance to be traveled. In this case, each city is assigned an integer number, and the order of these numbers defines the sequence in which the salesman will visit the cities [Merz and Freisleben, 1997].

- **Real-value encoding**: For some problems, the direct use of real values is preferred, as for example, to optimize the weights of a neural network. In this case, each gene of a candidate solution represents the value of each of the corresponding weights of the neural network [Obradovic and Srikumar, 2000].

- **Tree encoding**: It is used mainly to evolve programs or regular expressions, with every solution being encoded as a tree of objects, such as functions or commands of a given programming language. For example, given a set of input and output data samples, it can be used to find a function that maximizes the mapping between them [Koza et al., 2003].

The permutation and the binary encoding schemes are historically used by GAs to solve combinatorial problems, as confirmed by the given examples. The real-value encoding is usually employed by ES and DE on continuous optimization problems; while the tree representation scheme is often used within GP to automatically generate or optimize programs. More details about each of the mentioned EAs will be given in Section 2.4.

### 2.3.2 Representation-specific Components

Some of the components of an EA, namely the initialization procedure and the variation operators, are representation-dependent, *i.e.*, they need to be defined according to the chosen representation model. This is in fact one of the reasons why EAs are successfully applied to so many different domains of application (see Section 2.5 for a few examples): given an appropriate initialization procedure and variation operators, any kind of search space can be tackled [Eiben and Schoenauer, 2002]. Such representation-specific components will be briefly described in the following.

**Initialization**

According to the representation being used, the initial population is usually created after a random sampling of the search space. A uniform sampling is commonly used when the search space is finite and its bounds are known, *e.g.*, in the binary, permutation and tree-based representations. For the real-value representation, the uniform sampling can be used for the initialization if the search space bounds are provided, a Gaussian distribution being used otherwise.

Furthermore, in case some prior knowledge is available, it might be used in the initialization process, *e.g.*, by directly including a known good solution. But, on one hand, such manipulated initialization might result into a wrong bias to the search process, what is clearly much worst than having no bias at all [Eiben and Schoenauer, 2002]. On the other
2.3 Components

hand, this extra effort is usually not very well paid-off as, when starting from a random population, the same EA would typically need just very few generations to achieve the same level of solution quality [Eiben and Smith, 2003].

Variation Operators

Mutation operators are *asexual* variation operators, i.e., a single parent individual is considered to generate an offspring. These operators are responsible for introducing non-existing (or re-introducing missing) characteristics into the population, thus augmenting the so-called genetic diversity. A complementary view for their purpose is that of fine-tuning: individuals might improve their respective qualities after suffering slight variations (e.g., mutation of a single gene). Traditional mutation operators for each of the four popular representations mentioned in Section 2.3.1 can be listed as follows.

- For bit-strings, the bit-flip mutation operator flips each bit with probability $1/\ell$ by default (although a different probability can be employed), $\ell$ being the length of the bit-string. Another popular mutation operator is the $x$-bit operator, which flips $x$ randomly chosen bits each time it is applied.

- For real-valued vectors, the most common mutation operator is the addition of a random value to each vector component or gene. It is mainly used within ES, with the random value being usually extracted from a normal distribution with zero mean and a pre-defined standard deviation (also referred to as the mutation step-size). In the case of DE, several mutation strategies exist, using the differences between two or more vectors (individuals) in different ways for perturbing the vector population.

- For permutation and tree encoding schemes, a popular mutation operator is the order changing: two genes are randomly selected and have their values exchanged. In the case of trees, not just the values of the chosen nodes, but also both sub-trees (or branches) attached to them, are usually switched. A simpler alternative is the exchange of the value found in the chosen gene or node by another value randomly chosen from the finite search space.

Crossover or recombination operators are *sexual* variation operators: parts of the genetic material of two or more different parent individuals are recombined somehow, creating one or more new offspring. Their use is justified by the building blocks assumption: supposedly, the good fitness scores of the parents are related to some portions of their genetic material; with strictly positive probability, the good portions (building blocks) of both parents are recombined, consequently creating a fitter individual. Accordingly, the most common crossover operator for all the representations is the $x$-point crossover, which divides each parent individual into $x$ building blocks, forming the offspring by different recombinations of these portions. A more exploratory variant is the uniform crossover, which uniformly selects which genes are taken from each parent to constitute the new offspring. In the same way than for the mutation operators, there are other ways of doing so as well, according to the representation being used, as follows.
• In the case of real-valued vectors, arithmetical operations might be done between the genes of both parents.

• For tree-like representations, different branches of the parents trees can be exchanged.

It is important to note that the effect of the crossover operators on the search process is automatically adapted, by construction, according to how converged the population is: while there is a good level of diversity, it helps into exploring the search space; the less diversity there is, the more exploitation-like will be its behavior, up to the total inefficiency as, differently from the mutation counterparts, it can not introduce any novelty into the population.

The standard mutation and crossover operators are pure stochastic transformations that receive as input one or more (parent) individuals, and generate as output one or more new (offspring) individuals, not using any feedback about the impact of their application on the search; due to this, their application is usually referred to as a blind variation. Then, it is usually up to the replacement selection mechanism to accept or not the generated offspring, and consequently guide the search process. Differently from that, in some well-known application domains, the available information about the problem might be used into the design of specialized or “intelligent” operators. For example, the flipping of a bit might be prevented by the fact that it is known (by simulating the outcome of its flipping) to be already set to a good value, as done in the SAT domain (see, e.g., the GASAT [Lardeux et al., 2006]); or by intelligently choosing which building blocks should be exchanged, as in most of the specialized crossover operators for the TSP problem (see, e.g., [Chan et al., 2005]). By using trial-and-error and feedback from the search in order to decide its move, what is being done by these operators is in fact what is usually referred to as “local search”. In other cases, existing meta-heuristics are also used as inspiration for local search variation operators within EAs, as in [Branke et al., 2003], which proposes the use of crossover operators based on the Ant Colony Optimization algorithm for the same TSP problem. EAs that employ local search techniques as variation operators are commonly called as Memetic Algorithms (MA) [Krasnogor, 2002].

2.3.3 General Components

Agreeing with the “unified view” of EAs proposed in [De Jong, 2006], some components are general, not being affected by the kind of representation used. These components can be listed as follows:

• the size of the parent population $m$;

• the size of the offspring population $n$;

• the procedure for selecting parents $pselect$;

• the procedure for producing offspring $prod$;

• the procedure for selecting survivors $sselect$;
2.3 Components

- and the stopping criterion $cstop$.

Each of these representation-independent components will be briefly described in the following.

### Parent and Offspring Population Sizes

The parent population size $m$ defines the level of parallel search done by the EA, as it contains the starting points for the new solutions explored in the search space at each generation [De Jong, 2007], while the offspring population size $n$ determines the number of trials done by the algorithm at each generation.

Their definition is mainly related to (i) how rugged the fitness landscape is thought to be, as a bigger population will enable a better parallel exploration of multiple peaks; and to (ii) the available computational budget, as at each generation it is required to evaluate the fitness of all newly generated individuals, which is by far the most expensive step of the evolutionary cycle, as previously discussed.

### Parental or Reproduction Selection

One of the Darwinian representation-independent steps that “guide” the evolution engine, the **parental selection** $pselect$, as its name says, is the procedure responsible for selecting which of the individuals will be chosen for reproduction.

A very simple and popular parental selection method is the **proportional** one: for each individual, the probability of being selected is proportional to its fitness score, and the selection is performed by a roulette wheel-like method over these probabilities. In some application domains, however, an individual might have a fitness value that is orders of magnitude higher than the others. By using the proportional method in such a case, this super-individual will very probably be always selected for reproduction, thus quickly taking over the entire population, consequently leading to (possibly) premature convergence. Oppositely, in case the fitnesses of the individuals have similar values, similar selection probabilities will be assigned to each of them; consequently, there will not be enough selection pressure to guide the search towards the most promising regions.

To avoid such kind of problem related to fitness ranges, other methods widely used nowadays are: (i) the **rank-based** selection, in which the selection probability of an individual is proportional to the ranking of its fitness value with respect to the other individuals in the population; (ii) the **tournament** selection, in which $T$ individuals are randomly chosen from the population, and the best between these $T$ individuals is selected, with $T \in [2, m]$, $m$ being the parent population size; and a different variant, the (iii) **stochastic tournament** selection, in which 2 individuals are randomly chosen, and the best between them is selected with user-defined probability $t \in [0.5, 1]$.

### Survival or Replacement Selection

The other representation-independent procedure that enforces the simulation of the Darwinian natural selection process is the **survival selection**, $sselect$. It defines how the pop-
ulation of the next generation will be constituted, based on the current parental and offspring populations, *i.e.*, which individuals of both populations are going to survive for the next generation.

Broadly speaking, there are two categories of replacement methods: (i) individuals from both populations are considered, thus disputing between each other for survival, as the number of available “places” in the next generation is limited (the main population size \( m \)); or (ii) just the offspring population is considered, and the best \( m \) out of \( n \) individuals are maintained for the next generation. The former is referred to as the “plus” strategy in the Evolution Strategies (ESs) context, and is elitist by default, as the best individuals out of both populations are maintained. The latter is called as the “comma” strategy; if the risk of possibly losing the best solution found can not be assumed, elitism should be externally added in this case (remembering that the maintenance of the best in the population will always create a bias in that direction, what might be good, or not).

In case both parent and offspring populations have the same size \( (m = n) \) and a comma-like survival selection is used, the EA is said to be *generational*, *i.e.*, the entire population is replaced after each generation. When \( m > n \) and \( n = 1 \), with a plus-like survival selection, the algorithm is referred to as being *steady-state*, behaving in a much greedier way.

**Termination Condition**

The termination condition for EAs is commonly related to the available budget, *e.g.*, elapsed time, number of generations or fitness evaluations. However, part of this budget is usually wasted somehow, as follows. As soon as the population *converges*, *i.e.*, most of the individuals are very similar (no much diversity can be found into the population), the search becomes very inefficient, with the best solution being improved just by a lucky move (random sampling = Monte Carlo).

A more intelligent stopping criterion relies on the convergence measure: as soon as it is detected, the search can be stopped, possibly restarting from a new random initial population in order to provide more opportunities for the algorithm to find a better solution within the same available computational budget. There are several ways to account for population convergence, the simplest one is the number of generations since the last time a better solution was found (stagnation). A more complex mechanism that can be found in the GA literature is the “bit-wise average convergence measure” [Goldberg *et al.*, 1995], which estimates, for each gene, the percentage of individuals presenting the same value, with the final measure being the average of the values found on all genes; convergence is detected when this average exceeds some user-defined threshold.

**General Representation of Special Cases**

After the definition of these general/representation-independent components, it becomes possible to easily describe EAs well-known to the community, and also to create new variations [De Jong, 2007]. Besides facilitating the human comprehension, this general description is also beneficial in practice, when implemented into existing toolkits, as in
2.4 Popular EA Variants

Several EA variants exist, the most popular ones will be briefly described in the following, namely, Evolution Strategies (ESs), Evolutionary Programming (EP), Genetic Programming (GP), Genetic Algorithms (GAs) and Differential Evolution (DE). A more detailed description will be done for the latter two, as they are the EAs used in the experimental section of this manuscript. For the sake of correctness, the “historical” view of the techniques will be used in their description; however, as already discussed, given the frequent hybridizations and exchanges between the areas, nowadays it is becoming more and more difficult to differ between them, a “unified view” of EAs being recommended [De Jong, 2006].

2.4.1 Evolution Strategies

Evolution Strategies (ESs) [Rechenberg, 1972; Schwefel, 1981] are very popular EAs. They are mostly applied to numerical (continuous) optimization problems, thus using real-valued vectors to represent the solutions.

The common alternatives for the crossover are the exchange or the linear recombination of components, although historically no crossover is used. Differently, mutation is always applied, being usually a Gaussian noise with zero mean and user-defined standard
deviation (also referred to as mutation step-size). Both comma and plus selection strategies are considered, as well as different sizes for the parental and offspring populations, according to the characteristics of the problem.

It is worth noting that the state-of-the-art continuous optimizer to date is the Covariance Matrix Adaptation - ES (CMA-ES) [Hansen and Ostermeier, 2001], an ES with a very efficient dynamic control of the mutation step-size, shape and direction. The CMA and other schemes for automatically adapting the mutation step-size will be briefly discussed in Section 3.3.4.

2.4.2 Evolutionary Programming

Evolutionary Programming (EP) was originally applied to the evolution of finite state automata for machine learning problems [Fogel et al., 1966], with representation and variation operators being specially designed to this kind of search space. More recently, however, it was adapted to also tackle numerical optimization problems [Fogel, 1995], with just few details differing it from the ESs, e.g., stochastic instead of deterministic replacement. Given the higher popularity of ESs, EP is rarely mentioned in the recent literature; most authors consider it nowadays as a special case of ES.

2.4.3 Genetic Programming

Genetic Programming (GP) [Koza, 1992; Koza, 1994] is known as the EA variant to be used when evolving programs and logical expressions, using trees with varying sizes to represent them. Some Lisp-like languages that naturally embody tree structures are frequently used within GP; although other functional languages can also be adapted somehow to do so.

The evolution engine is very similar to GAs (which will be presented in Section 2.4.4), except for the crossover and mutation operators, which are specially designed in order to be able to cope with the tree representation. As the individuals do not have a fixed size, a very common problem in GP is the so called bloat, the uncontrolled growth of an individual, with a comprehensive body of literature being dedicated to its control (see, e.g., [Luke and Panait, 2006]).

2.4.4 Genetic Algorithms

Genetic Algorithms (GAs) [Holland, 1975; Goldberg, 1989] are by far the most popular methods. Traditionally, they are used to address combinatorial problems, using the very general bit-string representation. Other representations can also be employed to facilitate the translation from phenotype to genotype, consequently making easier the manipulation of the candidate solutions. For example, for sequencing problems such as the Traveling Salesman Problem (TSP), the permutation-based representation is used, in which each gene i corresponds to the object considered as being in the ith position. A very general representation of a GA in the form of a pseudo-algorithm is shown in Algorithm 2.1.

Each new offspring is usually generated as follows. Firstly, the parents are selected according to the parental selection method used, e.g., the tournament selection method.
2.4 Popular EA Variants

Algorithm 2.1: General pseudo-algorithm for a Genetic Algorithm

1: Generate the initial population
2: Evaluate the fitness of all individuals
3: while the stopping condition is not satisfied do
4:   for $i = 1$ to $n$ do
5:      parent1 = ParentalSelection(parentPop)
6:      if rndreal[0,1) < $p_c$ then
7:         repeat
8:            parent2 = ParentalSelection(parentPop)
9:         until parent2 != parent1
10:        offspringPop[i] = Crossover(parent1, parent2)
11:   else
12:      offspringPop[i] = Copy(parent1)
13:   end if
14:   if rndreal[0,1) < $p_m$ then
15:      offspringPop[i] = Mutation(offspringPop[i])
16:   end if
17:   end for
18: Evaluate the fitness of all the generated offspring
19: parentPop = SurvivalSelection(parentPop, offspringPop)
20: end while

A crossover operator is then applied with probability $p_c$ over the two selected parent individuals; the most common crossover operators in this case are the $x$-point and the uniform ones. The resulting offspring (a copy of one of the parents in case crossover is not applied) is then subject to a mutation operator, e.g., a bit-flip or a $x$-bit mutation, with probability $p_m$. Finally, for the survival selection, usually a generational procedure is used, i.e., the entire parental population is replaced by the newly generated offspring population; another popular method for survival selection in GAs is the steady-state one: each time an offspring is generated, it instantaneously replaces one of the individuals of the parental population according to a given criterion, such as fitness or age.

From this description, it can be seen that GAs are very general methods, with too many degrees of freedom with respect to their parameter choices. We refer the reader to [Whitley, 1994; Mitchell, 1998] for a more comprehensive introduction and an extensive theoretical analysis of GAs as known in the early days.

2.4.5 Differential Evolution

Differential Evolution (DE) [Storn and Price, 1997; Price et al., 2005] is a more recent method proposed for global numerical optimization. As in ES, the solutions are represented by vectors of real-values. The pseudo-code of the standard DE algorithm is shown in Algorithm 2.2, where $d$ is the number of decision variables (also referred to as the dimension of the problem). Following the terminology commonly used for this technique, $NP$ refers to the population size, $F$ is the mutation scaling factor, $CR$ is the crossover rate, $x_{i,j}$ is the $j$-th variable of the solution $x_i$, and $u_i$ is the offspring generated after $x_i$.  

23
Algorithm 2.2: The Differential Evolution algorithm with DE/rand/1/bin strategy

1: Generate the initial population
2: Evaluate the fitness for each individual
3: while the stopping condition is not satisfied do
4:   for $i = 1$ to $NP$ do
5:     Select uniform randomly $r_1 \neq r_2 \neq r_3 \neq i$
6:     $j_{rand} = \text{randint}(1, d)$
7:     for $j = 1$ to $d$ do
8:       if $\text{rndreal}_j[0, 1) < CR$ or $j$ is equal to $j_{rand}$ then
9:         $u_{i,j} = x_{r_1,j} + F \cdot (x_{r_2,j} - x_{r_3,j})$
10:       else
11:         $u_{i,j} = x_{i,j}$
12:       end if
13:     end for
14:   end for
15:   for $i = 1$ to $NP$ do
16:     Evaluate the offspring $u_i$
17:     if $f(u_i)$ is better than or equal to $f(x_i)$ then
18:       Replace $x_i$ with $u_i$
19:     end if
20:   end for
21: end while

Although achieving very good performance on a wide range of problems (see, e.g., all the successful applications listed in [Price et al., 2005]), it is a very simple algorithm. To start with, there is no parental selection: each individual in the population is used to generate one offspring; and there is only one (deterministic) method for survival selection: each offspring is compared only with its parent, replacing the parent in case it has a better fitness, sometimes the age of the parent also being considered as a penalty factor.

The generation of an offspring is done by means of mutation and crossover operators. But, differently from the historical convention used in EAs, the mutation takes into account the genetic material of two or more individuals, doing some form of sum of weighted (by the scaling factor $F$) differences between their components (genes); while the crossover considers only the parent and the intermediary solution generated after the application of the mutation operator, usually referred to as the mutant vector.

Many reproduction schemes have been proposed in the literature, using different mutation and/or crossover operators [Price et al., 2005; Storn and Price, 2008]. In order to distinguish among these schemes, the notation “DE/a/b/c” is commonly used, where “a” specifies the base vector to be mutated; “b” is the number of difference vectors used by the mutation strategy; and “c” denotes the crossover scheme, binomial or exponential. As an example, in Algorithm 2.2, the reproduction scheme used is the “DE/rand/1/bin” [Price et al., 2005] (see lines 8-12), i.e., the classical “DE/rand/1” mutation strategy, with the binomial crossover.

Other well-known mutation strategies can be listed as follows:
2.5 Application Areas

1. “DE/best/1”: \( v_i = x_{\text{best}} + F \cdot (x_{r_2} - x_{r_3}) \)
2. “DE/best/2”: \( v_i = x_{\text{best}} + F \cdot (x_{r_2} - x_{r_3}) + F \cdot (x_{r_4} - x_{r_5}) \)
3. “DE/rand/2”: \( v_i = x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) + F \cdot (x_{r_4} - x_{r_5}) \)
4. “DE/current-to-best/1”\(^1\): \( v_i = x_i + F \cdot (x_{\text{best}} - x_i) + F \cdot (x_{r_2} - x_{r_3}) \)

where \( x_i \) is the current individual or parent, \( x_{\text{best}} \) represents the best individual in the current generation, and \( x_{r_1}, \ldots, x_{r_5} \) are different individuals randomly chosen from the current population.

Concerning the crossover operators, the binomial crossover is similar to the uniform crossover used in GAs: for each variable of the problem, the offspring receives with probability \( CR \) the value of the mutant vector, of the parent vector otherwise. The exponential crossover is similar to some extent to the GA two-point crossover: components from the parent vector are used up to the first crossover point, randomly selected from \( \{1, \ldots, d\} \); then \( L \) consecutive components, counted in a circular manner, are copied from the mutant vector, \( L \) being a user-defined parameter; and the rest is taken again from the parent vector [Zaharie, 2009]. Although the exponential crossover was used in the seminal DE publication [Storn and Price, 1995], the binomial crossover is much more frequently used nowadays, being said to be “never worse than exponential” [Storn and Price, 2008].

From this description, it becomes clear that one of the main advantages provided by DE is its simplicity when compared to other EAs. In addition to the scheme to be used for reproduction, only three parameters are left to be defined by the user: the population size \( NP \), the mutation scaling factor \( F \) and the crossover rate \( CR \).

2.5 Application Areas

EAs have been successfully applied to many different application fields, as extensively presented in a recent book completely dedicated to this topic [Yu et al., 2008]. Besides the different “flavors” of optimization, which are by far their most important areas of application, EAs have also been used as a “source of creativity” in many other areas. Some of these very diverse application examples can be briefly described as follows.

**Combinatorial** problems have attracted the attention of EC researchers since the early days of the field, as many important real-world problems can be modeled in this way, what makes it a very profitable area. For example, EAs have been used for diverse **scheduling** problems, such as crew and train scheduling [Semet and Schoenauer, 2006], task **planning** [Bibai et al., 2010], etc. Most of its success, however, comes from its hybridization with local search and Operational Research (OR) heuristics, as exemplified in the problems of TSP [Merz and Freisleben, 1997], university **time-tableing** [Abdullah et al., 2007] and graph-coloring [Porumbel et al., 2010].

In **continuous** optimization problems, EAs have also greatly shown their value. Especially after the advent of the state-of-the-art, almost parameter-less, CMA-ES technique

---

\(^1\)“DE/current-to-best” is also referred to as “DE/target-to-best/” or “DE/local-to-best/” [Price et al., 2005; Das et al., 2009].
[Hansen and Ostermeier, 2001], many researchers have been applying it on their very own problems. A very comprehensive and up-to-date list of applications of CMA-ES to continuous optimization problems can be found in [Hansen, 2009b], incredibly counting up to 120 references as of nowadays: very diverse application fields have already been tackled, such as the optimization of gas turbine combustors [Hansen et al., 2009c], or the search for the best craniofacial super-imposition for forensic identification [Ibanez et al., 2009]

Another domain of increasing attention, specially since the beginning of this decade, is that of multi-objective optimization. By the use of special fitness evaluation and selection methods [Deb, 2001], EAs are known to efficiently find the best set of solutions that satisfies all the objectives under consideration, the so-called Pareto front. An interesting application example in this context is that of [Singh, 2006], in which EAs are used to optimize several criteria for the automatic estimation of seismic velocity, a measure used for the possible discovery of petrol in the underground.

Needless to say, the abilities of (i) handling mixed search spaces, and (ii) having solutions with variable-length (specially true for GP), enable the use of EAs on problems out of reach of standard methods. Besides, as in this way there is no constraint in terms of representation of the solutions, a much more comprehensive (and unbiased) exploration of huge search spaces can be done, possibly leading to the discovery of solutions that could never be imagined by biological intelligence. In this context, several examples of results automatically achieved by EAs that are competitive (and often better) than human performance are presented (and awarded) every year in the so called Humies competition [Koza, 2010], sometimes even resulting in patentable products, as presented in [Koza et al., 2000].

For the same reason, EAs have shown interesting results in terms of creativity in the art and design domains [Bentley and Corne, 2002; Romero and Machado, 2007], with examples ranging from architecture up to music automatically generated by evolution. In such kind of applications, in which the evaluation of the solutions is “subjective” somehow, a human-in-the-loop is often used to make the role of the fitness function, the so-called Interactive Evolutionary Computation (IEC). In [Quiroz et al., 2007], for example, the IEC paradigm is used for the optimization of user interfaces. One of the main problems of this approach is that of “fatigue”: differently from computers, humans do get tired; different proposals have been done in order to reduce such effect, as in [Kamalian et al., 2005].

Lastly, an application domain that is inevitably becoming more and more relevant nowadays is that of sustainable development, where EAs have also shown both their exploration and optimization efficiencies. As combinatorial examples, we can mention the optimization of strategies for pollution prevention [Tan, 2007] and the efficient planning of solid waste management [Yeomans et al., 2003]. It has also been used in the design of “green” buildings, optimizing all the multiple objectives in mixed search spaces that such a project might contain, e.g., the reduction of energy and water consumption, as well as waste and landfill generation [Pitman and King, 2009]. Another example in the same context is the optimization of the topology and parameters of the electronic components that constitute a Heating, Ventilating and Air-Conditioning system in order to achieve better energy efficiency [Angelov et al., 2003].
2.6 Discussion

As reviewed in the previous Section, EAs are very general and robust methods, outperforming other approaches and achieving interesting results on very different application domains. It is always unfair, however, to compare their performance with sophisticated problem-tailored methods on the problems to which the latter were specified to: the strength of such special methods is always related to the exploration of problem-specific knowledge, while EAs are general search methods that treat the problem as a black-box function, using the fitness function as the sole source of information. Consequently, the problem-tailored methods, as their name says, perform very well just on their very own problems, while EAs are able to achieve reasonable performance in a much wider range of problems, what might be preferred depending on the situation.

Anyway, given the mentioned characteristics of EAs, the general recommendations for their use can be summarized into the following cases, based in [Eiben, 2002]:

- The search space is very big: in such a case, the brute-force approach becomes prohibitive, consequently turning a (directed) randomized search into a good alternative.

- The search space is mixed, i.e., the variables of the problem have different types (integer and real for example): as discussed in Section 2.3.2, EAs do not have any restriction with respect to the representation of the candidate solutions, whenever corresponding variation operators and initialization procedure are provided.

- The variables of the problem interact with each other in a complex non-linear way, resulting in an objective function of same nature: it is unusual to have specific sophisticated methods for such cases, as extracting some knowledge from the gradient of the search is not a trivial task, while possibly leading to a highly multi-modal fitness landscape (what links to the following item).

- The search space is multi-modal, i.e., with many local optima: the population-based approach employed by EAs enable the exploration in parallel of several promising regions, consequently augmenting the probability of discovering the global optimum; while standard local search methods would tend to get prematurely trapped into local optima. However, whenever a higher (possibly local) optimum is found, the selection pressure will bias the entire population towards it; to avoid this undesired convergence, the maintenance of some diversity into the population needs to be enforced somehow (see, e.g., the niching methods [Horn, 1997]).

- The optimization problem is dynamic, changing over time: the evolution process follows the direction being currently given by the fitness function, no matter if it is static or dynamic, automatically adapting to eventual changes in a transparent way.

- The evaluations are noisy: the evolution is not guided by the evaluation of a single point, but rather by the “trend” gathered from the evaluation of the many points considered in the current population, significantly reducing the noise effect. Besides,
in such cases, several re-evaluations might also be performed until a higher confidence is achieved – indeed, this is a quite common approach, not exclusive to EAs.

It is worth noting, however, that the use of EAs and other stochastic meta-heuristics does not guarantee to find the truly optimal solution for the problem at hand. Accordingly, the fact that more computational time possibly means the discovery of better solutions explains why one of the main drawbacks for the use of EAs is their high computational cost. But, whatever the available budget, a solution as good as possible is always available at the end; such important property is commonly referred to as anytime behavior [Eiben and Smith, 2003].

Another drawback that prevents their broader use is that of parameter setting: the performance of an EA is directly related to how efficiently it explores the search space, while also being able to exploit the most promising regions. The so-called Exploration versus Exploitation (EvE) balance is controlled by several parameters, that usually need to be defined by the user according to the problem or class of problems at hand. A more comprehensive discussion on parameter setting will be presented in Chapter 3, while Chapter 4 will focus on the parameter setting sub-problem that is addressed by the contributions proposed in this thesis, the Adaptive Operator Selection.
Chapter 3

Parameter Setting in EAs

Contents

3.1 Introduction .................................................. 30

3.2 Parameters Influence and Possible Settings ............... 31
  3.2.1 Parent and Offspring Population Sizes ............... 31
  3.2.2 Selection Procedures .................................. 32
  3.2.3 Offspring Production ................................... 32
  3.2.4 Stopping Criterion ...................................... 33
  3.2.5 Representation ........................................... 34

3.3 Classification of Parameter Setting Techniques ............ 34
  3.3.1 Which parameter is changed? .......................... 34
  3.3.2 How the changes are made? ............................ 35
  3.3.3 Which evidences guide the changes? .................. 40
  3.3.4 Which is the scope of the change? .................... 40

3.4 Discussion .................................................. 41
In this Chapter, we present a survey about parameter setting in EAs. The possible influence of the setting of some parameters in the EA performance is discussed, together with some proposals for their automatic setting found in the literature. A well-known classification of methods for parameter setting in EAs is also described.

3.1 Introduction

In order to efficiently apply an EA to a given problem, there are several parameters that need to be defined by the user, as surveyed in Chapter 2. In the early days of research in the area, such parameters were seen as an advantage for the EAs, enabling their “personalization” according to the characteristics of the problem at hand.

The optimal values for these parameters were usually defined by intuition, based on rules-of-thumb well known to the community. At the same time, it was believed that researchers would be able to find problem-independent (or universal) “winner” settings, i.e., parameters values that would provide efficient performance to the EAs, independently of the application field. In the context of GAs, two very popular (although very different) “universal settings”, published in [Grefenstette, 1986; De Jong and Spears, 1990], are compared in Table 3.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>[De Jong and Spears, 1990]</th>
<th>[Grefenstette, 1986]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Number of generations</td>
<td>1000</td>
<td>not specified</td>
</tr>
<tr>
<td>Crossover type</td>
<td>(typically) 2-point</td>
<td>(typically) 2-point</td>
</tr>
<tr>
<td>Crossover rate</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Mutation types</td>
<td>bit-flip</td>
<td>bit-flip</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.001</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3.1: Two examples of static sets of parameters for GAs

However, many further works were published in the following, presenting very different settings for particular problems, consequently putting into question the scientific relevance of the mentioned works. The No Free Lunch (NFL) theorem [Wolpert and Macready, 1997] put an end to this quest for an universal setting, by stating that, roughly, there is no “best algorithm” that solves all problems better than any other. This might be seen as the unique very well-accepted contribution brought by the establishment of the NFL theorem; indeed, it is hardly considered in practice, and several extensions and contradictions have been published since then (see, e.g., [Auger and Teytaud, 2010; Christensen and Oppacher, 2001; Whitley and Watson, 2005]).

Accordingly, the contemporary view of EAs acknowledges that specific (sometimes classes of) problems require specific setups for satisfactory performance [Eiben et al., 2007]. This is the main reason to the fact that EAs are very rarely used
3.2 Parameters Influence and Possible Settings

outside the “evolutionary research labs”. Although being tempted by the several empirical demonstrations of the efficiency of EAs on many difficult problems out of reach of other optimization methods, scientists from other domains very often fail in getting interesting results, mainly because of the lack of general methods for tuning at least some of the involved parameters, and also because they are not, and do not want to become, “Evolutionary Engineers”.

From these observations, parameter setting in EAs appears actually as a major issue that has deserved much attention during recent years [Eiben et al., 1999], and research is still very active nowadays, as witnessed by a complete edited book that has been recently published [Lobo et al., 2007], and by the numerous recent references cited in this document. These contributions, proposed to automate or guide somehow the definition of the parameter values, intend to make the EAs to Cross the Chasm [Moore, 1991], enabling the whole scientific community to benefit from their use without their main burden, that of parameter setting.

The following of this Chapter summarizes the research in the field of parameter setting within EAs. Firstly, an overview of the influence of some of the main parameters on the performance of the EAs, and some examples of what has already been done to automate their setting, are presented in Section 3.2. Then, in Section 3.3, a well-accepted classification of the different ways of doing parameter setting within EAs is described. The Chapter is concluded in Section 3.4, with some further discussions about parameter setting in EAs.

3.2 Parameters Influence and Possible Settings

Although having been conceived for different purposes and presenting different behavior, the EAs have some common structures and parameters that are independent from the representation being used, as presented in Section 2.3.3. Since the parameter setting techniques presented in the following of this document are not meant to work with only one kind of EA, we will briefly discuss here about the influence of these common representation-independent parameters on the search process.

All these parameters affect somehow the Exploration versus Exploitation (EvE) balance: intuitively, as discussed throughout the text, the EA should explore the search space in the early stages of evolution, gradually migrating to a more focused exploitation of the promising regions. The objective of this Section is not to describe these parameters, as it was already done in Section 2.3.3, but rather to extend a bit on their influence on the EvE balance, and consequently on the performance of EAs, based on [De Jong, 2007]. Besides, some possibilities of parameter setting are also presented, including a brief bibliographic review for each of them.

3.2.1 Parent and Offspring Population Sizes

A small parent population size $m$ does not enable a good exploration of the (usually very big) search space, possibly converging prematurely to a local optimum; while a bigger $m$ provides a higher probability of finding a global optimum, as multiple peaks might be
simultaneously explored. Having a bigger population, however, might slow down the con-
vergence by the fact that more evaluations are needed at each generation. A compromise
between both needs to be found.

Following the EvE balance intuition, ideally, the population should be big in the be-
ginning, enabling a better exploration of the search space, with its size decreasing (thus
focusing on the most promising regions) as the search goes on. However, the dynamic
adaptation of the population size was found to be a difficult task [De Jong, 2007], due
to several interacting factors, such as selection pressure [Eiben et al., 2006], noisy fitness
landscapes [Goldberg et al., 1992], the fact that generations overlap (“plus” replacement)
or not (“comma”/generational replacement) [Schwefel, 1995], etc [Arabas et al., 1994;
Smith, 1993; Eiben et al., 2004; Bäck et al., 2000].

Still from the EvE balance point-of-view, the parent population represents which re-
gions of the search space are being currently explored, while the ratio between its size
$m$ and the offspring population size $n$ defines the amount of exploration done by it at
each generation. The ideas and methods proposed for the adaptation of the parent pop-
ulation size are also valid for the size of the offspring population; we refer the reader
to [Jansen et al., 2005] for a comprehensive analysis and some specific proposals for the
dynamic adaptation of this parameter.

### 3.2.2 Selection Procedures

As for the other parameters, the migration from exploration to exploitation is also related
to the level of selection pressure that is exerted, i.e., less selection results into more explo-
ratinon, and vice-versa. In the case of the standard tournament selection, for example, as
described in Section 2.3.3, the smaller the tournament size $T$ (i.e., the less individuals are
chosen to participate in the tournament), the more random the selection is; oppositely, the
higher is $T$, the higher is the chance of considering the best individuals, consequently the
more elitist is the selection process. Along the same lines, the probability $t$ of selecting the
best individual between the two chosen individuals in the stochastic tournament selection
can range from a totally random selection ($t = 0.5$) to a completely elitist strategy ($t = 1$).

However, the control of such selection pressure is not ruled simply by the setting of
these parameters. Indeed, it is defined by the combined effects of both parent and replace-
ment selection procedures, not mentioning other interacting effects, such as the population
size [Eiben et al., 2006]. This complexity might be the reason why so few references can be
found on the dynamic adaptation of the selection pressure. In [Herrera and Lozano, 1998],
a fuzzy model is used to automatically control the tournament size, based on the geno-
typic and phenotypic diversity measures; more recent works propose [Eiben et al., 2006]
and better empirically validate [Vajda et al., 2008] a hybrid self-adaptive tournament size,
which achieves much better results than the fuzzy model. It is referred to as hybrid by the
fact that the parameter control is done by a combination of self-adaptation and feedback-
based or adaptive control (see Section 3.3.2 for a brief overview of the different ways of
doing parameter control). After the proof-of-principle presented on these latter references,
the use of parameter control for selection methods has shown to be a viable path to be
further explored towards more efficient and easier-to-tune EAs [Vajda et al., 2008].
3.2 Parameters Influence and Possible Settings

3.2.3 Offspring Production

In the procedure for offspring production, \textit{prod}, the variation operators need to be defined according to the representation being used, in order to be able to generate feasible solutions. The application of these operators directly impact the EvE balance, consequently affecting the effects provided by all the previously mentioned parameters [De Jong, 2007], as follows: while the selection pressure tends to reduce the population diversity, variation operators are responsible for counter-balancing this effect by, as their name says, introducing variation into the population. The quantity of novelty to be possibly introduced, however, depends on the population size and on the level of diversity in the current population.

Such correlation between the parameters makes it very complex to decide which operators should be included in the EA algorithmic framework for a given problem, and how to set their sub-parameters. In addition to this possible correlation, there is the stochastic nature of the underlying algorithm. These issues make it very difficult to predict a priori how a given operator (with a given configuration) will behave during the search process. Besides, different operators (or different configurations of the same operator) might perform differently at different stages of the search, according to the characteristics of the region of the fitness landscape currently being explored by the population. See, for instance, how the performance of each operator varies on the simple OneMax problem, presented in Section 6.4.2.

An alternative is to take this decision out of the user’s burden by automatically adapting the internal parameters of a given operator. In the context of ESs, for example, the variance of the Gaussian mutation operator has been automatically adapted since the early days, starting with the “1/5th success rule” [Schwefel, 1975], until the advent of the very popular and current state-of-the-art CMA-ES [Hansen and Ostermeier, 2001]. This latter is, indeed, the most (and almost unique) successful case of a parameter control technique within EAs to date, after more than 30 years of research in the area [De Jong, 2007].

Another plausible approach is to maintain a collection of operators, and to dynamically select the ones that are affecting the search process in a more beneficial way [De Jong, 2007]. The selection of which operator among the several available operators should be used, what we here refer to as Adaptive Operator Selection (AOS), is representation-independent. Accordingly, the AOS methods proposed in this thesis can be applied to any of the existing or newly proposed EAs\footnote{Indeed, the proposed AOS techniques can also be extended to other local search heuristics, but this discussion is out of the scope of the current Section.} – as a representative set, in Chapter 6 we show their use within GAs and DE. Following the EvE intuition, ideally, the dynamic selection of operators should promote the use of the more exploratory operators in the beginning, preferring the less disruptive ones (exploitation) in the later stages of the search. An extensive bibliographic review on this, which is the central topic of this thesis, will be presented in Chapter 4.
3.2.4 Stopping Criterion

The stopping conditions do not directly affect the EvE balance, although possibly affecting the setting of other parameters. This is a clear example of a representation-independent component that could be defined in a more autonomous way, although depending on a representation-based criterion, the population diversity. Anyway, works proposing the dynamic adaptation of this parameter were not found in the literature. Fixed strategies are commonly used, being defined after an expensive off-line tuning phase (see Section 3.3.2) or, more frequently, via intuition and/or based on the budget constraints.

3.2.5 Representation

One of the main choices is very probably how to represent (and consequently manipulate) the candidate solutions of a given problem. Although greatly affecting the performance of EAs, the representation is very often defined a priori, guided by a large body of literature [De Jong, 2007]. Such definition is often static, with very few works considering its dynamic adaptation during the search process.

The effects of the adaptation of the representation can be said to be two-fold. On the one hand, it can be used to improve the effectiveness of operators, by adapting the representation according to the characteristics of the operator. For instance, in the Messy-GAs [Goldberg et al., 1991], the position of the genes on the chromosome are constantly modified while solving the problem, in order to maintain the 1-point crossover operator at a good level of performance throughout the search process. On the other hand, it can also be used to bring (or contribute into maintaining) invariance properties to the EAs, as in the recent Adaptive Encoding approach [Hansen, 2008]. Based on the CMA-ES, this method provides to any continuous search algorithm the invariance property with respect to rotation over a given problem function.

3.3 Classification of Parameter Setting Techniques

Very different parameter setting methods have already shown their usefulness in the literature by automatically setting representation-independent and also algorithm-specific parameters of EAs. A classification of these techniques, proposed in [Eiben et al., 1999], and later revised in [Eiben et al., 2007], is very well-accepted by the community, as acknowledged by the number of citations it received. Since it is used to classify the methods proposed in this work, it will be reminded in the following, for the sake of self-containedness.

It categorizes the parameter setting methods according to four aspects, listed as follows: (i) Which parameter is changed? (ii) How the changes are made? (iii) Which evidences guide the changes? (iv) And which is the scope of the change? The two former aspects are general, while the two latter regard only the on-line parameter control methods. Each of these aspects will be briefly discussed in the following.
3.3 Classification of Parameter Setting Techniques

3.3.1 Which parameter is changed?

The first criterion adopted for the classification concerns which component or parameter of the EA is being changed. Although there is no standard list of parameters, we consider here the parameters described in Section 3.2.

As already mentioned, each of the listed parameters might also have some sub-parameters, e.g., the number of bits to be flipped by the bit-flip mutation operator, the tournament size for the tournament selection, etc. These sub-parameters are neglected here; the objective of this classification is rather to be able to easily locate, within the standard EA loop, which steps are affected (hopefully improved) by the proposed changes.

The Adaptive Operator Selection techniques proposed in this work provide to the user an autonomous control of the use of the available variation operators. This can be seen as an adaptation of their application rates (despite the fact that the proposed bandit-based techniques, presented in Chapter 5, do not rely on probabilities for the operator selection).

3.3.2 How the changes are made?

The changes in the parameter values can be made, mainly, in two different ways, as illustrated in Figure 3.1: before the main run of the algorithm on the given problem, referred to as off-line or external parameter tuning; or during the run, while solving the problem, referred to as on-line or internal parameter control. A brief description about each of them will be presented in the following.

![Figure 3.1: Classification of parameter setting methods, from [Eiben et al., 1999].](image)

**Off-line or External Parameter Tuning**

Methods that perform off-line or external tuning determine a priori the appropriate parameter values, based on the results of several runs of the given algorithm. The algorithm to be tuned is usually considered as a black box, with the tuning method guiding the exploration of the search space of the parameter values. Off-line tuning methods can be further sub-divided into two main classes: pure statistical methods, and optimization methods, which treat the parameter tuning as an optimization problem itself.
Starting with the statistical methods, the most basic (and computationally expensive) way of doing so lies in the execution of a complete Design of Experiments (DoE) process, which is also referred to as a full factorial design, or even as a brute-force approach: the range of possible values for the parameter under consideration are discretized into \( m \) candidate configurations, each of them is independently assessed \( n \) times, and the best configuration is extracted according to some ANOVA-like statistical test over this \( m \times n \) performance data. In practice, it becomes very computationally expensive to tune even just a few parameters. For instance, by considering only 4 parameters, each parameter with 5 possible values, it will already lead to \( 5^4 = 625 \) candidate configurations to try out (not considering possible cross-influences between parameters).

The Racing techniques [Birattari et al., 2002; Yuan and Gallagher, 2004] do basically the same, but in a much less time-consuming way, as follows. As in DoE, the parameter values are also discretized into \( m \) candidate configurations. But, as soon as a candidate configuration is statistically found to be significantly worse than the current best configuration (after some runs, depending on the variance of the achieved results), there is no need to keep further assessing it; this configuration is thus eliminated from the tuning process. In this way, the computational resources are more efficiently used, focusing just on the most promising candidate configurations, consequently leading to lower variance performance estimates for them. The use of this approach results into important time savings, as illustrated in Figure 3.2. The x-axis \( \Theta \) represents the number of remaining candidate configurations, and the y-axis “\( i \)” shows the number of evaluations or “racing laps” done for each of them; the amount of computation needed for both, the F-Race [Birattari et al., 2002] and the brute-force approaches, are represented by the areas of their respective surfaces.

![Figure 3.2: A visual representation comparing the amount of computation needed by the brute-force approach (dashed rectangle) and the F-Race method (shadowed area), reproduced from [Birattari et al., 2002].](image)

A prominent example of Racing techniques is the F-Race [Birattari et al., 2002], which uses the “Friedman’s two-way analysis of variance by ranks” as statistical test to eliminate candidate configurations. This is the method used to tune all the hyper-parameters of the proposed and baseline AOS techniques for the empirical comparisons that will be presented in Chapter 6.

Although saving a significant amount of computational budget, the use of the F-Race
3.3 Classification of Parameter Setting Techniques

technique can become computationally prohibitive whenever there is a large number of
parameters and a wide range of possible values for each parameter, as some initial runs
need to be done for each candidate configuration before the first elimination round. A
simple alternative proposed to this problem is the use of a sampling of the whole set of
configurations [Balaprakash et al., 2007]. In case a priori knowledge about the configura-
tion search space is available, it can be used to define the probabilities of sampling each
configuration; however, as this is usually not the case (and remembering that a priori
information might also include a wrong bias in the search), the authors propose the use
of a completely random sampling of the configurations. The resulting method is referred
to as Random Sampling Design F-Race (RSD/F-Race) [Balaprakash et al., 2007].

A different kind of approach for the parameter tuning problem, as previously men-
tioned, is to consider it as an optimization problem on its own: by varying the pa-
rameter values, the objective might be to optimize some measure such as the perfor-
manace of the algorithm over a given problem or class of problems, or its robustness
with respect to several problems, etc. Based on this, it becomes straightforward to
think about the use of optimization methods for this task, thus at a higher level of
abstraction, commonly referred to as the “meta” or “hyper” level. EAs themselves
have already been used to do so, defining the so-called Meta-EAs [Clune et al., 2005;
Yuan and Gallagher, 2007]. The problem in this case lies in how to define the parameters
of the EA in the meta-level.

The ParamILS [Hutter et al., 2009] method uses an iterated local search algorithm to
explore the neighborhood of the best parameter values found so far, using some random
perturbations and restarting the search from time to time (according to a user-defined
probability) to enforce a better coverage of the search space. Its very general idea, com-
bined with an adaptive limit of the time spent for evaluating individual configurations,
enables it to be used on very different situations. Indeed, it was already shown to efficiently
tune algorithms with up to $10^{37}$ possible configurations [Hutter et al., 2009].

Along the same lines, another popular optimization method used to off-line tuning,
the Sequential Parameter Optimization (SPO) [Bartz-Beielstein et al., 2005], combines
modern statistical approaches for deterministic algorithms, as the Design and Analysis
of Computer Experiments (DACE), with classical regression techniques, in order to tune
stochastic algorithms such as EAs. The set of candidate configurations being assessed is
constantly refined during the tuning procedure, what is done by means of Gaussian pro-
cesses, with some configurations being eliminated and new ones being inserted in the pool
according to the current model of the parameter space. At a higher level of abstraction,
rather than simply a tuning method, SPO can be seen as a methodology for the empir-
ical analysis of stochastic optimization algorithms, providing to the experimenter a very
well-defined twelve-step procedure.

Another model-based optimization method applied to parameter tuning is the Iterated
F-Race (I/F-Race) [Balaprakash et al., 2007; Birattari et al., 2009], which is yet another
improved variant of the F-Race. Starting from the initial set of possible parameter values
for each parameter (as done in the original F-Race), at each iteration, some efficient
configurations are used to update a probabilistic model about the configurations search
space. This model is then used to generate new candidates, consequently biasing the search
Chapter 3. Parameter Setting in EAs

towards the most promising parameter configurations. The I/F-Race is more complex, but much more efficient than the RSD/F-Race approach.

A very different approach is implemented by the Relevance Estimation and Value Calibration (REVAC) method [Nannen and Eiben, 2007]. It uses Shannon and differential entropy in order to find parameters with higher impact on the performance of the algorithm, while also estimating the utility of their possible values. Thus, besides tuning the parameters of the algorithm, it provides to the user a high-level information about their relevance, which can in turn be used in order to better allocate the resources for their calibration, e.g., by providing more resources for the tuning of the most important or sensitive parameters.

All these methods have already proved their efficiency and usefulness in different ways in the literature. An advantage provided by them is that, as they use only generic performance measures, they are not limited to EAs, being possibly applied to many other stochastic algorithms, while being also very easily combined. In [Smit and Eiben, 2009], for instance, an extensive empirical comparison between different pure and hybrid off-line tuning methods is presented, including meta-EAs, REVAC, SPO and Racing. However, given the stochastic nature of EAs, each performance assessment corresponds in fact to the average of a few evolutionary runs, and this makes the off-line tuning a very expensive procedure. Furthermore, static settings are usually provided by these methods (the parameter value is fixed along the run), whereas the optimal setting likely depends on the local landscape being explored by the population (see, e.g., [Eiben et al., 2007, p.21] and references therein).

On-line or Internal Parameter Control

Internal parameter control methods work directly on the values of the parameters while solving the problem, i.e., on-line. Such kind of mechanisms for modifying parameters during an algorithm execution were invented early in EC history, and most of them are still being investigated nowadays. Indeed, there are at least two strong arguments to support the idea of changing the parameters during an EA run:

• As evolution proceeds, more information about the algorithm behavior within the current fitness landscape is known, so it should be possible to take advantage of it. This applies to global (e.g., knowing how rugged is the landscape) and to local properties (e.g., knowing whether a solution has been improved lately or not).

• As the algorithm proceeds from a global (early) exploration of the landscape to a more focused, exploitation-like behavior, the parameters should be adjusted to take care of this new reality. This is quite obvious, and it has been empirically and theoretically demonstrated that different values of parameters might be optimal at different stages of the search process (see [Eiben et al., 2007] and references therein).

The different approaches that have been proposed to internally adapt the parameters can be gathered into three categories, depending on the type of information used for the adjustment of the parameters values, as presented in Figure 3.1. Each category will be
briefly reminded in the following, including some examples in the context of adaptation of variation operators.

**Deterministic** parameter control methods implement a set of deterministic rules without any feedback from the search. This is, in general, hard to achieve, because of a simple reason: they rely heavily on knowing beforehand how long the EA will take to achieve a given target solution with the running algorithm, what can not be easily predicted. But even if it were, the way to balance exploration and exploitation can hardly be guessed. This situation is worsened by two facts: first, given the stochastic nature of EAs, there is usually a big variance between different runs on the very same problem; and second, these methods often require additional parameters that are used to tune the deterministic parameter itself (starting with the total number of generations the algorithm will run), and even though these parameters can be considered second-order, their influence is nevertheless critical. Given these difficulties, these methods were mainly used in the early days of research in the area, as in [Hesser and Männer, 1990], in which a theoretically optimal schedule was proposed to deterministically adapt the mutation application rate, based on the elapsed number of generations.

Since our knowledge about the way the search should behave is always limited, it is sometimes possible, and advantageous, to let evolution itself tune some of the parameters. This kind of parameter control approach, referred to as **Self-Adaptive**, adjusts parameters “for free”, *i.e.*, without any direct specification of the user. In other words, individuals in the population might contain “regulatory genes” that control some of the parameters, *e.g.*, the mutation and recombination rates; and these regulatory genes would be subject to the same evolutionary processes as the rest of the genome [De Jong, 2007]. During quite some time in the 90s, self-adaptation was considered as the royal road to success in EC. First of all, the idea that the parameters are adapted for free is very appealing, and the parallel with self-regulated genes is another suggestive argument. On the practical side, as self-adaptive methods require little knowledge about the problem and, what is probably more important, about the way the search should proceed, it sometimes remains as the only way to go when nothing is actually known about the problem at hand. As an early example, in [Bäck, 1992], the representation of each individual is extended by 20 additional bits, which are used to encode its own self-adapted mutation rate, a real value between 0 and 0.5. In [Spears, 1995], a single bit is used to represent which of two crossover operators (uniform or 2-points) should be applied, resulting in a much better performance than the one achieved by the static use of a single operator. A very recent and comprehensive review of the current state of research on self-adaptation methods, with special hints for its use within combinatorial problems, can be found in [Smith, 2008]; another review, more focused on continuous optimization, can be found in [Kramer, 2010]. Although having shown to be an efficient approach in many different situations, the main drawback is that the algorithm needs to explore, in parallel, the search space of the variables of the problem and also the search space of the parameter values, what potentially increases the complexity of the search.
Then, it becomes clear that whenever some decisions can be made to help the search following an efficient path, this should be done. **Adaptive** or **Feedback-based** methods follow this rationale, being the most successful approaches nowadays in on-line parameter control. These methods are based on the monitoring of particular properties of the search/optimization process, and use changes in these properties as an input signal to modify the parameter values. The most prominent example of adaptive methods, and one of the main successful stories of automatic parameter control within EAs, is that of CMA-ES [Hansen and Ostermeier, 2001]. In this method, information about the gradient and about the trajectory of the search are used to automatically adapt the step-size and the shape of the ES mutation operator. Lots of achievements have been reported also by the Operational Research and Local Search communities, in which the adaptive methods are referred to as “Reactive Search”; a recent survey on this can be found in [Battiti et al., 2008]. The main contributions of this work, the Adaptive Operator Selection (AOS) methods proposed in Chapter 5, are included in this latter category.

### 3.3.3 Which evidences guide the changes?

When using adaptive parameter control techniques, the parameter values are adapted based on the monitoring of some measures of the progress of the search. A further criterion commonly used to classify these techniques is the kind of evidences which are used to guide the changes done while solving the problem [Smith, 1998]. Using as example the AOS techniques, this feedback from the search progress can be provided in two different ways.

The most common **Credit Assignment** scheme for AOS considers the real values of the fitness improvements achieved by the application of each operator. Starting from the common Instantaneous and Average values, up to the use of Extreme values supported by us (see Section 5.2.2), all of them reward the operator (and consequently guide the changes in the operators preferences) based on raw values, which are **absolute** evidences.

Differently, the **Credit Assignment** schemes that use ranks over the raw values of the fitness improvements (see Section 5.2.4) control the choice of operators based on the ranking of the same fitness improvements. Thus, it is not the magnitude of the improvement brought by the application of the operator that matters, but rather how good it is with respect to the others, what is referred to as **relative** evidence.

### 3.3.4 Which is the scope of the change?

Yet another aspect used to classify parameter control techniques lies in the scope of the adaptation being done. According to [Angeline, 1995], the adaptation might happen at the level of the **population**, the **individual**, or the **component**. This factor is not only related to the parameter control method itself, but also to the parameter that is being adapted, as acknowledged in [Eiben et al., 2007].

At the **population** level there are the methods that adapt **global** parameters, *i.e.*, parameters affecting the whole population. A very early example of adaptive method at the population level is the so-called “1/5th success rule” [Schwefel, 1975]: if the applica-
3.4 Discussion

tions of the mutation operator succeed in generating offsprings that are fitter than their respective parents in more than 1/5 of the trials, the mutation step-size should be increased by an user-defined fixed ratio, decreased otherwise. The recent state-of-the-art CMA-ES [Hansen and Ostermeier, 2001], which adapts the step-size and the “shape” (defined by a Hessian matrix) of the mutation operator based on the results of its latest applications, is another example of an adaptive method affecting the whole population. The AOS techniques presented in this thesis also act at the population level, adapting the operators application rates that are globally used for the generation of every new candidate solution. Rather than adapting somehow the operators application step, in [Eiben and van Hemert, 1999], the SAW method is proposed to globally adapt the fitness function for constraint satisfaction problems. This is done by dynamically changing the weight of each gene (representing a constraint) on the fitness function, with the harder constraints affecting more highly the fitness evaluation, consequently resulting in a higher reward for the creation of individuals that succeed in satisfying them.

The methods at the individual level control parameters that locally affect each individual. As an example, the self-adaptive methods that encode the (GAs) operator application rates [Spears, 1995] or the (ESs) mutation step-size [Beyer, 1995], within the genotype of each individual, affect only the candidate solution to which they are attached to. A recent example of adaptive method that affects each solution individually is the Multi-Objective (MO) CMA-ES [Igel et al., 2007]: briefly, the same adaptation implemented by the original CMA-ES is used to adapt the mutation operator carried by each individual, whenever its application is successful, i.e., whenever it succeeds in generating a fitter offspring.

At the lowest level of abstraction considered here, there are also methods that adapt parameters within an individual, but at the so-called component level. Exemplifying, in [Schwefel, 1995], a self-adaptive ES was proposed, in which each element of the real-valued vector representation has a variance parameter attached to it; thus, each gene of the individual is mutated according to its own self-adapted variance parameter.

The hyper level could also be considered here, as recommended in [Maturana, 2009], aggregating the recent methods that control the usage of several heuristics in different ways. This kind of approach is commonly referred to as hyper-heuristics [Burke et al., 2010].

3.4 Discussion

As discussed throughout this Chapter, the performance of the EAs is directly related to how the Exploration versus Exploitation (EvE) balance is addressed by the algorithm: if too much exploration is done, the search will very probably take too long to achieve the optimum; while if too much exploitation is done, the search is very likely to prematurely converge to a local optimum. To achieve an acceptable performance, a compromise between both terms needs to be found, what is controlled by some of the EA parameters, as described in Section 3.2.

After the No Free Lunch theorem [Wolpert and Macready, 1997], it is acknowledged that there is no algorithm that performs best over all optimization problems. Considering
two instances of the same EA with different parameter settings as two different algorithms, it states thus that there is no winner universal parameter setting, i.e., specific problems require specific parameter settings. These findings lead to a kind of dilemma: in order to manually setup the parameters of the EA according to the problem at hand, the user would need to analyze and understand the characteristics of the problem; while one of the main reasons for the use of EAs and other meta-heuristics is, indeed, the lack of knowledge about the problem. Avoiding such dilemma, by automating the task of parameter setting, is thus one of the main motivations for research on this topic.

After the analysis of the influence of the parameters in the performance of EAs, however, it can be said that almost all of them affect the EvE balance somehow, with one parameter counter-balancing or intensifying the effects of the other. This interaction makes more complex the task of automatic parameter setting, being one of the main reasons why just very few works try to address more than one parameter at the same time.

Parameter setting in EAs can be done mainly in two different ways, as described in Section 3.3. Off-line parameter tuning methods consider the EA as a black-box, using only the performance of the algorithm (usually averaged over several runs given its stochastic nature) in order to choose the best set of parameters, which are usually “statically” used during the whole search process; while, intuitively, the EvE balance of the algorithm should be continuously modified while solving the problem, gradually switching between exploration and exploitation according to the progress of the search. This on-line dynamic adaptation is what is provided by the so-called parameter control methods, category in which the contributions proposed in this work are included. It is important to note that, even in the case of static problems, they are dynamic from the point of view of operator selection. Needless to say, an even higher payoff might be achieved in the cases in which the problem itself is dynamic, with its fitness landscape changing over time [De Jong, 2007].

The methods proposed for the automatic parameter setting, however, present their own parameters that also need to be defined by the user, referred to as hyper-parameters in the following of this text. Although it might seem not so interesting to replace some parameters by others, the hyper-parameters are at a higher level of abstraction, being thus (hopefully) more easily “understood” by the user and less sensitive than the original EA parameters with respect to their tuning. For example, in the case of the Adaptive Operator Selection techniques proposed in this work, described in Chapter 5, two or three hyper-parameters (depending on the method) need to be configured; while in the original EA framework the user would need to define a very complex and problem-specific scheduler in order to have the same kind of adaptive behavior. These hyper-parameters can then be tuned by off-line tuning methods, as done for the experimental comparison presented in Chapter 6; or extra layers of parameter control could be added, what is always worth whenever the assumptions about the easier comprehension and smaller sensitivity of the higher-level parameters with respect to the lower-level ones are held.

Although not being part of the scope of this thesis, another viable path for the parameter setting in EAs would be to try to build a knowledge base correlating somehow the parameters of the problem instances solved (the so-called problem descriptors) with the respective parameters used by the algorithm to achieve good performance. In this way, after logging data from a “sufficient” amount of instances and parameters, whenever a new
instance that needs to be solved is recognized as part of a certain class of problems, the parameter values that were already optimized to a previously seen instance of the same class can be re-used, thus no need of further tuning it. Such kind of approach has been successfully applied in the domain of SAT problems (see, e.g., [Hutter and Hamadi, 2005]). To the best of our knowledge, however, there does not exist yet a well-established set of descriptors for the kind of instances commonly tackled by EAs, being thus a possible path for further research.
Chapter 4

Adaptive Operator Selection

Contents

4.1 Introduction .................................................. 44
4.2 Adaptive Operator Selection .......................... 45
4.3 Credit Assignment ........................................ 46
   4.3.1 How to measure the Impact? ....................... 46
   4.3.2 How to assign Credit? ............................. 47
   4.3.3 Whom to assign Credit to? ....................... 48
   4.3.4 Compass: Aggregating Fitness and Diversity .... 48
4.4 Operator Selection ......................................... 50
   4.4.1 Probability Matching ............................... 50
   4.4.2 Adaptive Pursuit .................................. 52
4.5 Some Adaptive Operator Selection Combinations ...... 53
   4.5.1 Fitness-based Approaches ......................... 54
   4.5.2 Diversity-based Approaches ...................... 56
   4.5.3 Fuzzy-based Approaches .......................... 57
   4.5.4 Other Approaches ................................ 58
   4.5.5 AOS within Other Evolutionary Algorithms ....... 59
4.6 Discussion .................................................. 59

45
Chapter 4. Adaptive Operator Selection

In this Chapter, we focus on the parameter setting problem addressed by our contributions, referred to as Adaptive Operator Selection (AOS). The components needed to do AOS, namely, the Operator Selection and the Credit Assignment, are described, and some examples found in the literature are surveyed.

4.1 Introduction

In order to efficiently apply an EA to a given problem, there are commonly two design choices that need to be taken by the user concerning variation operators: (i) which of the existing operators should be used by the evolutionary scheme for the generation of new solutions, and (ii) at which rate each of the chosen operators should be applied. As discussed in Section 2.3.2, there are different kinds of operators for each representation scheme, namely mutation and crossover operators (not mentioning the problem-specific and/or the local search operators). Each one of them has its own characteristics, affecting the Exploration versus Exploitation (EvE) balance of the search process in its own manner, as also briefly discussed in Section 3.2.3. This scenario makes these operator-related choices very sensitive and complex, as follows.

First of all, the performance of a given operator usually depends on the characteristics of the problem being solved. Since it is very difficult to foresee how well a given operator will perform on the problem at hand, the natural choice in this sense would be to use an off-line tuning technique, such as the ones surveyed in Section 3.3.2, in order to find out which operator(s) should be used and how. Although being computationally expensive, these off-line methods usually succeed in providing to the user the best static strategy, consisting of one or a few operators that are applied at fixed rates during the whole search process.

The performance of the operators, however, does not solely depend on the global characteristics of the problem, but also on the local characteristics of the region of the search space that is being currently explored by the population, which can be more adapted or not to the characteristics of the operator. Finally, their performance also depends on the state of the search process, i.e., if it is approaching or not the optimum, how diverse the population is, etc. For example, following the very basic intuition of the EvE balance, more exploratory operators might achieve better performance in the early stages of the search, while more exploitation-like/fine-tuning operators might bring better improvements to the search when it is getting closer to the optimum. These issues are empirically confirmed by the results that will be depicted in Chapter 6: in most of the problems tackled, there is no single operator that is the best during all the search process. Based on this and on the stochastic nature of the underlying algorithms, the static strategies provided by off-line tuning methods tend to perform sub-optimally. Even if the search problem being tackled is static, the operator selection problem is dynamic; so, ideally, the choice of the best operator to be applied should be continuously adapted while solving the problem, i.e., in an on-line fashion.

46
4.2 Adaptive Operator Selection

On-line parameter setting methods are commonly referred to as Parameter Control [Eiben et al., 2007]. There are different ways of dynamically doing so, namely, in a self-adaptive or in an adaptive manner, as reviewed in Section 3.3.2. Self-adaptive methods have the advantage of tuning the parameters “for free”, by the evolution itself, adapting the best operator according to the region being “locally” explored by each individual solution; however, besides augmenting the overall complexity of the problem to be solved by aggregating the solutions search space with the parameters search space, these methods are intrinsically linked to the EA structure. Oppositely, adaptive methods might be more complex to implement, while presenting a few hyper-parameters that also need to be tuned; but they consider the problem search space as it is. Since the adaptation of the parameters is usually guided by general assessments of search progress, the adaptive methods can be easily extended to other meta-heuristics and/or stochastic local search algorithms.

4.2 Adaptive Operator Selection

Based on all the above arguments, we have decided to tackle the operator selection problem with adaptive parameter control methods, thus aiming at the on-line selection of the best operator, i.e., while solving the problem. We refer to such paradigm as Adaptive Operator Selection (AOS). Figure 4.1 depicts a high-level view of how AOS methods can be integrated within an EA, which can be read in a general way as follows.

1. For the generation of each new trial solution (or after $n$ trials or generations), the EA asks the AOS which of the available operators should be applied.
2. The AOS returns the operator to be used, according to its Operator Selection mechanism, which selects one operator based on the recent performance of all operators, usually represented by an estimate of their empirical qualities.
3. The selected operator is applied by the EA, a new solution is generated, consequently impacting somehow the search, e.g., generating an offspring better than its parent (fitness improvement), varying the mean diversity of the population, etc.

Figure 4.1: The Adaptive Operator Selection general scheme.
4. This impact assessment is transformed into a credit (also referred to as reward), according to the implemented Credit Assignment scheme.

5. This credit or reward is then used to update the empirical quality (or performance) estimates kept for each operator by the Operator Selection scheme, which will be used the next time it needs to select one of the operators.

6. This loop happens continuously while solving the problem, in an on-the-fly reinforcement learning fashion.

As can be seen from this description, designing an AOS method requires the definition of two components: (i) the Credit Assignment scheme, that assigns credit to an operator based on the impact brought by its recent application on the current search/optimization process; and (ii) the Operator Selection mechanism, which selects the next operator to be applied, based on the knowledge built by the stream of these empirical assessments. In the following, these components will be separately discussed, and some existing methods will be surveyed.

4.3 Credit Assignment

Several Credit Assignment mechanisms have been proposed in the literature, following Davis’ seminal paper [Davis, 1989]. They differ mainly in three aspects: (i) how the impact of the operator application should be measured; (ii) how to assign credit based on these impact assessments; and (iii) to which operator(s) the credit should be assigned to. Each one of these aspects will be briefly detailed and exemplified, respectively, in Sections 4.3.1 to 4.3.3.

Finally, although the most common impact measure is the fitness improvement, diversity becomes important as well when tackling multi-modal problems. The Compass [Maturana and Saubion, 2008a], which is a method to aggregate both measures, is used in our experimental section, within a GA applied to SAT problems (see Section 6.5). For the sake of self-containedness, this method will be described in Section 4.3.4.

4.3.1 How to measure the Impact?

In order to measure the impact of the application of an operator on the search process, most approaches consider the improvement achieved by the fitness of the generated offspring with respect to a reference value. This reference might be a local value (e.g., the fitness of its parents [Tuson and Ross, 1998; Wong et al., 2003; Ho et al., 1999]), or a global/population-based value (such as the fitness of the current best individual [Davis, 1989; Lobo and Goldberg, 1997], or the median or some other quantile fitness [Julstrom, 1995; Julstrom, 1997]). In [Barbosa and Sá, 2000], an aggregation of both local fitness improvement and global improvement (with respect to the 90% quantile of the current fitness distribution) is used to assess the productivity of the operators.

Instead of directly using the raw values of the fitness improvements to assess the impact, other approaches measure a relative value. For instance, in [Giger et al., 2007],
the improvement of the offspring with respect to the parent is divided by the gap between its fitness and the fitness of the current best operator (in the case of minimization; the inverse otherwise).

Other authors consider a much simpler version of impact measure: the fact that the operator application was successful or not. A successful application means that the generated offspring has a better fitness than its reference value. In [Niehaus and Banzhaf, 2001], the success over the parents is considered; while in [Luchian and Gheorghies, 2003], the success over the best, the success over the parents, plateau walks (same fitness than its parents) and worsenings (fitness lower than its parents), are all aggregated in order to accurately characterize the impact of an operator application. Although not being explicitly mentioned, in [Julstrom, 1995; Julstrom, 1997] only the measure of success is used, resulting in a 1 whenever an improved offspring is generated, 0 otherwise.

Different measures, such as the rank of the offspring within the current population, or the age of the solution in number of generations (in this case the adaptation happening every $n$ generations) can also be found in [Whitacre et al., 2006]. In most approaches, when there is no improvement, the offspring is simply discarded or, most commonly, the operator application is evaluated as having a null impact. This latter choice is the one employed by all AOS techniques developed in this work, unless stated otherwise.

In the case of multi-modal optimization, another relevant impact measure concerns the population diversity; a minimal level of diversity should be enforced in order to avoid premature convergence. To measure diversity, the Hamming or the Euclidean distances are commonly used. In [Giger et al., 2007], the relative fitness improvements and the mean Euclidean distance are independently used, depending on the needs of the search with respect to exploitation or exploration. Along the same lines, [Maturana and Saubion, 2008a] proposed an impact measure called Compass, defined as a weighted sum of fitness improvement and mean diversity (Hamming distance) variation. In [Maturana et al., 2010b], two different aggregation methods considering both impact measures were proposed, based directly on the Pareto Dominance paradigm.

### 4.3.2 How to assign Credit?

Based on the impact measures received, at some point a credit needs to be assigned to the operators, in order to update the empirical quality estimates that summarize their performance. This credit can be the instantaneous value, i.e., the impact measure received after its most recent application; but, given the stochastic nature of the underlying algorithm, this tends to be very unstable and noisy. This is often remedied by an aggregation of credits in the Operator Selection side, as done in [Lobo and Goldberg, 1997; Barbosa and Sá, 2000].

A more robust, and by far the most common approach is to use as credit the average of the latest $W$ applications of each operator, $W$ being the size of the sliding time window. The impact measures of the operator are hence aggregated over a given time period, as done in [Davis, 1989; Julstrom, 1995; Julstrom, 1997; Ho et al., 1999; Wong et al., 2003; Giger et al., 2007; Maturana and Saubion, 2008a; Maturana et al., 2010b]. In case the impact measure being used is the success, i.e., 0 or 1 depending if it succeeded in generating
a fitter offspring or not, the average is usually used, being simply referred to as the success rate [Niehaus and Banzhaf, 2001; Luchian and Gheorghies, 2003] of the operator. Though the instantaneous version can be viewed as an average over a window of size 1, both will be distinguished in the remainder of this text, termed respectively Instantaneous and Average Credit Assignment schemes.

A very different approach is the one proposed in [Whitacre et al., 2006], which assigns credit to the operators based on their ability to generate outlier solutions, following some statistics over the received impact measures. The underlying idea is that the generation of rare but highly beneficial improvements matters as much as, or even more than frequent small improvements. A simpler adaptation of this proposal was introduced into our AOS framework [Da Costa et al., 2008; Fialho et al., 2008; Fialho et al., 2009b], and will be considered here too: the credit is set to the maximum fitness improvement over a sliding time-window of size W.

For all these approaches, in case the mentioned statistics are done over the raw values of the received impact measures, the AOS methods tend to have a problem-dependent behavior, as different problems have different fitness distributions (what alters the range of the fitness improvements received), while also presenting different levels of modality (what also affects the magnitude of the diversity measures). In order to reduce such effect, a normalization over the raw methods can be used, e.g., the credit received by the given operator divided by the highest most recent credit received by all operators. Another yet more robust approach is to discard the raw values, considering their ranks instead. Both normalization and rank-based approaches, which are part of the contributions to AOS proposed in this thesis, will be described in detail and analyzed in Chapter 5.

4.3.3 Whom to assign Credit to?

Another independent issue that has been addressed in different ways in the literature is the choice of the operators that should be credited after the generation of a given offspring. It is unquestionable that the operator used to generate the offspring should be credited; but some authors consider that the operators used to generate its ancestors should also receive a share of its credit, somehow claiming that the generated offspring is as good as it is not only because of its parents and the current operator, but also because of how good were its ancestors and the operators used to generate them.

This is usually done following a kind of bucket brigade algorithm, the credit being assigned with a decay factor for each level of ancestry [Davis, 1989; Julstrom, 1995; Julstrom, 1997]. No clear indication, however, about the benefits of this approach can be found in the literature to the best of our knowledge. In [Barbosa and Sá, 2000], for example, the use of ancestors (up to 2 levels) was beneficial in some of the continuous benchmark functions considered, while resulting in worse results on other functions.

Hence, the methods developed during this thesis do not consider ancestry for the Credit Assignment: only the operator that has been applied to generate the given offspring is rewarded.
4.3 Credit Assignment

4.3.4 Compass: Aggregating Fitness and Diversity

Besides the fitness improvements, the diversity variation can also be considered for the Credit Assignment, specially when tackling multi-modal problems, in order to reward a possible tentative of escaping a local optimum. In an empirical analysis of the AOS schemes within a GA applied to SAT problems (see Section 6.5), we have explored Compass [Maturana and Saubion, 2008a], a method that assigns credit to the operators based on an aggregation of both fitness and diversity measures, which works as follows.

A steady-state scheme is used, i.e., the offspring generated after an operator application is instantaneously included into the main population, replacing another individual. Based on this, every time an operator is applied, three impact measures are gathered: (i) population mean diversity variation (\(\Delta D\)), calculated by means of Hamming distance, (ii) mean fitness or quality variation (\(\Delta Q\)), and (iii) execution time \(T\), as shown in Figure 4.2.a. The execution time becomes essential when dealing with complex operators, such as the local search ones used in the Compass original work [Maturana and Saubion, 2008a].

![Figure 4.2: Compass credit assignment](image)

Figure 4.2: Compass credit assignment: Sliding windows of three measures are maintained (a). Average measures of \(\Delta D\) and \(\Delta Q\) are plotted and distance of those points are measured according to a plane with a slope of \(\Theta\) (b). Finally, those distances are divided by the execution time, and the outcome is the credit to be assigned to the operator.

Originally, the average of these values over the last \(\tau\) applications of each operator is displayed in a “diversity versus fitness” plot (black dots in Figure 4.2.b, each dot representing one operator). A user-defined hyper-parameter \(\Theta\) defines the trade-off between the exploitation (fitness) and the exploration (diversity) criteria, consequently tuning the Exploration versus Exploitation (EvE) balance of the operators selection. In practice, such angle defines the plane according to which perpendicular distances from the dots are measured. Finally, the credit assigned to an operator is this measured perpendicular distance (between the dot representing its performance and the plane defined by \(\Theta\)), divided by the execution time (Figure 4.2.c). A complete representation of the Compass Credit Assignment technique in the form of a pseudo-algorithm is presented in Algorithm 4.1.

In the Compass original paper [Maturana and Saubion, 2008a], it is combined with the Probability Matching Operator Selection scheme (see Section 4.4.1); the resulting AOS combination is applied to SAT problems, selecting between 6 evolutionary and local search operators. Later on, we established a collaboration with them, in order to combine their sophisticated Credit Assignment scheme with our state-of-the-art (by that time) Operator Selection mechanism, the Dynamic Multi-Armed Bandit (DMAB), which will
Chapter 4. Adaptive Operator Selection

Algorithm 4.1: Credit Assignment: Compass \((K, \Theta)\)

1: \(D_{op} \leftarrow \frac{\text{Average}(\text{diversity}_{op})}{\max_{i=1...K} |\text{Average}(\text{diversity}_i)|} \)  // mean normalized by max
2: \(Q_{op} \leftarrow \frac{\text{Average}(\text{quality}_{op})}{\max_{i=1...K} |\text{Average}(\text{quality}_i)|} \)
3: \(V_{op} \leftarrow (D_{op}, Q_{op}) \)  // vector representing \(op\) in the plot
4: \(\alpha_{op} \leftarrow |\arctan(Q_{op}/D_{op}) - \Theta| \)  // angle between vector\(_{op}\) and plane defined by \(\Theta\)
5: \(\text{return } \frac{|V_{op}| \cos(\alpha_{op}) - \min_{i=1...K}|V_i| \cos(\alpha_i)|}{\text{Average(exectime}_{op})} \)  // distance to plane divided by time

be described in Section 5.3.2. A summary of the results achieved by this efficient AOS combination, applied to the same SAT problems, was published in [Maturana et al., 2009a; Maturana et al., 2010a]. These empirical results will be revisited in Chapter 6; other examples of schemes using the diversity to calculate the credit to be assigned to an operator after its application will be recalled in Section 4.5.2.

4.4 Operator Selection

Based on the credits received from the Credit Assignment mechanism after one or more operator applications, most Operator Selection schemes maintain an up-to-date empirical quality estimate for each operator, and use it to update their application rates. These probabilities are then used by the underlying algorithm to select the operator to be applied the next time it needs to generate an offspring, what is usually done by means of a roulette wheel-like process, as in the Probability Matching (PM) [Goldberg, 1990] and Adaptive Pursuit (AP) [Thierens, 2005; Thierens, 2007] methods. Both methods will be detailed in this Section.

Another possibility to Operator Selection will be introduced in this work: it is based on the so-called Multi-Armed Bandit framework [Auer et al., 2002], and uses directly the empirical quality estimate gathered by each operator together with an explorative term to deterministically choose amongst the different available operators. This approach is the basis of all Operator Selection schemes developed during this thesis work, which will be exhaustively described in Chapter 5.

4.4.1 Probability Matching

Because of its simplicity and reasonable performance, the most widely used to date Operator Selection scheme for AOS is the Probability Matching (PM) [Goldberg, 1990]. Although possibly presenting some very slight variations, and sometimes not being explicitly mentioned, PM is used in [Davis, 1989; Julstrom, 1995; Julstrom, 1997; Lobo and Goldberg, 1997; Barbosa and Sá, 2000; Niehaus and Banzhaf, 2001; Luchian and Gheorghies, 2003; Wong et al., 2003; Whitacre et al., 2006; Maturana and Saubion, 2008a], to mention a few. Its basic idea
4.4 Operator Selection

is that the probability of selecting a given operator is updated proportionally to its known empirical quality with respect to the others.

This can be mathematically formalized as follows. Let $K$ denote the number of available variation operators. PM maintains a probability vector $(p_{i,t})_{i=1,K}$ and an empirical quality estimate for each operator $j$ noted $\hat{q}_{j,t}$. At each time $t$:

1. The $j$-th operator is selected with probability $p_{j,t}$, via a roulette-wheel selection scheme.

2. The selected operator is applied, and a credit $r_{j,t}$ is computed after the Credit Assignment method at hand;

3. The empirical quality estimate $\hat{q}_{j,t}$ of the $j$-th operator is then updated to account for this credit received; this is done using an additive relaxation mechanism with adaptation rate $\alpha$ ($0 < \alpha \leq 1$, the memory span decreases as $\alpha$ increases):

   $$\hat{q}_{j,t+1} = (1 - \alpha) \hat{q}_{j,t} + \alpha \cdot r_{j,t} \quad (4.1)$$

4. And finally, the probabilities of application of each operator, $(p_{i,t})_{i=1,K}$, are updated to be proportional to the their respective empirical quality estimates, $(\hat{q}_{i,t})_{i=1,K}$:

   $$p_{i,t} = \frac{\hat{q}_{i,t}}{\sum_{l=1}^{K} \hat{q}_{l,t}} \quad (4.2)$$

   By updating the operators probabilities in this way, an operator that performs very badly during a long period of the search will have its application probability decreased to a very low value, or even zero. Such a situation should be avoided, as it would prevent the AOS from using this same operator in case it becomes efficient in a later stage of the search process. For this reason, a minimal selection probability $p_{min}$ is usually enforced. The update rule is then re-defined as follows:

   $$p_{i,t+1} = p_{min} + (1 - K \cdot p_{min}) \frac{\hat{q}_{i,t+1}}{\sum_{l=1}^{K} \hat{q}_{l,t+1}} \quad (4.3)$$

A complete representation of the PM Operator Selection technique in the form of a pseudo-algorithm is presented in Algorithm 4.2.

Discussion: After Equation 4.3, any ineffective operator (not getting any reward) would have at least a probability $p_{min}$ of being selected. The best operator (getting maximal rewards during some time) would be selected with probability $p_{max} = (1 - (K - 1) \cdot p_{min})$. In practice, however, all mildly relevant operators keep being selected, and this hinders the performance of PM (all the more so as the number of operators increases), as pointed out in [Thierens, 2005].

53
Chapter 4. Adaptive Operator Selection

Algorithm 4.2: Operator Selection: Probability Matching \((K, p_{min}, \alpha)\)

1: for \(i = 1\) to \(K\) do
2: \(p_i \leftarrow 1.0/K\) \hspace{1cm} // selection probability
3: \(\hat{q}_i \leftarrow 1.0\) \hspace{1cm} // empirical quality estimate
4: end for
5: while NotTerminated do
6: if one or more operators not applied yet then
7: \(op \leftarrow\) uniformly selected between the operators not applied
8: else
9: \(op \leftarrow\) ProportionalSelectOperator\((p)\) \hspace{1cm} // roulette-wheel
10: end if
11: Operator \(op\) is applied, impacting the search progress somehow
12: \(r_{op} \leftarrow\) CreditAssignment.GetReward\((op)\)
13: \(\hat{q}_{op} \leftarrow (1 - \alpha) \cdot \hat{q}_{op} + \alpha \cdot r_{op}\) \hspace{1cm} // relaxation update rule
14: for \(i = 1\) to \(K\) do
15: \(p_{i,t+1} \leftarrow p_{min} + (1 - K \cdot p_{min}) \left(\frac{\hat{q}_i}{\sum_{i=1}^{K} q_i}\right)\) \hspace{1cm} // proportional probability update
16: end for
17: end while

4.4.2 Adaptive Pursuit

Originally proposed for learning automata [Thathachar and Sastry, 1985], the Adaptive Pursuit (AP) method was ported to the AOS context [Thierens, 2005] in order to address the above shortcoming of PM. The first three out of the four steps describing PM in Section 4.4.1 are shared by AP: the operators are selected using a roulette-wheel process over their probabilities; and after receiving the credit from the operator application, the same relaxation rule is used to update the empirical quality estimates of the operators, as defined in Equation 4.1. The difference is that, in AP, instead of updating the probabilities proportionally to these estimates (see Equation 4.3), a winner-takes-all strategy is employed to push forward very quickly the application probability of the current best operator, noted \(i_t^*\), while consequently decreasing the others, as follows:

\[
\begin{align*}
    i_t^* &= \arg \max_{i=1 \ldots K} \{ \hat{q}_{i,t} \} \\
    p_{i,t+1} &= \begin{cases} 
        p_{i,t} + \beta (p_{\max} - p_{i,t}) & \text{if } i = i_t^* \\
        p_{i,t} + \beta (p_{\min} - p_{i,t}) & \text{otherwise}
    \end{cases}
\end{align*}
\]  \hspace{1cm} (4.4)

where \(\beta \in [0, 1]\) is the learning rate controlling the greediness of the winner-takes-all strategy. The two other hyper-parameters of AP are the same as the ones used in PM: \(p_{min}\), that enforces a minimal level of operators exploration, and the adaptation rate \(\alpha\), which controls the memory span of the Operator Selection scheme. A complete representation of the AP technique in the form of a pseudo-algorithm is presented in Algorithm 4.3.

To show the gain brought by the winner-takes-all strategy, in [Thierens, 2005], PM and AP were compared under the light of an artificially generated scenario, choosing between
4.4 Operator Selection

**Algorithm 4.3: Operator Selection: Adaptive Pursuit \((K, p_{\text{min}}, \alpha, \beta)\)**

1: \(p_{\text{max}} \leftarrow 1 - (K - 1) \cdot p_{\text{min}}\)
2: \(\text{for } i = 1 \text{ to } K \text{ do}\)
3: \(p_i \leftarrow 1.0/K\) // selection probability
4: \(\hat{q}_i \leftarrow 1.0\) // empirical quality estimate
5: \(\text{end for}\)
6: \(\text{while NotTerminated do}\)
7: \(\text{if one or more operators not applied yet then}\)
8: \(op \leftarrow \text{uniformly selected between the operators not applied}\)
9: \(\text{else}\)
10: \(op \leftarrow \text{ProportionalSelectOperator}(p)\) // roulette-wheel
11: \(\text{end if}\)
12: \(\text{Operator } op \text{ is applied, impacting the search progress somehow}\)
13: \(r_{op} \leftarrow \text{CreditAssignment.GetReward}(op)\)
14: \(\hat{q}_{op} \leftarrow (1 - \alpha) \cdot \hat{q}_{op} + \alpha \cdot r_{op}\) // relaxation update rule
15: \(op^* \leftarrow \arg \max_{l=1...K} (\hat{q}_l)\)
16: \(\text{for } i = 1 \text{ to } K \text{ do}\)
17: \(\text{if } i = op^* \text{ then}\)
18: \(p_{op^*} \leftarrow (1 - \beta) \cdot p_{op^*} + \beta \cdot p_{\text{max}}\) // winner-takes-all probability update
19: \(\text{else}\)
20: \(p_i \leftarrow (1 - \beta) \cdot p_i + \beta \cdot p_{\text{min}}\)
21: \(\text{end if}\)
22: \(\text{end for}\)
23: \(\text{end while}\)

5 different artificial operators whose reward distributions were modified every \(\Delta T\) steps. This artificial scenario, referred to as the Uniform scenario, was also used in our empirical comparisons, and it will be described into more detail in Section 5.4.1.

**Discussion:** Although AP showed a performance much superior to PM in the mentioned artificial scenario, both methods still suffer from two main drawbacks. Firstly, \(p_{\text{min}}\) defines a minimal level of exploration that is kept fixed during all the search process. Ideally, the surer the Operator Selection scheme is about one operator being the best one, the less exploration should be done by it, up to no exploration at all as far as the operator found to be the best remains sufficiently good. A second issue refers to another hyper-parameter, the adaptation rate \(\alpha\): it is also fixed during all the search process, and this means that the received credit always has the same fixed weight in the update of the empirical quality estimates of the operators. But, in case there is a long time one operator has not been applied, the assigned credit should have a higher weight, in order to quickly make its empirical quality estimate as up-to-date as possible. Conversely, in the case of an operator frequently applied, the reward weight should be smaller in order to avoid drastically affecting its already well-established performance estimate. These issues were part of the main motivations for the proposal of the Dynamic Multi-Armed Bandit (DMAB) and
Chapter 4. Adaptive Operator Selection

the Sliding Multi-Armed Bandit (SLMAB) Operator Selection techniques, which will be further discussed, respectively, in Sections 5.3.2 and 5.3.3.

4.5 Some Adaptive Operator Selection Combinations

After separately analyzing the AOS components, namely the Credit Assignment and the Operator Selection schemes, this Section will survey different approaches found in the literature for the AOS as a whole, although most of the cited works were already partially described in the two previous Sections.

The methods discussed here are divided into 5 categories: Section 4.5.1 reviews methods solely based on the fitness value and its derivations; Section 4.5.2 surveys methods that also consider diversity, on its own, or aggregated with fitness; and, due to the number of papers found, methods that use Fuzzy Logic for their operator control are overviewed in Section 4.5.3 for the sake of completeness, although this kind of approach will not be addressed in this thesis. Finally, other approaches that do not match any of the mentioned criteria are surveyed in Section 4.5.4, while Section 4.5.5 gives some examples of the use of AOS within EAs other than GA.

4.5.1 Fitness-based Approaches

The seminal AOS method, to the best of our knowledge, was proposed in [Davis, 1989]. Davis’ method updates the probability of each operator according to how often its application helped improving the best fitness in the population. A complex decay mechanism is employed to assign credit to the operators that generated the ancestors of the new-born best individual, up to a pre-defined number of generations. Possibly due to the high computational complexity for that time, this technique was not assessed on-line, it was rather used to obtain a non-adaptive time-varying schedule (i.e., a deterministic parameter control scheme, as described in Section 3.3.2) for later use [Tuson and Ross, 1998], which showed to perform better than a GA with fixed operator probabilities.

A similar but much simpler method was proposed in [Julstrom, 1995] and further assessed in [Julstrom, 1997], referred to as Adaptive Operator Probabilities (ADOPP). The most significant differences with respect to [Davis, 1989] are: (i) instead of the best fitness, the median and the 90% quantile of the current fitness distribution are independently tried as reference values for the measure of the fitness improvement; (ii) the rewarding is not based on the raw value of the fitness improvement, but rather on the success rate (1 in case of improvement, 0 otherwise); and (iii) the decay mechanism for the credit assignment to the ancestors is simply done as \((\text{decay}^{\text{ancestry level}} \times \text{credit})\). In [Julstrom, 1995], the ADOPP seems to show good results on a bi-dimensional continuous problem, and on the Traveling Salesman Problem (TSP), although no comparisons with other techniques are presented. In [Julstrom, 1997], ADOPP is compared with a static strategy (probabilities for each operator fixed at plausible values) on the rectilinear Steiner problem, not being able to achieve better results than it. Note that in both cases, as well as in [Davis, 1989], in addition to the complex bucket brigade-like Credit Assignment scheme, the Operator
4.5 Some Adaptive Operator Selection Combinations

Selection mechanism used is somehow similar to the PM method, though it is not clearly mentioned.

In [Lobo and Goldberg, 1997], the PM method is used again. An operator is credited whenever improvements over the current best solution are achieved. Better performance is shown with respect to several (static) baseline techniques on the OneMax problem. The ancestors are not considered by the rewarding scheme, putting into question the use of this complex and expensive (in terms of memory) procedure. In [Barbosa and Sá, 2000], a similar method is tried on the continuous domain: the main difference lies in the aggregation of two fitness improvements, one with respect to the parents, and another in relation to the 90% quantile of the fitness values found in the current population. The use of ancestors up to two levels is tried, credited in the same way as done in [Julstrom, 1997], but no clear evidences are reported to support its use when applied on a set of continuous benchmark problems – indeed, in some cases it even degrades the performance of the underlying algorithm.

A very different approach is proposed in [Hatta et al., 1997]: the crossover operator to be applied is chosen according to the elite degrees of the individuals selected to be parents. Based on the assumption that an individual that has a large number of recent ancestors with a high fitness value also tends to have a high fitness value, the elite degree of an individual is basically the ratio of the sum of all its “elite ancestors” up to a pre-defined level, divided by the total number of ancestors considered. An individual or ancestor is considered to be an elite member if its fitness is higher than \((\mu + \alpha \times \sigma)\), where \(\mu\) is the average fitness of the current population, \(\sigma\) the respective standard deviation, and \(\alpha\) a user-defined hyper-parameter, referred to as the elite decision factor. Based on this engineered Credit Assignment scheme, the Operator Selection is deterministically performed as follows: in case the sum of the elite degrees of both parents is higher than another user-defined threshold, a less disruptive operator is applied (the 2-point crossover in this case) in order to try to maintain some of the good building blocks; a disruptive crossover is applied otherwise (the uniform crossover). This work is extended in [Hatta et al., 2001], in which some mutation operators are also considered, and a much more complex scheme is devised to measure the elite degree as a continuous value, instead of the original discrete one. In both works, better results are achieved with respect to the GA applying each operator independently, and to the uniform selection between the available operators. Besides, in [Hatta et al., 2001], the scheme implementing the continuous elite degree is shown to improve over the original discrete elite degree, assessed on the NK landscape, the TSP problem, and on a set of continuous benchmark problems.

The Cost Operator Based Rate Adaptation (COBRA) method, devised in [Tuson and Ross, 1998], uses as Credit Assignment the average fitness improvements achieved over the parents, divided by the computational cost of evaluating an offspring. No ancestors are considered. The Operator Selection is simply done as follows: prior to the experiments, the user defines a set of static probabilities; then, at every adaptation cycle, these probabilities are deterministically assigned to the operators, according to their ranking with respect to the perceived performance measures, the top-ranked operators receiving the highest probabilities. On the Credit Assignment side, it is not clear which is the influence of the computational cost, as the evaluation of an offspring is supposed to
have a constant cost, no matter the operator used to generate it. Furthermore, on the 
**Operator Selection** side, no guidelines are provided on how to define a priori the static set 
of probabilities. Evaluated on the OneMax, Royal Road, Long $k$-Path and on a deceptive 
problem, indeed (and not surprisingly), the performance of the COBRA method was found 
to be dependent on the quality of this user-defined set of probabilities.

A different method, the Probabilistic Rule-driven Adaptive Model (PRAM), proposed 
in [Ho et al., 1999], uses a sequence of learning/production phases to adapt the opera-
tors application rates. During the learning phase, operators are uniformly selected, and 
their performances are estimated based on the fitness improvement of the offspring with 
respect to its parent. On the following production phase, operators are selected by the 
PM method, according to the empirical knowledge gathered in the first period. The 
PRAM method achieves better results than a fixed strategy and a self-adaptive scheme. 
In [Wong et al., 2003], the PRAM method is used in combination with an external mech-
anism for diversity maintenance, which gives a higher survival probability to individuals 
located in sparsely populated regions of the search space. The resulting method, referred 
to as APAGAIN, consistently achieves better solution quality than several other static evolu-
tionary schemes within the same computational budget, on a set of continuous benchmark 
problems. However, as pointed out in [Maturana and Saubion, 2008a], around 25% of 
the generations are devoted to the learning phase, in order to try to accurately follow the 
changes in the operators performances during the search process; this might severely harm 
the population and the progress of the search in case disruptive operators are considered.

The Integrated-Adaptive GA (IAGA) [Luchian and Gheorghies, 2003], as its name 
says, integrates several impact measures to adapt the operator application rates: the 
frequency of absolute improvements (over the best), simple improvements (over the par-
ents), plateau walks (same fitness than its parents), and worsenings (fitness lower than its 
parents) achieved by the applications of each operator within a generation. The operators 
are selected via a PM-like scheme based on the ranks of the operators with respect to 
the measured frequencies. Besides, the IAGA method also implements an adaptation of 
some internal parameters of the operators, but their description is out of the scope of this 
Section. The IAGA method shows to perform much better than the GA using different 
sets of static probabilities on different instances of the Royal Road problem.

### 4.5.2 Diversity-based Approaches

Besides the Compass [Maturana and Saubion, 2008a], described in detail in Section 4.3.4, 
other similar approaches have been proposed in the literature, aggregating the fitness and 
diversity measures, or using only the diversity as an impact measure after an operator 
application.

The Adaptive GA (AGA) proposed in [Srinivas and Patnaik, 1994] is, to the best of 
our knowledge, the first method proposed for the adaptation of the operators application 
rates that also take into account the diversity in the decision process, motivated by the 
difficulty of solving multi-modal problems. Its adaptation method can be briefly described 
as follows. The crossover and mutation rates are varied, for each individual, according 
to the difference between the fitness of the best and the fitness of the current individual,
4.5 Some Adaptive Operator Selection Combinations

divided by the difference between the fitness of the best and the average fitness of the
current population. The numerator measures how close to the best individual is the cur-
rent individual; the fitter the individual is, the less it will be disrupted by the operators.
Conversely, the denominator roughly measures the level of convergence of the population;
the more converged it is, the higher is the variation that should be introduced, in order to
possibly escape from local optima. The balance between both intensification and diver-
sification (the same as exploitation and exploration, respectively) factors is controlled by
a user-defined hyper-parameter for each operator. AGA is shown to significantly outper-
form a standard GA with fixed operator probabilities on a set of continuous benchmark
problems; a much higher gain is achieved in the highly multi-modal problems, as expected.

A similar approach, the Diversity-Guided EA (DGEA) [Ursem, 2002], is, as its name
says, completely guided by the level of diversity in the population. The main objective is
again to avoid premature convergence. A special diversity measure is proposed, referred to
as the “distance-to-average-point”, which takes into account the size of the search space,
the size of the population, and the sum of the differences between the genes of each indi-
vidual, in order to evaluate how converged the population is. Once every generation, the
algorithm switches between exploration and exploitation behaviors, based on the assessed
diversity level. Intuitively, exploration is performed by the generation of an entire pop-
ulation via the sole use of a mutation operator, while exploitation is done by crossover.
Compared to a set of non-adaptive GA schemes on some continuous benchmark problems,
the DGEA presents better performance, consistently reducing in around 25% the number
of fitness evaluations to attain a given target solution.

The Adaptive Operator Rate Controlled EA (AORCEA) [Giger et al., 2007] is an in-
teresting AOS method, although quite complex. It is very different from the previously
mentioned approaches in what concerns the update of the operators application rates. To
start with, the criterion to evaluate the impact of an operator application depends on the
level of stagnation of the search process, which is calculated based on the frequency dis-
tribution of the fitness values of the current population. In case more diversity is needed,
applications of operators are evaluated based on how different are the offsprings they gen-
erate with respect to their parents (Euclidean distance); the relative fitness improvements
are used otherwise. The operators are ranked according to how well they perform in
average with respect to the chosen criterion during the given adaptation cycle. Their ap-
plication rates are then linearly updated, by taking into account these ranks and the ratio
between the level of stagnation and a user-defined hyper-parameter. This hyper-parameter
exerts a function analogous to the greediness control $\beta$ parameter used by AP (described
in Section 4.4.2). The AORCEA presents significantly better results when compared to
a non-adaptive GA on a set of continuous benchmark functions, and also on a real-world
problem, the optimization of the structure of a tubular steel frame for a motorcycle.

Guided by the same motivations than those of the Compass fitness and diver-
sity aggregation method [Maturana and Saubion, 2008a] (see Section 4.3.4), two other
Credit Assignment mechanisms have been later proposed in [Maturana et al., 2009b; 
Maturana et al., 2010b], directly inspired by the concepts of Pareto dominance. Con-
sidering the diversification and the intensification as two criteria to be optimized, the first
scheme, referred to as Pareto Dominance (PD), evaluates the operator according to the
number of other operators it dominates, i.e., operators that performed worst in average than the operator under assessment on both objectives. Oppositely, the second scheme, *Pareto Rank* (PR), accounts for the number of operators that dominate the operator under assessment. The main difference between both is that the latter encourages only the use of non-dominated operators, while the former rewards more accurately all the operators that are performing well on average: the PD scheme is thus the best choice, as empirically shown in the cited references. Combined with an external scheme that dynamically changes the set of available operators while solving the problem, the PD *Credit Assignment* scheme with the PM *Operator Selection* mechanism achieves better performance than other adaptive combinations on a set of SAT instances.

4.5.3 Fuzzy-based Approaches

Another kind of approach with several examples found in the literature is the use of Fuzzy Logic Controllers (FLC) to control the selection of operators. Some of these methods will be briefly reviewed now.

The seminal paper on this matter, to the best of our knowledge, is the work by [Lee and Takagi, 1993], in which the operators application rates are deterministically controlled according to fuzzy rules based on population-wise measures: the average, best, and worst fitness values found in the current population. The FLC itself is off-line tuned by another GA at the meta level, according to its performance on the control of the operators of the main GA while solving the well-known set of “5 DeJong functions”. The resulting tuned algorithm is later applied to the optimization of another FLC solving the inverted pendulum problem, outperforming a simple static GA in terms of number of evaluations to achieve a given target solution. Although being out of the scope of this thesis, it seems worthy to highlight the several levels of efficient hybridizations between fuzzy and evolutionary techniques that can be found in this work. In summary, a GA is used to optimize an FLC, that controls the operator rates of another GA, which is used to optimize another FLC, that is finally applied to a control problem. Besides, an important motivation for using the kind of human-comprehensive knowledge representation employed by FLCs is that experts can try to incrementally enhance the controller with their own understanding about the problem.

In [Herrera and Lozano, 2001], an FLC optimized by a meta-GA is used to control the use of 12 different operators by a GA, that is applied to continuous optimization problems. But, in this work, the controller is optimized while solving the problem, by means of a separate GA that co-evolves with the GA to be controlled. A gain with respect to non-adaptive schemes is not shown in terms of efficiency, but rather in terms of robustness when applied to continuous benchmark problems with very different levels of difficulty.

Another work using FLC to control the operators application within a GA is presented in [Maturana and Saubion, 2007b; Maturana and Saubion, 2007a; Maturana and Saubion, 2008b]. Similarly to the previously mentioned PRAM method [Ho et al., 1999], the adaptation method is divided into two periods, a learning phase, during which the FLC is improved after the empirical knowledge gathered from several trials of all the operators; and a production phase, when the learned rules are actually employed.

60
4.5 Some Adaptive Operator Selection Combinations

to deterministically select which operator should be applied, according to the feedbacks (diversity and quality variations) received from the search. However, around 55% of the generations are spent by the learning phase, what greatly deteriorates the performance of the algorithm, specially if disruptive operators are employed (as also previously discussed for the PRAM method). The presented results do not compare the developed fuzzy-based AOS scheme with other methods from the literature, but rather with different variants of its own, on an instance set of the Quadratic Assignment Problem (QAP).

4.5.4 Other Approaches

Most of the previously cited works use slight variations of the PM method for the adaptation of the rates and further Operator Selection, while a lot of effort is invested on the many different alternatives mentioned for the Credit Assignment part. There is no clear evidence to support the preference for enhancing just one of the modules, but given the difficulty of the task, it might seem relevant to separately investigate both issues. A first step along this line is taken by [Thierens, 2005], which proposes a new mechanism for Operator Selection, the Adaptive Pursuit (AP) (see Section 4.4.2), while assessing it on an artificial dynamic scenario, assuming the reward associated to each operator to be known. The reward distributions are modified every $\Delta T$ steps, with AP showing a much superior performance than PM. This artificial scenario will be described in detail in Section 5.4.1.

In [Whitacre et al., 2006], attention is given to the Credit Assignment scheme again, while the Operator Selection is the well-known PM, autonomously selecting between 9 operators. A 10th operator is applied according to a deterministic scheduler in order to enforce some level of diversity. Several alternative impact measures are compared, e.g., the rank of the generated offspring within the current population, and the age of the generated solution; for the latter, the adaptation needs to happen once every many generations (20 in this case). From these impact measures, the main novelty proposed in this work is the use of a Credit Assignment mechanism that rewards the production of outlier solutions, which are found out based on statistics over the whole set of generated solutions. This method is shown to be superior to the common Average scheme on a set of continuous benchmark problems.

4.5.5 AOS within Other Evolutionary Algorithms

All the works reviewed in this Section up to now, as well as most of the literature on AOS, are proposed in the context of GAs. However, the concept is general enough to be applied to other EAs (as well as to other meta-heuristics).

For instance, in [Niehaus and Banzhaf, 2001], the PM method is used to select between special operators in the Genetic Programming (GP) framework, based on the success rate of each operator, a successful trial being defined as the generation of an offspring fitter than its parents. The proposed adaptive scheme presents superior performance than the standard GP using both randomly and empirically defined static application rates, on different problems of symbolic regression and classification.

Some works in the scope of Differential Evolution (DE), yet another EA, can also be
found in the literature. For instance, although being referred to as Self-Adaptive DE, the SaDE method [Qin et al., 2009] employs indeed the AOS paradigm, with the PM method selecting between DE mutation strategies according to the success rate of each operator. This scheme is combined with another method to dynamically adapt the crossover rate CR and the mutation scaling factor F. The SaDE method outperforms the DE independently applying each of the available mutation strategies, and other previously proposed adaptive and self-adaptive schemes, on a set of continuous benchmark problems.

In a collaboration with the China University of Geosciences [Gong et al., 2010; Gong et al., 2011], we have also used the PM method within DE, but this time using the average of the relative fitness improvements as Credit Assignment. Along the same lines, a large part of the empirical results that will be presented in Chapter 6 were achieved applying our Rank-based Multi-Armed Bandit AOS mechanisms (which will be described in Chapter 5) to DE on continuous benchmark optimization problems [Fialho and Ros, 2010; Fialho et al., 2010b].

4.6 Discussion

Most of the Adaptive Operator Selection (AOS) methods surveyed in this Chapter are employed to select between operators within Genetic Algorithms; a few approaches considering other variants, namely Genetic Programming and Differential Evolution, are also mentioned. Although all these works are in the scope of Evolutionary Algorithms, the adaptive paradigm, as reviewed in Section 3.3.2, is indeed very general. In fact, any stochastic algorithm can benefit from this kind of approaches. At a higher level of abstraction, AOS schemes can also be used to select between different algorithms at the hyper level, what is nowadays commonly referred to as Hyper-Heuristics (we refer to reader to [Burke et al., 2010] for a recent very comprehensive review on this).

The ideal scenario for the use of the adaptive paradigm can be briefly characterized as follows:

1. The algorithm has some choice to be made among different options that directly affect the search process;
2. This choice is supposed to not have only a single best component during the whole search process; instead, different components perform best during different stages of the search;
3. It is possible to have an instantaneous feedback from the search process as a result of the choice.

The first and the third issues are, in fact, requirements to be able to use adaptive methods in general, that indeed quite always hold in the case of stochastic algorithms. The second issue can be relaxed a bit: even if there is only one unique best option for the given choice, it is usually unknown to the user, and generally problem-dependent. Even in this case, thus, the use of an adaptive method can be justified anyway, in order
4.6 Discussion

to automatically find the best option while solving the problem. The price to pay is a small loss in terms of performance (the time taken in order to find the best option), which is compensated by the fact that everything is done during an optimization run, while several runs would be needed in order to apply an off-line tuning procedure, as discussed in Section 3.3.2.

In the context of AOS, the input is the feedback received from the search, and the output is the operator to be applied, as depicted in Figure 4.1. The generality of the method, however, depends on how general is the information used to constantly update it. The methods surveyed in this Chapter use very general information, such as the fitness and the diversity. In some cases, however, one might want to explore some prior knowledge about the characteristics of the problem in order to do a more efficient AOS. A lot of research on this has been done in the very competitive context of SAT problems. An example of a problem-specific AOS method is the recent NCVW (Non-, Clause, and Variable Weighting) [Wei et al., 2008], which uses SAT specific features, the variable and clause weights, in order to choose between three well-known variable selection heuristics. Although losing generality, exploring prior knowledge about the problem can be very beneficial for the search process. Indeed, if the motivation is to go for state-of-the-art results, this is very probably the path to be taken in any domain. On the opposite, if the motivation is to have a general method to adapt the operator selection and achieve good performance in very different situations and with different algorithms, such kind of problem-specific knowledge should be avoided.

Finally, as remarked in the bibliographic review presented in Section 4.5, most of the mentioned works concentrate a lot of effort on just one component of the AOS, usually using a quite common choice for the other one. For instance, lots of methods use a very complex Credit Assignment scheme, while implementing the standard PM for Operator Selection. In this thesis, we will present different contributions addressing both issues: the bandit-based approaches on the Operator Selection side; the Extreme fitness improvement and the Rank-based measures over the fitness for Credit Assignment. Besides, the Compass [Maturana and Saubion, 2008a] aggregation between fitness and diversity will also be considered, after some work done in collaboration with the authors [Maturana et al., 2009a]. The other options for Credit Assignment are the common Instantaneous and Average fitness improvement over the parents. All the proposed AOS combinations, presented in Chapter 5, will be compared with both PM and AP Operator Selection methods, combined with the above-mentioned Credit Assignment schemes.
Part III

Contributions
Chapter 5

Contributions to Adaptive Operator Selection

Contents

5.1 Introduction ................................................. 66
5.2 Contributions to Credit Assignment ......................... 68
  5.2.1 Basic Credit Assignment Scheme: Fitness Improvements . 68
  5.2.2 Extreme Fitness Improvement ............................ 69
  5.2.3 Normalized Fitness Improvement ......................... 70
  5.2.4 Rank-based Credit Assignment Schemes .................. 70
  5.2.5 Comparison-based Credit Assignment Schemes .......... 77
5.3 Contributions to Operator Selection ......................... 78
  5.3.1 Basic Operator Selection Scheme: Multi-Armed Bandit .... 78
  5.3.2 Dynamic Multi-Armed Bandit ............................. 81
  5.3.3 Sliding Multi-Armed Bandit .............................. 84
  5.3.4 Rank-based Multi-Armed Bandit .......................... 86
5.4 Contributions to Empirical Assessment ...................... 88
  5.4.1 Basic Artificial Scenario: Uniform ....................... 89
  5.4.2 Boolean and Outlier Scenarios .......................... 89
  5.4.3 Two-Value Scenarios .................................... 90
5.5 Discussion .................................................. 92
Chapter 5. Contributions to Adaptive Operator Selection

In this Chapter, we present our main algorithmic contributions to the Adaptive Operator Selection problem, namely, the extreme and the rank-based approaches for Credit Assignment, the bandit-based techniques for Operator Selection, and the artificial scenarios proposed to their empirical assessment.

5.1 Introduction

As discussed throughout Chapter 4, in essence, the goal of Adaptive Operator Selection (AOS) is to select on the fly the operator maximizing some measure of quality, usually, though not exclusively, reflecting the fitness improvement brought by its application. AOS thus raises two main tasks, referred to as Operator Selection and Credit Assignment (Section 4.2). This Chapter describes the contributions developed during this thesis work for each of these tasks, as well as for their empirical assessment. A brief summary of these contributions, in a chronological order, is presented as follows.

Starting with the first task, the Operator Selection might be seen as yet another instance of the Exploration versus Exploitation (EvE) dilemma: the operator that is currently known to be the best should be used as much as possible (exploitation), while the other operators should also be tried from time to time (exploration). The exploration needs to be done, on the one hand, because some seemingly poorly-performing operators might just have been unlucky on its recent trials; and on the other hand, due to the dynamics of the evolutionary process, i.e., one of the other operators might eventually become the new best operator at a further moment of the search.

The EvE trade-off has been intensively studied in the context of Game Theory, in the framework of the so-called Multi-Armed Bandit (MAB) [Lai and Robbins, 1985; Auer et al., 2002]. The use of MAB algorithms to solve the EvE dilemma has been investigated in the selection between different algorithm portfolios to solve decision problems [Gagliolo and Schmidhuber, 2008], before being extended to the AOS context in the work presented here. Our preliminary attempt to do so was by directly using a slightly modified version of the Upper Confidence Bound (UCB) algorithm [Auer et al., 2002] (described in more detail in Section 5.3.1), which was chosen by providing asymptotic optimality guarantees with respect to the total cumulative reward in MAB problems. However, these guarantees hold only in a stationary context; some modifications need to be proposed in order to efficiently use the UCB algorithm in the dynamic context of AOS – this is where most of the contributions developed in this thesis are concentrated.

A first proposal, referred to as Dynamic Multi-Armed Bandit (DMAB) [Da Costa et al., 2008], is presented in Section 5.3.2. The DMAB proceeds by coupling the original UCB technique with the Page-Hinkley statistical change-detection test [Hinkley, 1970]: upon detecting a change in the operator quality distribution, the MAB process is restarted from scratch.

Concerning the Credit Assignment, most of the AOS combinations found in the literature use some simple statistics over the fitness improvements. Instead of using the
common Instantaneous and Average Credit Assignment schemes (see Section 5.2.1), we proposed the use of Extreme fitness improvements [Fialho et al., 2008], based on the assumption that rare but high improvements might be even more important than frequent but moderate ones (Section 5.2.2).

The combination of Extreme Credit Assignment with DMAB Operator Selection referred to as the Ex-DMAB AOS technique, showed to be very efficient, outperforming the baseline approaches on different benchmarking scenarios [Fialho et al., 2008; Fialho et al., 2009a; Maturana et al., 2009a]. However, directly using the raw values of the fitness improvements (Credit Assignment) to update the preferences (Operator Selection) of the AOS technique showed not to be a very good approach: as different problems have different fitness ranges, this AOS scheme needs to have its hyper-parameters tuned for every new problem in order to achieve good performance. For this reason, on the Credit Assignment side, we proposed the use of a simple normalization of these raw values [Fialho et al., 2009b] (described in Section 5.2.3).

On the Operator Selection side, even with the normalized rewards, the hyper-parameter of the DMAB controlling the change-detection test continued to be very problem-dependent, as the restarting mechanism is directly related to the dynamics of the fitness landscape. This was the main motivation for the proposal of a smoother way to account for dynamic environments in the MAB framework, referred to as Sliding Multi-Armed Bandit (SLMAB) [Fialho et al., 2010a], presented in Section 5.3.3. Briefly, it uses a sliding time window to gracefully update the operator quality estimates, discarding ancient events while preserving the information from recent operator applications. Contrasting with DMAB, SLMAB does not call upon an external monitoring of the evolution process and involves only one hyper-parameter, as the original MAB technique, while DMAB has two.

By the use of normalization, the effects of problem-dependency on the Extreme Credit Assignment are smoothed, but not eliminated. This is what led us to the proposal of the two last Credit Assignment measures, completely based on ranks, the Area-Under-Curve (AUC) and the Sum-of-Ranks (SR) [Fialho et al., 2010c] (Section 5.2.4). In addition to the gain in robustness achieved by the use of rank-based measures, the use of ranks over the exact fitness values rather than ranks over the fitness improvements preserves the important invariance property of the method with respect to monotonous transformations (i.e., comparison-based), as presented in Section 5.2.5. These rank/comparison-based Credit Assignment schemes were combined with a simplified version of the UCB, to which we refer to as the Rank-based Multi-Armed Bandit (RMAB) (Section 5.3.4).

By the time this manuscript is being written, the AOS technique constituted by the RMAB Operator Selection method with the AUC Credit Assignment scheme is our final and recommended proposal in case one wants to implement the AOS paradigm: it achieves state-of-the-art (or equivalent) performance while being very robust with respect to their hyper-parameters, as confirmed by the results presented in [Fialho et al., 2010c; Fialho et al., 2010b], which will be detailed in Chapter 6.

Additionally, while developing these AOS combinations, we have also proposed some new artificial scenarios for their empirical assessment. The Boolean and the Outlier scenarios [Da Costa et al., 2008] were introduced to evaluate the AOS schemes in situations
Chapter 5. Contributions to Adaptive Operator Selection

involving five artificial operators with different reward distributions than the previously existing Uniform scenario [Thierens, 2005]. The latter is described in Section 5.4.1, while the two newly proposed scenarios are presented in Section 5.4.2. Besides, another family of artificial scenarios was proposed to simulate different situations with respect to the mean and variance of rewards given by two artificial operators, referred to as the Two-Values (TV) benchmarks [Fialho et al., 2010a], presented in Section 5.4.3.

Each of these mentioned contributions will be detailed in the following of this Chapter, divided into three categories. Section 5.2 presents the contributions for Credit Assignment; while the new methods for Operator Selection are described in Section 5.3. At the end of the presentation of each Operator Selection scheme, the corresponding AOS combinations are reminded. Finally, Section 5.4 details the newly proposed artificial scenarios for the empirical assessment of the developed AOS combinations. For each Section, the basic or initial approaches are also reminded prior to the presentation of the contributions, for the sake of self-containedness.

The last contribution of the present thesis consists in a principled and systematic empirical comparison of the proposed AOS methods, compared with one another and with some baseline approaches. Several experiments were done on some different benchmarking scenarios; they will be analyzed in detail in Chapter 6.

5.2 Contributions to Credit Assignment

Credit Assignment is the name given to the scheme that assesses the performance of an operator regarding the progress of the search, which can be measured in different ways, as reviewed in Section 4.3. Starting from the existing Instantaneous or Average of the fitness improvements brought by the application of a given operator (Section 5.2.1), we have proposed the use of Extreme values (Section 5.2.2). In the quest for a higher robustness, a simple normalization over these raw values was next proposed (Section 5.2.3), before the development of the most recent and very robust rank-based Area-Under-Curve (AUC) and Sum-of-Ranks (SR) schemes, both described in Section 5.2.4. These latter methods, when using the fitness values instead of the fitness improvements as raw measures of impact, become fully Comparison-based, as emphasized in Section 5.2.5.

It is worth remembering that, after discussion in Section 4.3.3, for all the Credit Assignment schemes considered (the baseline and the schemes proposed by us), no ancestors are rewarded: the credit is only assigned to the operator that was used to achieve the given fitness improvement. Besides, by convention, all schemes assign a null credit (=0) in case the computed credit is negative.

5.2.1 Basic Credit Assignment Scheme: Fitness Improvements

As discussed in Section 4.3, a Credit Assignment scheme is defined by three aspects: (i) how to measure the impact of an operator application; (ii) how to assign credit to the operator based on this measured impact; and (iii) to which operator the credit should be assigned to.
Concerning the first issue, the Credit Assignment schemes proposed in this thesis use as a measure of the impact after an operator application, unless stated otherwise, the fitness improvement brought by the generated offspring over its parent, if a mutation operator is used, or over the best of its parents in the case of crossover. Formally, let $F$, $o$ and $x$ respectively denote the fitness function, a variation operator, and an element of the current population. The impact of an operator application on the search at time $t$, i.e., the fitness improvement, is measured as $\delta(t) = \max((F(o(x)) - F(x)), 0)$ when the objective is to be maximized, $\delta(t) = \max((F(x) - F(o(x))), 0)$ otherwise. For the description of the Credit Assignment schemes throughout this Section, we assume the objective function to be maximized.

The most common ways of assigning credit will be used as baseline for comparison, namely: (i) the Instantaneous, which credits the operator according to the fitness improvement received after its most recent application; and (ii) the Average, which assigns as credit the average of the fitness improvements achieved by its last $W$ applications, $W$ being a hyper-parameter (the size of the sliding window for each operator) that needs to be defined by the user. We refer the reader to Section 4.3.2 for a more extensive discussion including some references for both approaches.

5.2.2 Extreme Fitness Improvement

Our first proposal for the Credit Assignment problem, referred to as the Extreme value-based scheme [Fialho et al., 2008], is inspired by the following remark. Let us consider an operator bringing frequent, small improvements, and compare it to an operator bringing rare, large improvements. The latter operator will hardly be considered if the reward reflects the Average fitness improvement, as the average estimated over a few trials is likely to be 0 (and this becomes even worse in case the Instantaneous values are considered), implying that very few further trials of this operator will take place, although it might be the current best operator.

Hence, as advocated in [Whitacre et al., 2006], attention should be payed to extreme, rather than average, events. Incidentally, the role of extreme events in design has long been acknowledged in numerical engineering (e.g., taking into account rogue waves when dimensioning an oil rig). Additionally, they have been receiving an ever growing attention in the domain of complex systems, as extreme events govern diffusion-based processes ranging from epidemic propagation to financial markets.

The proposed Extreme value-based Credit Assignment mechanism proceeds as follows. When operator $o$ is selected after the Operator Selection rule under examination, $o$ is applied on the current individual $x$; the fitness of the generated offspring is computed and the improvement achieved over its parents is added to the sliding window of operator $o$ with size $W$, implementing FIFO order. Finally, the credit to be assigned to this operator is set to the maximum fitness improvement found in its sliding time window.

More formally, let $t$ be the current time step, and $t_{o,1}, t_{o,1}, \ldots, t_{o,i-1}$ denote the time step where operator $o$ was used for the last time (respectively, the last time before $t_{o,i-1}$). If $\delta_o(t)$ denotes the fitness improvement observed at time $t$, then the credit to be
assigned to operator $o$ is computed as:

$$r_o(t) = \max_{i=1...W} \{\delta(t, o, i)\}$$  \hspace{1cm} (5.1)

Hence, the Extreme value-based mechanism involves a single hyper-parameter, the window size $W$, as does the previously mentioned Average scheme. This hyper-parameter is meant to reflect the time scale of the process. If too large, the switches between two different situations with respect to operators qualities might be delayed, i.e., operators will be applied after their optimal epoch. If $W$ is too small, operators causing large but infrequent jumps might be ignored, as successful events will probably not be observed at all, or too rapidly forgotten. The Extreme value-based scheme will be simply referred to as Extreme in the following.

### 5.2.3 Normalized Fitness Improvement

Although alleviating the user from the need of selecting which operators should be applied to the problem at hand, and doing so while solving the problem, each of the AOS methods presented in this thesis involves some hyper-parameters that also need to be tuned. The three Credit Assignment schemes previously mentioned, namely, Instantaneous, Average, and Extreme, reward the operators based on simple statistics over the fitness improvements achieved by their application. The use of the raw values of the fitness improvements, however, makes these schemes (and consequently the hyper-parameter tuning of the AOS techniques implementing them) to be very problem-dependent.

Firstly, different problems have fitness ranges with different variance and at different orders of magnitude. Hence, a given AOS setting is efficient just when applied to the problem used during the off-line tuning phase of its hyper-parameters. Additionally, and even more importantly, the fitness variance, as well as the magnitude of the possible rewards received, tend to reduce as the search approaches the optimum, while improvements themselves tend to become more and more scarce. Thus, even when the AOS is very carefully tuned for the problem at hand, its behavior might not be optimal during all the search process.

A proposal to improve the robustness of the mentioned Credit Assignment schemes over different problems was to use a simple Normalization scheme [Fialho et al., 2009b]: the credit to be assigned to the operator is priorly divided by the maximum credit that would be assigned to any of the operators, according to the Credit Assignment scheme under employment (e.g., the mentioned Instantaneous, Average and Extreme schemes). In this way, no matter the moment of the search or the problem that is being tackled, all the rewards are in the real-value interval between 0 and 1, and the current best operator always receives a reward of 1.

The utilization of all these basic Credit Assignment schemes described in Sections 5.2.1 to 5.2.3, namely, the Absolute or Normalized output of Instantaneous, Average, or Extreme schemes, calculated over the fitness improvements, is exemplified in the form of a pseudo-algorithm in Algorithm 5.1, considering a maximization function. It is important to remember that, for all these “basic” schemes, there is one FIFO window of size $W$ for each of the $K$ operators.
5.2 Contributions to Credit Assignment

Algorithm 5.1: Credit Assignment Schemes over $\Delta F$ (op, type, norm, W, K)

1: if type = Instantaneous then
2:   reward $\leftarrow$ last($wRewards_{op}$)
3: else if type = Average then
4:   reward $\leftarrow$ avg($wRewards_{op}$)
5: else if type = Extreme then
6:   reward $\leftarrow$ max($wRewards_{op}$)
7: end if
8: if norm then // normalization
9:   normfactor $\leftarrow$ max$_{i=1...K}${this.GetReward(i, type, norm=false, W, K)}
10: reward $\leftarrow$ reward/normfactor
11: end if
12: return $\max\{reward, 0\}$

5.2.4 Rank-based Credit Assignment Schemes

The normalization over the fitness improvement contribute into reducing the mentioned effects of problem-dependency, but do not eliminate the problem. An important flaw of this approach is that, as the normalization factor depends on the region of the landscape that is currently being explored, the same gain might have different weights in the update of the empirical estimates throughout the search process, and this is likely to still lead to problem-dependent hyper-parameter configuration for the AOS schemes.

Inspired by the gain in robustness achieved by GAs when employing rank-based parental selection schemes (e.g., the tournament selection) instead of selection schemes over the raw fitness values (as the fitness-proportional roulette-wheel method), it seems clear that the use of ranks instead of raw values for the Credit Assignment is a way towards robust AOS techniques. Following this path, we have proposed two Credit Assignment schemes totally based on ranks, namely, the Area-Under-Curve (AUC) and the Sum-of-Ranks (SR) [Fialho et al., 2010c; Fialho et al., 2010b], which will be now described in turn.

Sliding Window

Besides the fact that ranks are used to assign credit, another major difference is that these newly proposed schemes maintain the gains achieved by all the operators in a single sliding window of size $W$, still being updated in a FIFO way; while in the previously described Credit Assignment schemes, there is one separate window for each operator. Each slot in this unique window is a structure that contains the index of the operator that was applied, the fitness improvement (or other impact assessment) achieved by this operator application, and the corresponding ranking of this fitness improvement with respect to all the other fitness improvements stored in the current window.

The motivation for using a unique FIFO window for all operators is related to the already discussed dynamics of the AOS problem. By doing so, the oldest result stored in
Chapter 5. Contributions to Adaptive Operator Selection

the window is as old as \( W \) operator applications; while in the case of multiple windows, the results of very old applications of a given operator not applied during a long period might still be used for the estimation of its empirical quality, although not reflecting the reality anymore. Besides, as the inclusion of a new value of fitness improvement in the window alters the ranking of all the values worse than it (and in case the window contained already \( W \) values before this inclusion, the oldest value is deleted from the window; then the ranking of all the results worse than the excluded value are also updated), the application of one operator might also affect the empirical quality estimates of the other operators (in addition to its own). In this way, thus, the dynamics with respect to the performance of the operators are already handled to some extent on the Credit Assignment side.

Decaying Mechanisms for Rank-values

Following, and somehow smoothening the intuition of the Extreme Credit Assignment presented in Section 5.2.2, special mechanisms were proposed for the assignment of rank-values. The overall idea is that the top-ranked rewards should exert a higher influence on the calculation of the credit assigned to each operator.

Firstly, each slot in the window is ranked according to the values of the fitness improvements stored in it, in a descending order, with the slot \( i \) receiving a rank-value of \( R(i) \). A decay factor \( D \in [0,1] \) is then applied over these rank-values. The final decayed rank value \( \text{Decay}(R(i)) \), which defines the weight of each operator application \( i \) in the AUC and SR Credit Assignment schemes, is then calculated as:

\[
\text{Decay}(R(i)) = D^{R(i)}(W - R(i))
\] (5.2)

The hyper-parameter \( D \) defines how skewed the ranking distribution is. The smaller the value of \( D \), the faster the decay (i.e., the more Extreme it is). The use of \( D = 1 \) is equivalent to a linear decay (i.e., \( W - R(i) \)). We refer to this as the Decay approach.

Another way of providing a decaying mechanism is the Normalized Discounted Cumulative Gain (NDCG), originally proposed in the context of Information Retrieval (IR) [Järvelin and Kekäläinen, 2000; Burges et al., 2005]. The motivation for using it is very similar: the discovery of highly relevant documents should receive a higher weight than the marginally relevant documents in the evaluation of the effectiveness of an IR method. Formally, the original NDCG method assign to element \( i \) (documents in the original context, operator applications in ours) the following rank-value:

\[
\text{Original NDCG}(R(i)) = \frac{2^{R(i)} - 1}{\log(1 + i)}
\] (5.3)

It seemed to be an interesting method for assigning decayed rank-values to the operator applications, mainly due to the fact that it does the same job, but without requiring the definition of any hyper-parameter. In the original context, however, the NDCG measure does not consider only the ranking (with respect to relevance) of the corresponding document, but also the order (the \( i \) in the denominator of Equation 5.3) in which it appears in the original document list: it is important for the IR methods to bring the most relevant
5.2 Contributions to Credit Assignment

documents firstly. But in the AOS context, the order in which the top-ranked impact measures are achieved is not important, what matters is how many top-ranked impact measures are brought by each operator, in the time scale limited by the sliding window size $W$. Therefore, in order to apply the NDCG method in the AOS context, it was re-written as follows:

$$\text{Adapted NDCG}(R(i)) = \frac{2^{(W-R(i))} - 1}{\log(1 + R(i))}$$  \hspace{1cm} (5.4)

In practice, however, we found out that the NDCG method is rather equivalent to the Decay scheme defined in Equation 5.2 when using $D = 0.4$, as shown in Figure 5.1. Anyway, the NDCG and the Decay approaches, combined with each of the two rank-based Credit Assignment schemes that will be presented in the following, will be independently considered in the experiments presented in Chapter 6. It is true that this choice complicates the experimental setup and the analysis of results (by adding another degree of freedom for the rank-based approaches), but we decided to do so because, in this way, it becomes possible to verify how much can be gained in terms of performance by trying some different values for $D$ in the Decay scheme, with respect to the fixed parameter-less NDCG approach.

![Figure 5.1: Comparison between different decaying mechanisms for $W = 15$.](image)

**AUC method: Rank-based Area-Under-Curve**

The Area-Under-Curve (AUC) Credit Assignment method, as its name says, borrows ideas from the Area Under the ROC Curve, a criterion originally used in Signal Processing and later adopted in Machine Learning to compare binary classifiers, with the property of being
robust with respect to class imbalance [Bradley, 1997]. Originally, the Receiver Operating Characteristic (ROC) curve depicts how the true positive rate varies with the false positive rate. This indicator is adapted to the rank-based assessment of operators as follows.

Let us consider the list of fitness improvements achieved in a given time window, and let the list be ranked after the raw values of these fitness improvements. The ROC curve associated to a given operator \( o \) is drawn by scanning this ordered list, starting from the origin: a vertical segment is drawn when the given fitness improvement has been generated by \( o \), a horizontal segment is drawn otherwise, and a diagonal segment is drawn in case of ties. The quality associated to operator \( o \) finally is the Area Under this Curve.

As an example, considering the list of 15 operator applications presented in Table 5.1, the ROC curve corresponding to the quality of operator 1 with respect to the others is the solid line, upper bound of the area illustrated in Figure 5.2a. The grey area corresponds to the AUC evaluation for this operator. In this example, for the sake of clarity, all rank positions have the same weight, \( i.e. \), all horizontal and vertical segments have length 1. However, it makes sense to give more weight to the top-ranked values, as previously discussed. A decay factor can be applied, with each segment of the ROC curve being re-scaled accordingly; in this case, there is the need of defining the hyper-parameter \( D \) (or use the parameter-less NDCG scheme). Figure 5.2b presents the AUC for the same set of rewards, but using decay factor \( D = 0.4 \). The corresponding weight of each segment, after application of the decaying mechanism, is depicted in the last column of Table 5.1.

<table>
<thead>
<tr>
<th>( \Delta F )</th>
<th>Op.</th>
<th>( R(i) )</th>
<th>( D1.0 )</th>
<th>( D0.4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>12.5</td>
<td>2.4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>12.5</td>
<td>2.4</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>0.69</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>0.31</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>0.077</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>0.077</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>0.077</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td>0.012</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>11</td>
<td>5</td>
<td>0.0049</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>12</td>
<td>4</td>
<td>0.002</td>
</tr>
<tr>
<td>0.4</td>
<td>4</td>
<td>13</td>
<td>3</td>
<td>7.3e-4</td>
</tr>
<tr>
<td>0.2</td>
<td>4</td>
<td>14</td>
<td>2</td>
<td>2.4e-4</td>
</tr>
<tr>
<td>0.1</td>
<td>3</td>
<td>15</td>
<td>1</td>
<td>6.1e-5</td>
</tr>
</tbody>
</table>

Table 5.1: Sample list of 15 operator applications, ordered by fitness improvement \( (\Delta F) \) in descending order, with respective original \( (R(i)) \) and decayed \( (D1.0 \) and \( D0.4 \) rank values.

It is important to note that, although using rank-based measures, the range of the credit values provided by AUC is sensitive to the window size \( W \): the bigger the window,
5.2 Contributions to Credit Assignment

In order to avoid this situation, we propose again the use of a normalization scheme. In the preliminary empirical results of the AUC Credit Assignment scheme combined with the Rank-based Multi-Armed Bandit (RMAB) Operator Selection mechanism (which will be presented in Section 5.3.4), published in [Fialho et al., 2010c], the normalization was done on a per-axis basis, i.e., in the AUC plot (exemplified in Figure 5.2), the $x$ coordinates (respectively the $y$ ones) were divided by the maximum value to be plotted in $x$ (respectively $y$). What happens in this case, however, is that different operators might have a different number of rewards being assigned to each axis; therefore, they can be normalized by different values. Consequently, by using this scheme, it was later verified that good performance was attained just in situations involving only 2 operators. Later on, in [Fialho et al., 2010b; Fialho and Ros, 2010], we used a much simpler normalization scheme that eliminates this effect. Basically, the AUC credit of a given operator is normalized by the sum of the credits of all operators, so that their sum is equal to 1. This is the current version of the AUC Credit Assignment scheme, simply referred to as AUC in the remainder of this text; the preliminary version will be called as AUCv1 for the sake of distinction. An empirical comparison between both versions on the OneMax problem will be presented in Section 6.4.2.

Figure 5.3 depicts the AUC for each operator, considering the same sample data presented in Table 5.1. In case the same weight is considered for all the rewards (Figure 5.3a), the area corresponding to operator 1 is equivalent to (and the credit to be assigned to it is evaluated as) 56% of the sum of the areas. If a linear decay is used (Figure 5.3b; see the corresponding weights in the fourth column of Table 5.1), the area of operator 1 is increased to 64%. Remembering that the idea is to exploit as much as possible the best operator, Figure 5.3c is a supporting example for the use of a strong decaying factor: with $D = 0.4$, the AUC of operator 1 receives a score of 92 out of 100.
Figure 5.3: Comparison between different decaying mechanisms for the calculation of the AUC credit. The final credit assigned to each operator is its AUC divided by the sum of AUCs of all operators, i.e., the proportion of the sum of the areas that is occupied by it.

A complete and detailed representation of how to calculate the AUC credit is presented in the form of a pseudo-algorithm in Algorithm 5.2.

**SR method: Sum-of-Ranks**

The Sum-of-Ranks (SR) is a much simpler method, that credits the operators with the sum of the ranks of the rewards given after its applications, subject to the same decaying mechanisms previously described. For instance, considering the sample data in Table 5.1, the SR value for operator 1 using decay factor $D = 0.4$ would be: $15 + 7 + 2.4 + 0.69 + 0.077 + 0.012 = 25.179$

As for the AUC, the final credit assigned to the operator is its SR value normalized by the SR values of all operators, so that the sum of the credits assigned to all operators is equal to 1. Following the same example, the credit to be assigned to operator 1 would be
5.2 Contributions to Credit Assignment

Algorithm 5.2: Credit Assignment: Rank-based Area-Under-Curve (W, D, op)

1: area ← x ← y ← 0
2: for rank position r ← 1 to W do // loop on window (just one window for all operators)
3:     area ← D*r(W − r) // calculate weight of rank position in the area
4:     tiesY ← CountTiesTargetOp(r) // # rewards equal to reward ranked r given by op
5:     tiesX ← CountTiesOtherOps(r) // # rewards equal to reward ranked r given by others
6:     if tiesX + tiesY > 0 then // if ties, proportional diagonal trace
7:         for rank position s ← (r + 1) to (r + tiesX + tiesY) do
8:             ∆r ← ∆r + (D*s(W − s)) // sum weights of tied ranks, divided by # ties
9:         end for
10:     x ← x + tiesX · ∆r // sum the rectangle below
11:     area ← area + y · tiesX · ∆r
12:     y ← y + tiesY · ∆r // sum the triangle below slanted line
13:     area ← area + 0.5 · ∆r² · tiesX · tiesY
14:     r ← r + tiesX + tiesY − 1
15:     else if Op_r == op then // if op generated r, vertical segment
16:         y ← y + ∆r
17:     else // if another operator generated r, horizontal segment
18:         x ← x + ∆r
19:     area ← area + (y · ∆r)
20: end if
21: end for
22: return area/∑K{i=1} CreditAssignment.GetReward(W, D, i) // credit = normalized area

evaluated as 25.179/28.050931 = 89.76 out of 100. Formally, the operator i is rewarded at time t as follows:

\[ SR_{i,t} = \frac{\sum_{op_r=i} D^r(W - r)}{\sum_{r=1}^{W} D^r(W - r)} \]  (5.5)

A more complete view of the SR Credit Assignment scheme is presented in Algorithm 5.3, which also includes the handling of ties, not represented in Equation 5.5.

5.2.5 Comparison-based Credit Assignment Schemes

Coming back to the discussion about the robustness of the Credit Assignment schemes in Section 5.2.4, by the use of ranks, both AUC and SR methods are invariant with respect to linear scaling of the fitness function, i.e., their behavior, when applied on a given fitness function \( F \), is exactly the same than when applied to a fitness function defined by \( (a \cdot F) \), with a real value \( a > 0 \). Nevertheless, as the raw rewards that are used here are actual values of fitness improvements, some monotonous transformations will indeed modify the ranking of such values, and, hence, the outcome of the whole algorithm (see some empirical examples on this in Section 6.7.2).

The complete invariance with respect to monotonous transformations can be attained with a very simple modification: the replacement of the fitness improvements by the
Algorithm 5.3: Credit Assignment: Sum-of-Ranks (W, D, op)

1: \( \text{sum} \leftarrow 0 \)
2: \( \text{for} \text{ rank position } r \leftarrow 1 \text{ to } W \text{ do} \) // loop on window (just one window for all operators)
3: \( \Delta r \leftarrow D'(W - r) \) // calculate weight of rank position in the sum
4: \( \text{tiesY} \leftarrow \text{CountTiesTargetOp}(r) \) // \# rewards equal to reward ranked \( r \) given by \text{op}
5: \( \text{tiesX} \leftarrow \text{CountTiesOtherOps}(r) \) // \# rewards equal to reward ranked \( r \) given by others.
6: \( \text{if } \text{tiesX} + \text{tiesY} > 0 \text{ then} \) // if ties
7: \( \text{for} \text{ rank position } s \leftarrow (r + 1) \text{ to } (r + \text{tiesX} + \text{tiesY}) \text{ do} \)
8: \( \Delta r \leftarrow \Delta r + \left( \frac{D'(W - s)}{\text{tiesX} + \text{tiesY}} \right) \) // sum weights of tied ranks, divided by \# ties
9: \( \text{end for} \)
10: \( \text{sum} \leftarrow \text{sum} + \text{tiesY} \cdot \Delta r \)
11: \( r \leftarrow r + \text{tiesX} + \text{tiesY} - 1 \)
12: \( \text{else if } \text{Op}_r = \text{op} \text{ then} \) // if \text{op} generated \( r \)
13: \( \text{sum} \leftarrow \text{sum} + \Delta r \)
14: \( \text{end if} \)
15: \( \text{end for} \)
16: \( \text{return} \frac{\text{sum}}{\sum_{i=1}^{K} \text{CreditAssignment.GetReward}(W, D, i)} \) // credit = normalized sum

fitness values of the newborn offspring as a raw impact measure. By doing this, the AUC and the SR Credit Assignment schemes become fully comparison-based, as only sorting some fitness values is required. This means that, in addition to the linear scaling of the fitness functions, they also become invariant to all the family of fitness functions defined by monotonous transformations over the original function.

These comparison-based Credit Assignment schemes will be referred to as Fitness-based Area-Under-Curve (FAUC) and Fitness-based Sum-of-Ranks (FSR) in the following. As in the original schemes, only the fitnesses of the offspring that improved over their parents are considered, a null reward is assigned otherwise. To date (and to the best of our knowledge), they are the most robust methods for evaluating the operators performance, although being a bit less efficient than the simple rank-based schemes in some cases, as acknowledged in the experimental comparisons that will be presented in Chapter 6.

5.3 Contributions to Operator Selection

Let us turn now to the other component of AOS, the Operator Selection. This is the process used to select the next operator to be applied, based on the credits received from the Credit Assignment scheme under employment during the current search process, as reviewed in Section 4.4. In this thesis, we have proposed and extended the use of the Multi-Armed Bandit (MAB) paradigm for Operator Selection. Starting with a slightly modified version of the Upper Confidence Bound (UCB) algorithm, which will be surveyed in Section 5.3.1, we have proposed two modifications in order to enable it to account for the dynamics of the AOS problem: the Dynamic Multi-Armed Bandit (DMAB) and the Sliding Multi-Armed Bandit (SLMAB). They will be described, respectively, in Sections 5.3.2 and Section 5.3.3. Lately, we have proposed the use of a simplified version of the UCB
algorithm, that directly uses the output of the rank/comparison-based Credit Assignment schemes as the empirical quality estimate of each operator. This technique, referred to as the Rank-based Multi-Armed Bandit (RMAB), will be presented in Section 5.3.4.

5.3 Contributions to Operator Selection

5.3.1 Basic Operator Selection Scheme: Multi-Armed Bandit

A very important concept for efficient problem solving within EAs is that of the Exploration versus Exploitation (EvE) balance: as discussed throughout Chapter 2, the EA needs to efficiently exploit as much as possible the most promising regions of the search space, while it also needs to explore the search space as a whole, in order to have higher chances of finding the true global optimum. In the context of Operator Selection, the same EvE problem exists, but at a higher level of abstraction: the most promising operator needs to be exploited as much as possible, while the other operators also need to be explored from time to time, as the problem of operator selection is dynamic and one of the other operators might become efficient at a further stage of the search. The EvE dilemma has been intensively studied in the context of Game Theory, in the so-called Multi-Armed Bandit (MAB) framework [Lai and Robbins, 1985; Auer et al., 2002]. Based on these works, we have decided to consider the MAB algorithms as possible solutions for the Operator Selection problem, as presented in the following.

To start with, the general paradigm for solving Multi-Armed Bandit (MAB) problems can be formalized as follows. A MAB problem involves $K$ arms; the $i$-th arm is characterized by its fixed, unknown reward probability $p_i \in [0, 1]$. At each time step $t$, the player selects an arm $j$; with probability $p_j$ it gets reward $r_t = 1$, otherwise $r_t = 0$.

At any point $T$ during the game, the performance of the MAB strategy is measured by its cumulative reward $\sum_{t=1}^{T} r_t$. An equivalent criterion, more amenable to theoretical analysis, is the so-called regret of the strategy, defined as the difference between its performance and the best possible performance. Clearly, the best possible performance is achieved by playing at each time step the (unknown) best arm, i.e., the arm with maximal probability $p^*$ of getting a reward. Hence, the regret of the strategy after $T$ time steps is:

$$L(T) = T \cdot p^* - \sum_{t=1}^{T} r_t \quad (5.6)$$

Classically, it is assumed that arms are independent from each other; and that the rewards associated to each arm are independently and identically distributed (i.i.d.). Under these assumptions, it can be shown that the minimal regret increases logarithmically with time ($L(T) = O(\log(T))$) [Lai and Robbins, 1985]. One of the solutions proposed for the MAB problem, the so-called Upper Confidence Bound (UCB) algorithm [Auer et al., 2002], achieves this optimal regret rate through an Exploration versus Exploitation-based criterion. Formally, the $i$-th arm is associated to its empirical quality estimate $\hat{q}_i$ (the average of the rewards $r_i$ obtained up to the given time instant); and to a confidence interval that depends on the number of times $n_i$ the $i$-th arm has been tried.

The UCB algorithm deterministically selects, at each time step, the arm with best upper bound of the confidence interval defined in Equation 5.7.
Select $\arg \max_{i=1 \ldots K} \left( \hat{q}_{i,t} + \sqrt{\frac{2 \log \sum_{k=1}^K n_{k,t}}{n_{i,t}}} \right)$ \hspace{1cm} (5.7)

with $n_{i,t+1} = n_{i,t} + 1$ \hspace{1cm} (number of times used) \hspace{1cm} (5.8)

and $\hat{q}_{i,t+1} = \left( \frac{\sum_{i=0}^{t} r_{i,t}}{n_{i,t+1}} \right)$ \hspace{1cm} (empirical quality estimate) \hspace{1cm} (5.9)

Clearly, the left term in Equation 5.7 favors the arm with best empirical quality (exploitation), while the right term promotes the trial of the other arms (exploration). The UCB algorithm works thus by choosing mostly the arm that can possibly give the best reward, while giving a chance from time to time to infrequently tried arms. Its efficiency comes from the fact that, although every arm is selected an infinite number of times, the lapse of time between two selections of an under-optimal arm increases exponentially with respect to the number of time steps. For this reason, the UCB is frequently phrased as “Be optimistic in front of the Unknown”.

However, the mentioned optimality of the UCB algorithm with respect to the balance between the Exploration and Exploitation terms is ensured only in the original MAB context, in which rewards are Boolean. In the AOS context, the rewards are usually in between some real-value interval, depending on the Credit Assignment scheme being employed, and this “breaks” this balance. In order to correct it, a Scaling factor was introduced by us into the original formula, being represented by the $C$ term in Equation 5.10 [Da Costa et al., 2008]. Besides, in order to avoid storing all rewards received by each operator up to time $t$, the averaging procedure of the empirical quality estimate presented in Equation 5.9 can be re-written as shown in Equation 5.12.

Select $\arg \max_{i=1 \ldots K} \left( \hat{q}_{i,t} + C \cdot \sqrt{\frac{2 \log \sum_{k=1}^K n_{k,t}}{n_{i,t}}} \right)$ \hspace{1cm} (5.10)

with $n_{i,t+1} = n_{i,t} + 1$ \hspace{1cm} (number of times used) \hspace{1cm} (5.11)

and $\hat{q}_{i,t+1} = \left( \frac{(n_{i,t}-1) \cdot \hat{q}_{i,t} + r_{i,t}}{n_{i,t}} \right)$ \hspace{1cm} (empirical quality estimate) \hspace{1cm} (5.12)

A complete representation of this Operator Selection technique is presented in the form of a pseudo-algorithm in Algorithm 5.4.

It must be noted that the MAB paradigm differs in two aspects from the mainstream framework concerned with learning optimal strategies, namely Reinforcement Learning (RL). On the one hand, MAB aims to select the best action, whereas RL is concerned with finding the best sequence of actions. On the other hand, while RL is concerned with learning the optimal sequence, it does not pay attention to the rewards it gets during the training phase. Quite the contrary, MAB simultaneously wants to learn the best action, and to optimize the cumulative reward it gets during learning. Clearly, the latter objective is the only one relevant in the context of AOS: the goal is to continuously learn which operator should be applied while maximizing the fitness improvement in the course of evolution.
5.3 Contributions to Operator Selection

Algorithm 5.4: Operator Selection: Multi-Armed Bandit (K, C)

1: for \( i = 1 \) to \( K \) do
2: \( n_i \leftarrow 0 \) \hspace{1cm} // number of operator trials
3: \( \hat{q}_i \leftarrow 0.0 \) \hspace{1cm} // empirical quality estimate
4: end for
5: while NotTerminated do
6: if one or more operators not applied yet then
7: \( op \leftarrow \) uniformly selected between the operators not applied
8: else
9: \( op \leftarrow \arg \max_{i=1...K} (\hat{q}_i + C \sqrt{\frac{2 \log (\sum_{j=1}^K n_j)}{n_i}}) \)
10: end if
11: Operator \( op \) is applied, impacting the search progress somehow
12: \( r_{op} \leftarrow \) CreditAssignment.GetReward(\( op \))
13: \( n_{op} \leftarrow n_{op} + 1 \)
14: \( \hat{q}_{op} \leftarrow \frac{(n_{op}-1)\hat{q}_{op} + r_{op}}{n_{op}} \)
15: end while

Although we acknowledge that MAB is the name given to the problem, for the sake of convenience, in the remainder of this manuscript we will refer to the UCB selection strategy with scaling factor \( C \) (as shown in Equation 5.10) as the MAB technique. As this technique is the basis of all the Operator Selection methods developed during this thesis, it will be always used as baseline for empirical comparison, being combined with one of the Credit Assignment schemes that use the raw values of fitness improvements to measure the impact of operator applications, namely, the absolute and the normalized versions of the Instantaneous, Average, and Extreme schemes.

5.3.2 Dynamic Multi-Armed Bandit

As previously discussed, the MAB algorithm has been designed in order to minimize the regret, i.e., the loss compared to the cumulative reward obtained by the (unknown) optimal strategy [Auer et al., 2002]. This makes it compulsory to determine the best arm (say with reward \( r \)): in case the algorithm settles on the second best arm (say with reward \( r' \)), it incurs some loss \( r - r' \) at each time step, and its regret increases linearly with the number of time steps. In the meanwhile, unpromising arms are tried exponentially less and less; since the reward distribution is assumed to be stationary, the chances of mistaking the best arm for an unpromising one decreases exponentially with the number of trials. The main rationale behind the MAB exploration (trying other arms than the one with best empirical \( \hat{q} \)) is thus to determine the best arm among the most promising ones.

AOS faces quite different priorities. The main need for exploration comes from the dynamics of the environment: one cannot assume the reward distribution to be stationary, the quality of any operator is bound to vary along evolution (see in Section 6.4.2, e.g., the
variation in the quality of some mutation operators on the OneMax problem). Henceforth, mistaking the best and second best operator has moderate consequences as the loss is small (provided $r$ and $r'$ are sufficiently close) compared to the cost of exploration. The point thus becomes to identify as fast as possible a sufficiently good operator.

Note that if the reward distribution is not stationary, the MAB regret cannot but linearly increase with the number of time steps in the worst case. The worst case is when the reward distribution of the supposedly best arm does not change, whereas a previously bad arm covertly becomes the best one. The only way to detect such a worst-case change would be to try all arms sufficiently often, i.e., to define a minimal selection probability, along the same lines as the Probability Matching (PM) Operator Selection technique. In the evolutionary framework, however, such a worst-case change scenario is unlikely to occur.

On the one hand, the average reward of every operator tends to decrease as evolution goes on (diminishing returns): in the One-Max problem, for instance, the best mutation operator is the 5-bit mutation when the population is far away from the optimum; but the reward of the 5-bit mutation decreases as the population goes to more fit regions, and at some point the 3-bit mutation operator catches up (more details on this can be found in Section 6.4). This suggests that when a good operator has been identified, there is no need for exploration as long as this operator remains sufficiently good.

On the other hand, even without employing a minimal selection probability, and with the lapse of time between two exploration trials increasing exponentially as the search goes on, there is still some exploration being done by the original MAB algorithm. It should thus be able to handle the AOS dynamic scenarios to some extent, although not having been devised to do so. The problem, however, lies in the update rule of the MAB empirical quality estimate $\hat{q}_i$: the simple averaging formula presented in Equation 5.12 can be re-written as:

$$\hat{q}_{i,t+1} = \hat{q}_{i,t} + \frac{1}{n_{i,t}} \cdot (r_{i,t} - \hat{q}_{i,t}) \quad (5.13)$$

From Equation 5.13, it becomes clear that the weight of the reward received by operator $i$ from the Credit Assignment under employment at time $t$ ($r_{i,t}$) is inversely proportional, in the update of the operator empirical quality estimate $\hat{q}_{i,t}$, to the number of times operator $i$ was applied ($n_{i,t}$). Therefore, in case the current best operator has been selected $n_i$ times and its reward falls down by $\delta$, it will need roughly $n_i \cdot \delta/\varepsilon$ trials before recovering an accurate quality estimate of operator $i$ up to precision $\varepsilon$. In other words, the longer is the steady-state of the quality of the best operator, the longer it will take for the MAB process to correct its empirical knowledge in case the situation changes. This inertia is what significantly degrades the performance of the original MAB algorithm on the dynamic AOS context, as assessed in the experimental comparisons that will be presented in Chapter 6.

The above remarks motivated the proposal of the Dynamic Multi-Armed Bandit (DMAB) approach: the original MAB algorithm is coupled with a statistical test, the Page-Hinkley (PH) change-point detection test [Page, 1954]. Briefly, upon the detection
5.3 Contributions to Operator Selection

of a change in the reward distribution, the MAB process is restarted from scratch. After describing the PH test, more details about the DMAB will be given in the following.

**Page-Hinkley Change-Point Detection Test**

Given a series of observations \((r_1, \ldots, r_\ell)\), a frequently asked question is whether these observations can be attributed to a same statistical law (Null hypothesis), or if some change in the law underlying the observations has occurred at some point (Change hypothesis). A standard test for the change hypothesis is the Page-Hinkley (PH) test [Page, 1954], which can be formally described as follows.

Let \(\bar{r}_\ell\) denote the average of \((r_1, \ldots, r_\ell)\) and let \(e_\ell\) denote the difference \((r_\ell - \bar{r}_\ell + \delta)\), where \(\delta\) is a tolerance parameter [Page, 1954]. The PH test considers the random variable \(m_t\) defined as the sum of \((e_1, \ldots, e_t)\). The maximum value \(M_t\) of the \(|m_t|\) values for \((\ell = 1, \ldots, t)\) is also computed and the difference between \(M_t\) and \(|m_t|\) is monitored. When this difference is greater than some user-specified threshold \(\gamma\), the PH test is triggered, i.e., it is considered that the Change hypothesis holds. This can be formally written as follows:

\[
\begin{align*}
\bar{r}_\ell &= \frac{1}{\ell} \sum_{i=1}^{\ell} r_i \\
M_t &= \sum_{\ell=1}^{t} (r_\ell - \bar{r}_\ell + \delta) \\
M_t &= \arg\max_{\ell=1, \ldots, t} \{|m_\ell|\} \\
PH_t &= M_t - |m_t| \\
\text{Return } (PH_t > \gamma)
\end{align*}
\] (5.14)

The PH test involves two hyper-parameters. Parameter \(\gamma\) controls the trade-off between false alarms and unnoticed changes; and parameter \(\delta\) enforces the robustness of the test when dealing with slowly varying environments. Following early experiments [Da Costa et al., 2008], the latter has been kept fixed to 0.15; while the former is a hyper-parameter that needs to be defined by the user.

**MAB + PH = DMAB**

The hybridization of the original MAB algorithm (UCB with Scaling factor) with the PH statistical test results in the so-called Dynamic Multi-Armed Bandit (DMAB). It maintains four indicators for each arm \(i\): from the MAB side, the number \(n_{i,t}\) of times this arm has been tried up to time \(t\), and its current (average) empirical quality estimate \(\hat{q}_{i,t}\); from the PH test side, there are the average and the maximum deviation measures \(m_i\) and \(M_i\). In addition to these indicators, the DMAB also consequently inherits the hyper-parameters of both components, which need to be tuned beforehand by the user, namely, the MAB scaling factor \(C\) and the PH change-detection threshold \(\gamma\). A complete and detailed representation of DMAB, with its indicators and hyper-parameters, is presented in Algorithm 5.5.

Note that the DMAB combination was firstly proposed in another dynamic context by [Hartland et al., 2007], being originally applied in AOS by us in [Da Costa et al., 2008].
Chapter 5. Contributions to Adaptive Operator Selection

Algorithm 5.5: Operator Selection: Dynamic MAB ($K, C, \gamma, \delta = 0.15$)

1: for $i = 1$ to $K$ do
2:   $n_i \leftarrow 0$ // number of operator trials (MAB)
3:   $\hat{q}_i \leftarrow 0.0$ // empirical quality estimate (MAB)
4:   $m_i \leftarrow M_i \leftarrow 0.0$ // cumulative and max. cumulative difference (PH)
5: end for
6: while NotTerminated do
7:   if one or more operators not applied yet then
8:      $op \leftarrow$ uniformly selected between the operators not applied
9:   else
10:      $op \leftarrow \arg \max_{i=1,...,K} \left( \hat{q}_i + C \sqrt{\frac{2 \log(\sum_{j=1}^{K} n_j)}{n_i}} \right)$
11:   end if
12:   Operator $op$ is applied, impacting the search progress somehow
13:   $r_{op} \leftarrow \text{CreditAssignment.GetReward}(op)$
14:   $n_{op} \leftarrow n_{op} + 1$
15:   $\hat{q}_{op} \leftarrow \frac{(n_{op}-1)\hat{q}_{op} + r_{op}}{n_{op}}$ // identical to the MAB algorithm up to here
16:   $m_{op} \leftarrow m_{op} + (r_{op} - \hat{q}_{op} + \delta)$ // then the PH change-detection test is performed
17:   if $|m_{op}| > M_{op}$ then
18:      $M_{op} \leftarrow |m_{op}|$
19:   else if $M_{op} - |m_{op}| > \gamma$ then
20:      Restart MAB and PH variables ($n, \hat{q}, m, M$)
21:   end if
22: end while

In the latter paper, the absolute instantaneous value of rewards given by some artificial scenarios (described in Sections 5.4.1 and 5.4.2) was used as the Credit Assignment scheme, in order to study the Operator Selection techniques independently. Being fed by the Extreme value of fitness improvements (and also by the other Credit Assignment schemes based on fitness improvements, for the sake of empirical comparison), it was later assessed on some EA binary benchmark problems [Fialho et al., 2008; Fialho et al., 2009a; Fialho et al., 2009b]. It was also evaluated on some SAT instances [Maturana et al., 2009a], within an aggregation of fitness and diversity, the Compass (described in Section 4.3.4), being used as Credit Assignment. All these experiments will be reminded in detail in Chapter 6.

5.3.3 Sliding Multi-Armed Bandit

Although showing to be very efficient, the DMAB Operator Selection technique presents two main weaknesses. On the one hand, the change-point detection test is triggered only in the case of an abrupt change, whereas the reward of an operator usually decreases gradually: this makes it very difficult to calibrate this test (namely, its change-detection
5.3 Contributions to Operator Selection

threshold $\gamma$). On the other hand, upon triggering the test, the whole memory of the MAB process is lost, and the exploration of the operators must start anew.

These remarks motivated the introduction of a new bandit-based Operator Selection technique, referred to as the Sliding Multi-Armed Bandit (SLMAB) [Fialho et al., 2010a]. The underlying idea of this method is to be able to gracefully follow the dynamics of the AOS scenario, without needing the very sensitive and somehow controversial (although efficient when correctly tuned) restart mechanism employed by DMAB.

Several heuristics have been proposed to update statistical estimates in a non-stationary context. The most natural heuristic is the so-called relaxation update rule, used by the PM and AP Operator Selection schemes, in which the weight of the instant reward $r_t$ on the update of the empirical quality estimate is defined by some constant learning rate $\alpha$ ($0 < \alpha \leq 1$):

$$\hat{q}_{i,t+1} = (1 - \alpha) \cdot \hat{q}_{i,t} + \alpha \cdot r_{i,t}$$

(5.15)

The difficulties with the above rule are that, besides introducing the extra hyper-parameter $\alpha$ to be tuned by the user, it defines a constant weight for the instant reward $r_{i,t}$, regardless of how frequently the $i$-th operator has been applied in the last time steps. In the AOS framework, however, different operators are applied with different frequencies; if an operator has not been applied for a long time, the weight of the instant reward it received should be higher, everything else being equal, in order to enable a more rapid adjustment of its $\hat{q}_{i,t}$.

The update rule must thus take into account the number of time steps elapsed since the previous time step $t_i$ in which the $i$-th operator has been applied. Finally, in order to preserve the MAB trade-off between exploration and exploitation, one must also maintain the $n_{i,t}$ counters reflecting the frequency of application of operators up to time step $t$.

Considering a window of size $W$, the proposed update of the sliding exploitation and exploration terms (respectively, $\hat{q}$ and $n$), which is performed every time operator $i$ is applied, is defined as:

$$\begin{align*}
\hat{q}_{i,t+1} &= \hat{q}_{i,t} \cdot \frac{W}{W + (t - t_i)} + r_{i,t} \cdot \frac{1}{n_{i,t} + 1} \\
n_{i,t+1} &= n_{i,t} \cdot \left( \frac{W}{W + (t - t_i)} + \frac{1}{n_{i,t} + 1} \right)
\end{align*}$$

(5.16)

The above update rule is designed in such a way that, if an operator is applied with frequency $W/n_t$, then $n_t$ is constant. The rationale for this update scheme can be explained as follows.

If an operator is performing well and is almost always applied, counter $n_{i,t}$ rapidly increases up to $W$ and sticks to this value, while its empirical quality estimate $\hat{q}_{i,t}$ accurately reflects the reward expectation for the current stage of the search. The main difference compared to the MAB and DMAB settings is that $n_{i,t}$ is upper bounded by $W$. Equivalently, the inertia of the reward estimate is bounded: the weight of the instant reward cannot be less than $1/W$.

Oppositely, if an operator is rarely applied, its $\hat{q}_{i,t}$ can be seen as an outdated estimation of the actual reward expectation. On the other hand, the fact that the operator is
rarely applied means that its empirical quality estimate is lower than that of the current best operator. With the averaging update rule employed by the original MAB scheme, presented in Equation 5.12, it would take a long time to correct the empirical quality estimate of this operator in case it had become the new best one. With the proposed sliding update rule, however, this outdated estimation of the operator quality is more efficiently corrected: if the operator has not been tried in the previous $W$ time steps, $n_{i,t}$ is low, consequently the weight given to the instant reward is high in the update formula, thus rapidly shifting the empirical quality estimate towards its actual value.

Besides the scaling factor $C$, that is needed by all bandit-based Operator Selection mechanisms, the other hyper-parameter that needs to be defined in SLMAB is the window size $W$, used in the proposed window-based relaxation update mechanism. But, as most Credit Assignment schemes found in the literature, including the schemes proposed in this thesis, rely on windowing the operator reward distribution, this latter hyper-parameter can be said to be parameterized “for free” (by using the same window size $W$ employed by the Credit Assignment scheme being used). A complete presentation of the SLMAB in the form of a pseudo-algorithm is presented in Algorithm 5.6.

<table>
<thead>
<tr>
<th>Algorithm 5.6: Operator Selection: Sliding Multi-Armed Bandit ($K, C, W$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: $times \leftarrow 0$ // number of total time steps</td>
</tr>
<tr>
<td>2: for $i = 1$ to $K$ do</td>
</tr>
<tr>
<td>3: $n_i \leftarrow 0$ // number of operator trials</td>
</tr>
<tr>
<td>4: $\hat{q}_i \leftarrow 0.0$ // empirical quality estimate</td>
</tr>
<tr>
<td>5: $last_i \leftarrow 0$ // last time it was applied</td>
</tr>
<tr>
<td>6: end for</td>
</tr>
<tr>
<td>7: while NotTerminated do</td>
</tr>
<tr>
<td>8: if one or more operators not applied yet then</td>
</tr>
<tr>
<td>9: $op \leftarrow$ uniformly selected between the operators not applied</td>
</tr>
<tr>
<td>10: else</td>
</tr>
<tr>
<td>11: $op \leftarrow \arg \max_i \left( \hat{q}_i + C \cdot \sqrt{\frac{2 \log(K)}{n_i} \cdot \frac{1}{n_i}} \right)$</td>
</tr>
<tr>
<td>12: end if</td>
</tr>
<tr>
<td>13: Operator $op$ is applied, impacting the search progress somehow</td>
</tr>
<tr>
<td>14: $r_{op} \leftarrow$ CreditAssignment.GetReward($op$)</td>
</tr>
<tr>
<td>15: $\hat{q}<em>{op} \leftarrow \hat{q}</em>{op} \cdot \left( \frac{W}{W + (times - last_{op})} \right) + r_{op} \cdot \left( \frac{1}{n_{op} + 1} \right)$</td>
</tr>
<tr>
<td>16: $n_{op} \leftarrow n_{op} \cdot \left( \frac{W}{W + (times - last_{op})} + \frac{1}{n_{op} + 1} \right)$</td>
</tr>
<tr>
<td>17: $last_{op} \leftarrow times$</td>
</tr>
<tr>
<td>18: $times \leftarrow times + 1$</td>
</tr>
<tr>
<td>19: end while</td>
</tr>
</tbody>
</table>

The SLMAB Operator Selection technique, combined with the absolute Instantaneous, Average and Extreme Credit Assignment schemes, was assessed on some artificial scenarios, described in Section 5.4, being also used to select between some mutation and
crossover operators within a real EA applied to the Royal Road benchmark problem [Fialho et al., 2010a]. All these experiments will be detailed in Chapter 6.

5.3.4 Rank-based Multi-Armed Bandit

The main criticism with respect to the bandit-based AOS schemes described up to now is related to the high sensitivity of their hyper-parameters, what was an important motivation factor for most of the further developments presented throughout Sections 5.2 and 5.3. From the Operator Selection point-of-view, the SLMAB represented a further step towards more robust schemes, by eliminating one of the two very sensitive hyper-parameters present in the DMAB approach (the Page-Hinkley change-detection threshold $\gamma$), while still being able to achieve equivalent performance in following the AOS dynamics on some artificial benchmark functions [Fialho et al., 2010a]. But, even for the SLMAB, there is still the need to tune the scaling factor $C$, which is indeed a very sensitive hyper-parameter common to all bandit-based Operator Selection approaches previously presented. The difficulty for tuning this parameter comes from the Credit Assignment scheme being employed, as follows.

By the time the DMAB and SLMAB techniques were proposed, the Credit Assignment schemes under consideration were the basic Instantaneous, Average (Section 5.2.1), and Extreme ones (Section 5.2.2), as well as their Normalized versions (Section 5.2.3). All these schemes assign credit based directly on some statistics over the raw values of the fitness improvements. Whenever these raw values are used, the tuning of the AOS hyper-parameters tends to be highly problem-dependent, as the range of fitness values varies widely from one problem to another, as well as in the course of an optimization run. Hence, the scaling factor $C$, which has as original role to tune the balance between the exploitation and exploration terms of the UCB formula (Equation 5.10), also needs to play a radically different role, that of accounting for the scale of the rewards received. This double role is the reason why $C$ shows to be such a sensitive hyper-parameter.

These issues motivated the proposal of the rank-based and further comparison-based Credit Assignment schemes, presented in Sections 5.2.4 and 5.2.5, respectively. However, the direct combination of these schemes with the previously described bandit-based Operator Selection techniques showed a rather poor performance after some preliminary experiments. The reason for this is that the AUC and the SR indicators already provide, on the Credit Assignment side, an empirical statistic over the last $W$ offspring generated; while the MAB techniques do another aggregation of rewards in the Operator Selection side (the $\hat{q}$ in Equation 5.10) – the two layers of statistics were somehow diluting the interesting characteristics of the proposed performance measurements. Therefore, as the outputs of the AUC and SR indicators already reflect accurate and up-to-date performance measures of one operator with respect to all others, they can be used directly as the exploitation term in the MAB formula, i.e., $\hat{q}_{i,t} = AU C_{i,t}$ or $SR_{i,t}$, depending on the scheme being used.

This simple adaptation of the MAB scheme brings another very important benefit for Operator Selection. As the mentioned rank/comparison-based Credit Assignment schemes maintain just one sliding window for the rewards received by all operators, the inclusion
of a new reward in the sliding window, achieved by a given operator, affects the quality estimates of all the other operators. Consequently, the AOS dynamics are already handled on the Credit Assignment side in a transparent way, without needing an external observer, as the change-detection test in the DMAB technique, or a relaxation update rule, as the one employed in the SLMAB scheme.

Finally, in order to ensure a minimal level of exploration, the MAB term $n$ is modified to reflect the number of times each operator appears in the sliding window. A complete representation of this simplified version of the MAB algorithm, specially adapted to be used with the rank/comparison-based Credit Assignment schemes, thus referred to as the Rank-based Multi-Armed Bandit (RMAB), is presented in the form of a pseudo-algorithm in Algorithm 5.7.

**Algorithm 5.7: Operator Selection: Rank-based Multi-Armed Bandit ($K, C$)**

1: for $i = 1$ to $K$
2:   $n_i \leftarrow 0$ // number of times operator $i$ appears in the current credit window
3:   $\hat{q}_i \leftarrow 0.0$ // empirical quality estimate = normalized CreditAssignment output
4: end for
5: while NotTerminated do
6:   if one or more operators not applied yet then
7:     $op \leftarrow$ uniformly selected between the operators not applied
8:   else
9:     $op \leftarrow \arg \max_i \left( \hat{q}_i + C \cdot \sqrt{\frac{2 \log(K)}{n_i}} \right)$
10: end if
11: Operator $op$ is applied, impacting the search progress somehow
12: for $i = 1$ to $K$ // the application of one operator might affect the others
13:   $\hat{q}_i \leftarrow$ CreditAssignment.GetReward($i$)
14:   $n_i \leftarrow$ CreditAssignment.GetTimes(creditWindow, $i$)
15: end for
16: end while

In order to evaluate the efficiency and robustness of the AOS techniques derived from the combination of the RMAB Operator Selection scheme with the rank/comparison-based Credit Assignment schemes, they were firstly assessed within a GA applied to the OneMax problem and to other three fitness functions defined by monotonous transformations over this problem [Fialho et al., 2010c]. Later on, their performances were also evaluated on a set of 24 single-objective continuous problems [Fialho et al., 2010b; Fialho and Ros, 2010], selecting between different mutation strategies within a DE algorithm. As expected, these combinations showed to be very robust with respect to their hyper-parameters, namely, the scaling factor $C$, the decay factor $D$, and the window size $W$; while achieving the same level of state-of-the-art performance of the previously presented (efficient but very problem-dependent) approaches. The experimental results on the OneMax and on the continuous problem will be detailed, respectively, in Sections 6.4 and 6.6.
5.4 Contributions to Empirical Assessment

By the time this manuscript is being written, the use of RMAB as Operator Selection, with the Area-Under-Curve (AUC) Credit Assignment scheme, is our final recommended choice in case one wants to use the AOS paradigm on his own optimization algorithm/problem. The reason for this choice will be extensively discussed and justified in Section 6.8, based on the evidences brought by the comprehensive empirical analysis that will be presented in Chapter 6.

5.4 Contributions to Empirical Assessment

While developing the AOS schemes presented in this Chapter, some artificial benchmark problems were proposed to analyze different aspects of their behavior in a controlled environment. All these scenarios, that will be now described in turn, were used in part of the empirical analysis of the AOS schemes, which will be presented in Section 6.3.

Based on the Uniform artificial scenario, proposed in [Thierens, 2005] and described in Section 5.4.1 for the sake of self-containedness, we have introduced two other artificial scenarios. Referred to as the Boolean and the Outlier scenarios, they will be presented in Section 5.4.2. They involve the same switches between five operators, but with rewards coming from different distributions and with different probabilities.

More recently, we have extended this set of artificial benchmarks by proposing a new family of problems, referred to as the Two-Values (TV) benchmarks [Fialho et al., 2010a], which can be used to simulate different situations with respect to the mean and variance of rewards given by two artificial operators. The TV benchmarks will be described in detail in Section 5.4.3.

5.4.1 Basic Artificial Scenario: Uniform

The Uniform artificial benchmark, proposed in [Thierens, 2005], involves a set of 5 operators, in which the reward distribution associated to each operator is constant during an epoch (ΔT times steps). During every epoch, the operator reward is uniformly drawn from an interval, as follows: \{4, 6\} for the current best operator, \{3, 5\} for the second best, \{2, 4\} for the third, \{1, 3\} for the fourth, and \{0, 2\} for the worst operator.

The reward distributions associated to all operators are permuted at the end of every epoch, using pre-defined permutations to decrease the experimental noise, defined in [Thierens, 2005] as follows: 41203 01234 24301 12043 41230 31420 04213 23104 14302 40213. More precisely, the best operator in the first epoch is the \(op_4\), which becomes the worst one in the second epoch. The best operator in the second epoch is \(op_0\), which was the fourth one in the first epoch.

It is true to say that these abrupt changes in the quality of the operators are very unrealistic; they are used here as a kind of extreme scenario with respect to AOS dynamics. Different epoch lengths are considered, e.g., set to \(ΔT = 50\) for fast dynamics and \(ΔT = 200\) for slow ones. The performance associated to an AOS scheme is the cumulative reward obtained during this sequence of 10 epochs. As the reward expectation of the best operator is 5, the maximal cumulative reward is 2,500 in the fast case (5 \(*\) 10 \(*\) 50) and 10,000 in the slower one (5 \(*\) 10 \(*\) 200).
5.4.2 Boolean and Outlier Scenarios

Within the Uniform benchmark, an operator always gets a reward that is positive, while also being informative, i.e., it indicates (to some extent) which is the best operator, possibly mistaking just with the second best, as the intervals of their reward distributions overlap. In a real evolutionary context, however, the AOS task might be much more challenging. For instance, the probability of getting some useful information about the quality of the operator might be smaller than that, up to the situation in which no information whatsoever is provided to the AOS (specially true when the search is getting closer to the optimum); or, even if some rewards are frequently provided, the information gathered from them might not be so useful in order to efficiently differ between the available operators. Based on these two difficulties, we proposed [Da Costa et al., 2008] and further used [Fialho et al., 2010a] two variants of the Uniform benchmark, referred to as the Boolean and Outlier scenarios.

In the Boolean scenario, the best operator gets a reward of 10 with probability 50%, and 0 otherwise; the second best gets the same reward of 10 but with probability 40% and 0 otherwise; and so forth, until the worst operator, getting a reward of 10 with probability 10% and 0 otherwise. In this scenario, the difference between the operators is the probability of getting a non-null reward; the reward takes the same value in all cases. In particular, the best operator has the same reward expectation as in the Uniform scenario, though with a much higher variance.

Quite the contrary, in the Outlier scenario, all operators get a non-null reward with the same probability (10%). The difference lies in the reward value, set to 50 for the best operator, 40 for the second best and so forth, up to 10 to the worst operator. While the reward expectation is still the same as in the Uniform benchmark, the AOS is provided with much less information (only 10% of the trials produce some information), and the reward variance is much higher than in the Boolean scenario.

Summarizing, thus, the probability of getting some information is high (for the best operator) in the Boolean benchmark, while being low (for all operators) in the Outlier benchmark. But the Boolean scenario typically does not provide useful information (all rewards have the same values, only the probabilities differ), while the Outlier scenario involves very informative but rare rewards. Both use the same sequence of switches between reward distributions after every epoch, and the reward expectation for each operator is also the same as in the Uniform case.

5.4.3 Two-Value Scenarios

As discussed in the previous Section, any AOS is usually provided with some (more or less) informative results (the reward amount, everything else being equal); and it is more or less likely to be provided with any information at all. Typically, the MAB process is well equipped to deal with Boolean-like settings, where operators (arms) get the same reward in case of success and only the probability of success differs.

Along these lines, a framework for AOS benchmarks, referred to as Two-Values (TV) benchmarks, was proposed in [Fialho et al., 2010a], enabling a more precise control of
5.4 Contributions to Empirical Assessment

the two issues previously discussed. Briefly, every operator is assumed to get one out of two reward values, a small value noted $r$ and a large one noted $R$. Within these two parameters, it is possible to control the informativeness of the reward distribution, defined by the ratio $R/r$; while the third parameter, $p$, defines the probability of getting reward $R$; and $(1 - p)$ is the probability of getting $r$. Needless to say, the mean and the variance of the rewards received are also intrinsically managed by these parameters.

Formally, the reward distribution specified from the triple $(p, r, R)$ is defined as:

$$ TV(p, r, R) = \begin{cases} R & \text{with probability } p \\ r & \text{with probability } 1 - p \end{cases} $$

with expectation and variance respectively noted $E(p, r, R)$ and $V(p, r, R)$:

$$ E(p, r, R) = p \cdot R + (1 - p) \cdot r $$

$$ V(p, r, R) = p \cdot (1 - p) \cdot (R - r)^2 $$

It is clear that only the ratio $R/r$ impacts the results; thus, in the remainder of this manuscript, $r$ will be set to 1 and omitted in the notations for the sake of simplicity – a reward distribution will be noted $TV(p, R)$ instead of $TV(p, 1, R)$. Furthermore, a scenario involving $TV(p_1, R_1)$ and $TV(p_2, R_2)$ will be denoted by $ART(p_1, R_1, p_2, R_2)$. The respective roles of $p$ and $R$ are exemplified in Figure 5.4, displaying two samples of size 100 of distributions with the same expectation and high versus low variance.

![Figure 5.4: Two samples drawn from two $TV$ distributions with same expectation $E(.1, 10) = E(.9, .2) = 1.9$; the distribution on the left picture presenting a much higher variance ($V(.1, 10) = 7.29$) then the one on the right ($V(.9, 2) = .09$).](image)

Such a general framework will help us into analyzing very different aspects of the behavior of the AOS schemes proposed. The main question will always be how agile a given AOS combination is into switching between two different situations, what will be analyzed under the light of different variants of this scenario in Section 6.4. But other low level details could also be analyzed, e.g., does the AOS (or the Credit Assignment scheme
it implements) tend to favor operators with a high variance instead of the ones with a low variance [Fialho et al., 2010a].

In the same way as for the previously presented scenarios, the exchanges between the reward distributions, done after each epoch of $\Delta T$ time steps, obey a fixed sequence in order to decrease the experimental noise: $01 \leftrightarrow 01 \leftrightarrow 10 \leftrightarrow 01 \leftrightarrow 10 \leftrightarrow 10 \leftrightarrow 01 \leftrightarrow 01 \leftrightarrow 10$.

5.5 Discussion

In order to be well-accepted by the EA community, the main requirement of a good AOS technique is, of course, to be able to efficiently automate the control of the operators to be applied, in a dynamic way, during the search process. But it also needs to be (i) easily implementable and, even more importantly, to be (ii) computationally cheap.

The first issue can be alleviated by making available implementations of the proposed schemes (e.g., as open source libraries). The algorithms implemented in this work were coded in ANSI-C++. The Evolving Objects (EO) library\footnote{EO: C++ library for coding EAs, freely available at http://eodev.sourceforge.net/ as of today.} [Keijzer et al., 2002] was used to assist the implementation of the underlying EAs, controlled by the proposed AOS methods. All this experimental framework is freely available under request by email. And indeed, it was already requested by some researchers, and we acknowledge its recent use by one of them [Verel et al., 2010] on the assessment of a newly proposed method.

A lot of care should be taken, however, with the second issue: the more computationally expensive the AOS technique, the smaller the margin of possible benefits it can bring to the underlying algorithm whose operators are being controlled by it. These benefits are usually computed in terms of computational time or effort to achieve a given solution. Indeed, in some domains, such as in numerical engineering, the fitness evaluation is very expensive, what makes the AOS computational cost negligible. But in combinatorial problems, such as the Boolean Satisfiability problems (see Section 6.5.2), the cost of evaluating the fitness of a given solution is almost zero. In this latter case, it might become impracticable to consult and update the AOS technique every time an operator is applied: a trade-off should thus be found between the AOS granularity (how frequent it is consulted and updated, e.g., once every generation instead of once every application) and the AOS dynamics accuracy (the more frequent it is updated, the more reliable will be its estimation with respect to the operators empirical performance).

A third issue worth discussing concerns a very common critic that prevents people from using AOS schemes on their own algorithms and problems: although being proposed to automate some user choices, these schemes have their own (hyper-)parameters that need to be tuned in order to achieve acceptable performance, e.g., the scaling factor $C$ and the window size $W$, respectively, common to all Operator Selection schemes and Credit Assignment mechanisms we have proposed. What should be argued in this case is that the AOS schemes automatically take care of many shallow parameters ($i.e.$, which operators should be considered by the EA, and at which rate each of the chosen ones should be applied), besides the main fact that they control the rates in a dynamic way, during the
5.5 Discussion

search process, while involving only a few general hyper-parameters (2 or 3, depending on the AOS combination being implemented).

A very common approach to solve this hyper-parameter setting issue is to use off-line parameter tuning methods to automatically set them. Indeed, this has been done for all the experiments that will be presented in Chapter 6, in order to compare the proposed and the baseline techniques at their peak performance. But off-line tuning is an expensive procedure, as discussed earlier in Section 3.3.2; so, ideally, a good AOS technique should also be robust with respect to its hyper-parameters, \( i.e., \) whenever a new problem needs to be solved, the tuning of the AOS hyper-parameters should be required as rarely as possible.

Although a good level of efficiency with respect to the AOS dynamics was attained very early in this thesis work, with the proposal of the Ex-DMAB technique (combining the Extreme \textit{Credit Assignment}, presented in Section 5.2.2, with the DMAB \textit{Operator Selection}, described in Section 5.3.2), the hyper-parameters of such AOS combination showed to be very sensitive, a good performance being shown just in case the hyper-parameters were tuned for every new problem. All the further developments proposed in this thesis, for both \textit{Credit Assignment} and \textit{Operator Selection} issues, aimed at smoothening this effect by creating more robust \( i.e., \) less problem-dependent) techniques, while maintaining the same level of performance.

A big progress was achieved on this direction by the recent proposal of the RMAB for \textit{Operator Selection} (Section 5.3.4), combined with any of the rank-based \textit{Credit Assignment} schemes described in Sections 5.2.4 and 5.2.5. The simple fact that rewards are computed based on ranks instead of raw values of fitness improvements already provides a much higher robustness to the AOS technique implementing it, guaranteeing invariance with respect to any linear scaling of the original fitness function. Furthermore, the use of ranks over the fitness values of the generated offspring, instead of ranks over the fitness improvements, provides to the AOS technique the property of being fully comparison-based, \( i.e., \) invariant with respect to all monotonous transformations over the same original fitness function. This robustness gain was confirmed by performing independent off-line tuning procedures over the AOS combination constituted by the Fitness-based Area-Under-Curve (F AUC) as \textit{Credit Assignment} with the RMAB as \textit{Operator Selection} over very different problems – the same or very similar hyper-parameter settings were found to be the best, while state-of-the-art (or very close to that) performance was achieved, as reviewed in the empirical comparisons that will be presented in Chapter 6.

The proposed rank-based AOS techniques are efficient and robust, but it is important to note that their current implementations can not achieve good performance in problems with high multi-modality \( i.e., \) several local optima), because the only action being rewarded is the progress with respect to the fitness. As exemplified in Section 4.5.2, in order to efficiently tackle multi-modal problems, the maintenance of some diversity in the population should also be rewarded somehow. Further developments should be done in this direction in the near future by, \( e.g., \) trying to provide the same level of robustness and invariance properties to the \textit{Credit Assignment} schemes proposed in [Maturana \textit{et al.}, 2010b], which aggregate both fitness and diversity measures to evaluate the operator performance. A different alternative, in the case of multi-modal problems,
would be to let the proposed AOS techniques as they are, i.e., rewarding just exploitation (fitness), but implementing efficient convergence detection mechanisms in the underlying EAs, in order to restart the search process in case it gets trapped in a local optimum, consequently giving more opportunities to the algorithm to possibly achieve better solutions within the same computational budget; such kind of approach, briefly discussed in Section 2.3.3, is very common in Evolution Strategies [Auger and Hansen, 2005].

Finally, concerning the artificial scenarios proposed for the empirical assessment of the AOS techniques, it is true that many different and more general settings could have been proposed, such as considering more than two TV operators, or more complex reward distributions (e.g., taking uniformly drawn values in two intervals centered in the \( r \) and \( R \) values, or smoother transitions between epochs). But the preliminary motivation was the analysis of the effect of both, the reward level of informativeness \( R/r \) and the probability \( p \) of receiving a positive reward, on the AOS performance, as will be shown in the empirical comparisons presented in Section 6.3. More complex and realistic reward landscapes shall be considered in further studies, according to the needs of the empirical analysis.
Chapter 6
Experimental Results

Contents

6.1 Introduction .............................................. 97
6.2 General Experimental Settings .......................... 98
  6.2.1 AOS Combinations and Respective Hyper-Parameters .... 98
  6.2.2 Off-line Tuning of Hyper-Parameters ................. 100
  6.2.3 Performance Indicators and Results Presentation ...... 102
6.3 On Artificial Scenarios ................................. 103
  6.3.1 Experimental Settings ............................... 104
  6.3.2 Results on Uniform, Boolean and Outlier Scenarios ... 104
  6.3.3 Results on ART Scenarios ......................... 120
  6.3.4 Discussion .................................... 130
6.4 On Boolean Benchmark Problems ....................... 131
  6.4.1 Experimental Settings ............................... 131
  6.4.2 The OneMax Problem ................................. 132
  6.4.3 The Long k-Path Problem ......................... 138
  6.4.4 The Royal Road Problem ............................ 143
  6.4.5 Discussion .................................... 147
6.5 Collaboration On Satisfiability Problems ............... 148
  6.5.1 Compass + Ex-DMAB = ExCoDyMAB .................. 148
  6.5.2 Boolean Satisfiability Problems .................... 149
  6.5.3 Experimental Settings ............................... 150
  6.5.4 Architecture definition and tuning of hyper-parameters .... 151
  6.5.5 Empirical Results ................................ 152
  6.5.6 Discussion .................................... 154
6.6 On Continuous Benchmark Problems ..................... 156
  6.6.1 Black-Box Optimization Benchmarking .............. 156
  6.6.2 Experimental Settings ............................... 157
Chapter 6. Experimental Results

6.6.3 The PM-AdapSS-DE Method ........................................ 159
6.6.4 Empirical Results ....................................................... 160
6.6.5 Discussion ............................................................... 166

6.7 Hyper-Parameters Analysis .......................................... 168
   6.7.1 On the Sensitivity of the Hyper-Parameters ..................... 169
   6.7.2 On the Robustness of the Hyper-Parameters ..................... 174

6.8 General Discussion ...................................................... 179
6.1 Introduction

In this Chapter, we present some empirical evaluations of the proposed AOS contributions. The methods are compared with one another, and with other baseline approaches, on diverse benchmark scenarios. Besides the performance evaluation, their sensitivity and robustness in relation to their hyper-parameters is also analyzed.

6.1 Introduction

From the standard MAB Operator Selection technique with a Credit Assignment based on fitness improvements, up to the latest RMAB that is rewarded based on ranks, several contributions for Adaptive Operator Selection (AOS) have been proposed in Chapter 5. In this Chapter, a comprehensive empirical analysis for each of the AOS methods, resulting from the combination of the proposed Operator Selection mechanisms and Credit Assignment schemes, will be presented.

Their performance will be assessed and compared in diverse benchmark scenarios, with different characteristics and levels of complexity. Firstly (Section 6.3), the empirical analysis on the artificial scenarios will consider the selection between operators whose rewards come from different pre-defined artificial distributions that are deterministically changed after every $\Delta T$ iterations. Then (Section 6.4), experiments on some boolean EA benchmark problems (OneMax, Long $k$-Path and Royal Road) will be analyzed: in these cases, the AOS schemes are applied to a Genetic Algorithm, automatically selecting between actual mutation and crossover operators, with the rewards coming from the progress attained by the search process on the considered fitness landscapes. Additionally (Section 6.5), the results obtained in the scope of a collaboration with Université d’Angers will be reminded: in this scenario, only the Dynamic Multi-Armed Bandit (DMAB) Operator Selection technique (Section 5.3.2) is evaluated, combined with the Compass Credit Assignment method (Section 4.3.4). The resulting AOS combination is used by a GA to autonomously select between some crossover, mutation and local search operators, in the context of Boolean Satisfiability (SAT) problems. In the last set of experiments (Section 6.6), the AOS schemes are used within a Differential Evolution algorithm in order to control which of the available mutation strategies should be applied; the results will be assessed in a big set of single-objective continuous benchmark functions.

Prior to all these experiments, a preliminary off-line tuning of the hyper-parameters was done for each AOS technique. The details about this procedure will be presented in Section 6.2, together with some general experimental settings. Besides, the specific experimental settings for each scenario will also be detailed in the respective Sections, before the presentation of the results. In addition to the performance, the sensitivity of each hyper-parameter and the robustness of the AOS techniques with respect to their hyper-parameters will also be analyzed in Section 6.7. Finally, the conclusions and findings gathered from all these empirical data will be discussed in Section 6.8.
6.2 General Experimental Settings

The different AOS combinations proposed have some hyper-parameters that need to be set, as discussed throughout Chapter 5. On each of the considered benchmark scenarios, an off-line tuning procedure was preliminarily performed for the hyper-parameters of each technique, in order to compare them at their peak performance. A summary of the AOS combinations and their respective hyper-parameters is presented in Section 6.2.1, while Section 6.2.2 describes the values explored for each hyper-parameter, and the off-line tuning procedure used. Finally, Section 6.2.3 overviews the different performance measures and displays that have been used in the diverse empirical comparisons that will be presented in the following.

6.2.1 AOS Combinations and Respective Hyper-Parameters

In this Chapter, all the AOS combinations proposed in Chapter 5, namely, the bandit-based MAB, DMAB and SLMAB Operator Selection mechanisms, combined with Absolute and Normalized versions of the Instantaneous, Average, and Extreme Credit Assignment schemes, as well as the RMAB Operator Selection with the rank/comparison-based Credit Assignment schemes, are compared with one another on different benchmark scenarios.

Regarding the Credit Assignment, all the schemes, except for the Instantaneous one, have a common hyper-parameter, the size of the sliding window $W$. This parameter defines how many operator applications are taken into account to calculate the credit to be assigned to a given operator after its most recent application. It is important to remember that the rank/comparison-based schemes have only one window for all operators, while the other schemes use one window per operator. For the schemes based on the raw values of the fitness improvements, although the fact of normalizing the output or not could be considered as another (boolean) hyper-parameter, schemes employing the Absolute or the Normalized values are separately considered. The rank/comparison-based schemes have an hyper-additional parameter, the exponential decay factor of the ranking distribution, referred to as $D$. Here, again, the choice of use of the fitness values or the fitness improvements as impact measures for the ranking could be seen as a hyper-parameter, but the schemes implementing each of them are independently analyzed and compared with one another.

On the Operator Selection side, all the proposed bandit-based schemes have a common hyper-parameter, the scaling factor $C$. This parameter defines the balance between the UCB exploration and exploitation terms. In case one of the Credit Assignment schemes based on the raw values of fitness improvements is used, $C$ also accounts for the scale of the received rewards, as discussed in Section 5.3.4. Additionally, the DMAB also needs the setting of the threshold $\gamma$ for the Page-Hinkley change-detection statistical test, used by its restarting mechanism. Concerning the SLMAB, it requires the definition of its own sliding window size $w$, used by its update mechanism – after some preliminary experiments, this hyper-parameter will be tuned “for free” here, by using the same value than that of the Credit Assignment $W$.

As baseline AOS method for comparison, we consider the probability-based Adaptive
6.2 General Experimental Settings

Pursuit (AP) Operator Selection scheme [Thierens, 2005] (Section 4.4.2), combined with the same Credit Assignment schemes based on the raw values of fitness improvements. AP needs the setting of: the adaptation rate $\alpha$ to control the update of the empirical quality estimates of each operator; the learning rate $\beta$, which defines the level of greediness of the winner-take-all strategy for the update of the application rates of each operator; and the minimal application probability of each operator, referred to as $p_{\text{min}}$, in order to avoid inefficient operators to get lost by the process (i.e., have zero probability of being applied), as they might become useful in a further stage of the search. Experiments were also done considering the PM method (Section 4.4.1), but its results will be neglected here, due to the fact that it was always outperformed (significantly in most of the times) by AP.

The other methods used for comparison are: the “Naive” uniform strategy, i.e., the operator to be applied is randomly selected using a uniform distribution, which represents what would be a common choice for a naive user; and the “Oracle” strategy, available only to some of the benchmark problems considered, that represents what would be the optimal behavior with respect to operator selection on the problem at hand. Needless to say, these two latter methods do not have any hyper-parameter to be tuned. For the experiments on the boolean and on the continuous benchmark problems, there is an additional baseline technique: the probabilities of applying each operator were off-line tuned for each benchmarking scenario, and the winner configuration is used for comparison, being referred to as “Static”.

The lists of the considered Credit Assignment and Operator Selection schemes, with their corresponding hyper-parameters, are presented in Tables 6.1 and 6.2, respectively; while Table 6.3 summarizes the AOS combinations considered.

<table>
<thead>
<tr>
<th>Baseline Credit Assignment Schemes</th>
<th>Hyper-Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Instantaneous (AbsIns)</td>
<td>—</td>
</tr>
<tr>
<td>Normalized Instantaneous (NormIns)</td>
<td></td>
</tr>
<tr>
<td>Absolute Average (AbsAvg)</td>
<td>W Sliding window size</td>
</tr>
<tr>
<td>Normalized Average (NormAvg)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed Credit Assignment Schemes</th>
<th>Hyper-Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Extreme (AbsExt)</td>
<td>W Sliding window size</td>
</tr>
<tr>
<td>Normalized Extreme (NormExt)</td>
<td></td>
</tr>
<tr>
<td>Decay/Area-Under-Curve (Decay/AUC)</td>
<td>W Sliding window size</td>
</tr>
<tr>
<td>Decay/Fitness-based Area-Under-Curve (Decay/FAUC)</td>
<td></td>
</tr>
<tr>
<td>Decay/Sum-of-Ranks (Decay/SR)</td>
<td>D Decay factor</td>
</tr>
<tr>
<td>Decay/Fitness-based Sum-of-Ranks (Decay/FSR)</td>
<td></td>
</tr>
<tr>
<td>NDCG/Area-Under-Curve (NDCG/AUC)</td>
<td>W Sliding window size</td>
</tr>
<tr>
<td>NDCG/Fitness-based Area-Under-Curve (NDCG/FAUC)</td>
<td></td>
</tr>
<tr>
<td>NDCG/Sum-of-Ranks (NDCG/SR)</td>
<td></td>
</tr>
<tr>
<td>NDCG/Fitness-based Sum-of-Ranks (NDCG/FSR)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Baseline and proposed Credit Assignment methods, and their hyper-parameters
Chapter 6. Experimental Results

### Baseline Operator Selection Methods and Hyper-Parameters

<table>
<thead>
<tr>
<th>Method</th>
<th>Hyper-Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Oracle (when available)</td>
<td>—</td>
</tr>
<tr>
<td>Static (boolean and BBOB)</td>
<td>$P_l$, $P_k$</td>
</tr>
<tr>
<td>Adaptive Pursuit (AP)</td>
<td>$p_{\text{min}}$, $\alpha$, $\beta$</td>
</tr>
</tbody>
</table>

#### Proposed Operator Selection Methods and Hyper-Parameters

<table>
<thead>
<tr>
<th>Method</th>
<th>Hyper-Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Armed Bandit (MAB)</td>
<td>$C$ Scaling factor</td>
</tr>
<tr>
<td>Dynamic Multi-Armed Bandit (DMAB)</td>
<td>$C$ Scaling factor, $\gamma$</td>
</tr>
<tr>
<td>Sliding Multi-Armed Bandit (SLMAB)</td>
<td>$C$, $w$</td>
</tr>
<tr>
<td>Rank-based Multi-Armed Bandit (RMAB)</td>
<td>$C$ Scaling factor</td>
</tr>
</tbody>
</table>

Table 6.2: Baseline and proposed Operator Selection methods, and their hyper-parameters

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abs.</td>
<td>×</td>
<td>Ins.</td>
<td>×</td>
<td>AP</td>
<td>MAB</td>
</tr>
<tr>
<td>Ext.</td>
<td></td>
<td></td>
<td></td>
<td>NDCG</td>
<td>SR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 $\Delta$fitness-based combinations</td>
<td></td>
<td></td>
<td>8 rank-based combinations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: List of 32 AOS combinations considered in most experiments

### 6.2.2 Off-line Tuning of Hyper-Parameters

To promote a fair empirical comparison, it is generally desirable to evaluate the AOS schemes at their best. Accordingly, an off-line tuning was performed preliminarily to every experiment in order to determine, for each AOS combination, the best hyper-parameter configuration. Table 6.4 presents the ranges of values tried for each hyper-parameter, unless stated otherwise.

Briefly, the Credit Assignment settings involve 4 configurations for the schemes based on the raw values of fitness improvements and for the rank-based schemes using the NDCG decaying mechanism, while 20 configurations are explored for the other rank-based schemes (using Decay). The Operator Selection settings include: 64 possible configurations for AP, 7 for MAB and RMAB, and 49 for DMAB. For the SLMAB, there are 7 possible configurations, except for its combination with the Instantaneous, when the Credit Assignment window size is set to 1 but all the 4 values are also tried for the sliding window used by its update rule, thus summing up to 28 configurations. The final number of possible configu-
6.2 General Experimental Settings

| Param.          | Used by        | |values| | Range of Values |
|-----------------|----------------|---|--------------|----------------|
| $p_{min}$       | AP             | 4 |              | $\{0.05, 0.1, 0.2\}$ |
| $\alpha$        | AP             | 4 |              | $\{0.1, 0.3, 0.6, 0.9\}$ |
| $\beta$         | AP             | 4 |              | $\{0.1, 0.3, 0.6, 0.9\}$ |
| $C$             | all MABs       | 7 |              | $\{0.01, 0.1, 0.5, 1, 5, 10, 100\}$ |
| $\gamma$        | DMAB           | 7 |              | $\{10^{-3}, ..., 10^{3}\}$ |
| $W$             | all Credit Assignment | 4 |          | $\{10, 50, 100, 500\}$ |
| $D$             | Decay/rank-based Cr. Assign. | 5 |          | $\{0.25, 0.5, 0.75, 0.9, 1.0\}$ |
| $P_1 \ldots P_k$| Static strategy | 6 |          | $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ |

Table 6.4: Ranges of values tried for the corresponding hyper-parameters

rations for each AOS combination is attained by multiplying the number of configurations for its respective Credit Assignment and Operator Selection components.

Regarding the “Static” strategy, 6 possible values were tried for the application rate of each operator. Only configurations with application rates summing up to 1 were tried. This results in 56 different configurations for the scenarios involving 4 operators, and 126 configurations for the scenarios with 5 operators.

In order to find the optimal values of all hyper-parameters for each method on each of the analyzed scenarios, rather than a complete factorial Design of Experiments, we used the F-Race off-line parameter tuning method [Birattari et al., 2002]. As discussed in Section 3.3.2, the general idea of Racing techniques is to start with all configurations, discarding some of them as soon as there is enough statistical evidence showing that they will not likely be the best one. More specifically, the F-Race applies the Racing paradigm using the Friedman two-way analysis of variance by ranks [Conover, 1999] as statistical test to eliminate candidate configurations. In order to enable a fair comparison, as recommended in [Birattari, 2004b], all the experiments in this work use, for each epoch, the same initial population (the “blocking design” concept). Starting from a minimal number of 11 runs, after each run for all configurations, the elimination of inefficient configurations is performed with the statistical test being applied at a confidence level of 95%. Although 11 runs might be excessive (e.g., [Birattari, 2004a] recommends one run over each instance), we prefer to be conservative here, in order to be sure that the methods are really compared at their best, specially because there is a considerable variance in the results of some of the benchmark problems. The procedure is stopped when a single configuration remains, or when all “survivors” have been run on the maximal number of runs, set to 50 in these experiments. In the latter case, the retained configuration is the one with the best mean amongst the survivors, as done in [Birattari et al., 2002] (although different alternatives could be used, e.g., in a critical situation the configuration with best worst case could be considered). In all cases, 50 runs are launched for the retained configuration and the results presented in this paper are based on statistics over these runs, unless stated otherwise.

It is worth noting that other off-line tuning methods than the F-Race could have been tried, notably the methods that iteratively refine the set of candidate configurations, such as the Iterated F-Race [Balaprakash et al., 2007], the Sequential Parameter Optimization
Chapter 6. Experimental Results

[Bartz-Beielstein et al., 2005], and others, surveyed in Section 3.3.2. However, the objective here was rather to simply do some tuning of the hyper-parameters while being more computationally efficient. Although we acknowledge that the use of different off-line tuning techniques, or different sets of candidate configurations tuned by the same F-Race, could eventually lead to different winners on some benchmark scenarios, we believe that this issue does not affect the global conclusions gathered from the diverse benchmark situations that will be considered in the following.

6.2.3 Performance Indicators and Results Presentation

Several complementary views and indicators for analyzing the AOS performance will be used on each benchmark scenario. The main ones will be summarized now, distinguishing between off-line and on-line performance presentations.

Off-line Performance

The diverse benchmark scenarios considered in the following use different indicators to evaluate the performance of the AOS schemes. These indicators will be described in the corresponding Sections presenting the specific Experimental Settings for each scenario. The off-line performance measures are compared in a Table, for each analyzed scenario, in which each cell corresponds to a Credit Assignment and an Operator Selection, and indicates:

- the average and standard deviation of the performance measure at hand (according to the type of scenario); and
- the values of the best hyper-parameter configuration determined after the F-Race procedure, as described in Section 6.2.2.

Besides the numerical results, two different kinds of statistical comparison are shown in the same Table (see Table 6.5 for instance). The first comparison is the global one: the best performance achieved between all the considered AOS techniques is presented in bold-face with grey background, like this. The results which are not significantly different, according to at least one of both unsigned Wilcoxon rank sum and Kolmogorov-Smirnov non-parametric tests applied at confidence level 90%, are displayed with a grey background. Small result variations among the different AOS schemes thus translate like many grey cells. The second comparison takes place in the scope of each Operator Selection technique: the best performance achieved by it using one of the available Credit Assignment schemes is marked with a ★ symbol. Accordingly, all the schemes within the same Operator Selection technique that obtained equivalent performance, according to the same statistical tests, are marked with a ▲ symbol. Finally, the caption of the table indicates the performances of the Naive uniform strategy, and of the Oracle and Static strategies (whenever available).

In some scenarios, however, given the high dispersion of the results, it is not meaningful to present only the averages and standard deviations, i.e., no significant difference can be found, in terms of performance, between most of the AOS schemes. Because of this, in
order to be able to figure out the difference of performance between the techniques, the empirical distribution of the results over 50 runs is presented, depicted using Empirical Cumulative Distribution Functions (ECDFs): for each level of performance (on $x$-axis) the percentage of runs reaching this score is indicated (on $y$-axis); see, e.g., Figure 6.15b. The $x$-axis is limited by the average performance of the Naive uniform approach; thus, the $y$ value attained by the corresponding ECDF curve at the right border of the plot represents the percentage of runs of the method under assessment that performed better than the Naive uniform baseline. ECDFs are preferred in lieu of standard box-plot diagrams, mainly because they display the whole distribution, consequently enabling a fine-grained comparison of the different schemes, accounting for the fact that one scheme might outperform another scheme with regard to some quantile performance, although being outperformed with respect to the average or median value.

ECDFs are also used to assess the sensitivity of the schemes with respect to their hyper-parameters. More precisely, ECDFs aggregating series of runs corresponding to several hyper-parameter configurations will be presented for each AOS, to graphically show how fast the performance degrades when departing from the optimal parameter configuration.

### On-line Performance

The off-line performance, however, does not tell whether an AOS fails to detect the best operator due to an excess of exploration, or exploitation. In the former case, the AOS fails to stick to the best operator after the change. In the latter case, it fails to swiftly adapt whenever a change occurs.

For this reason, the on-line performance of the AOS schemes with respect to operator selection is also presented for some benchmark scenarios. The on-line performance plots (or behavior plots, e.g., Figure 6.2) depict, for each operator, its instant selection rate along time. For the sake of a smoother presentation, given the high variation of behavior between the many runs, the selection rates plotted represent the average of every 50 time steps for each run, further averaged over 50 runs. Additionally, on all such plots for DMAB, the small peaks below the x-axis indicate the frequency of restarts after the triggering of the PH change-detection statistical test, also averaged in the same way.

### 6.3 On Artificial Scenarios

In order to assess the proposed and the baseline AOS combinations in controlled environments, i.e., in benchmark problems in which the expected behavior is exactly known, experiments were done on the artificial scenarios described in Section 5.4. The specific experimental settings for these experiments, in complement to the general settings presented in Section 6.2, will be described in Section 6.3.1. The empirical comparison of the AOS schemes on the Uniform, Boolean and Outlier scenarios will be surveyed in Section 6.3.2, while Section 6.3.3 will present the results involving two different $ART$ instances. Finally, Section 6.3.4 will present a discussion about the highlights of these experiments. The results that will be analyzed here were partially published in [Da Costa et al., 2008; Fialho et al., 2010a; Fialho et al., 2010c].
6.3.1 Experimental Settings

On these artificial scenarios, the most natural performance indicator is the total gain brought by an AOS scheme, referred to as the Total Cumulative Reward (TCR). It is computed as the sum of the rewards gathered by the algorithm over the complete run.

Besides the average and standard deviation of the TCR, a related but not equivalent indicator, the percentage of times the best operator is selected, is also presented for an illustrative purpose in the comparison tables (see, e.g., Table 6.5); it will be referred to as \( p(\text{best}) \) in the remainder of this text. It is important to note that two techniques might present similar \( p(\text{best}) \), although presenting very different TCR performance, due to the difference between the sub-optimal choices done by each of them, the so-called error costs, which are not taken into account by this measure.

In all cases, the reward distributions are modified every \( \Delta T \) time steps. The epoch lengths considered are \( \Delta T \in \{50, 200, 500, 2000\} \), i.e., ranging from very fast to very slow dynamics. Ten epochs are considered, unless stated otherwise. Although being already defined in the corresponding Sections 5.4.2 and 5.4.3, the sequences used for the exchanges of reward distributions are reminded here: for the Uniform, Boolean and Outlier scenarios, the sequence is \( 41203 \rightarrow 01234 \rightarrow 24301 \rightarrow 12043 \rightarrow 41230 \rightarrow 31420 \rightarrow 04213 \rightarrow 23104 \rightarrow 14302 \rightarrow 40213 \); and for the \( \text{ART} \) scenarios, \( 01 \rightarrow 01 \rightarrow 10 \rightarrow 01 \rightarrow 10 \rightarrow 10 \rightarrow 10 \rightarrow 01 \rightarrow 01 \rightarrow 10 \).

An alternative view of off-line performance is also shown for each artificial scenario: how the performance (in terms of TCR and \( p(\text{best}) \)) of each Operator Selection technique, with its best Credit Assignment scheme found by the off-line tuning, scales with respect to the different epoch lengths considered (see, for instance, Figure 6.1). The motivation is not to compare the results on the different scenarios, but rather to compare the “trend” of the scaling of the performance of the AOS combinations on the different epoch lengths.

6.3.2 Results on Uniform, Boolean and Outlier Scenarios

The Uniform, Boolean and Outlier scenarios involve 5 operators, with rewards coming from different distributions, as described in Sections 5.4.1 and 5.4.2. Empirical results on each of these scenarios will be separately analyzed in the following.

Uniform scenario

On the Uniform scenario, results are very clear and almost in total accordance with respect to all the four different epoch lengths considered. Tables 6.5 and 6.6 present, respectively, the results obtained on \( \Delta T \in \{50, 200\} \) and \( \Delta T \in \{500, 2000\} \).

Given the high informativeness of the rewards received with respect to the quality of the operators, and their steady-state nature during each epoch, it becomes reasonably easy to detect which is the current best operator on this scenario. In the worst case, as the reward distributions of subsequent operators partially overlap between each other, the second best operator might occasionally be considered to be the best one; but the error cost is always very small for the same reason, therefore it does not greatly affect the Total
Cumulative Reward (TCR). The difference in performance thus lies mainly in how fast the AOS techniques are able to adapt to new situations, whenever a change occurs.

Accordingly, using Average or Extreme Credit Assignment schemes will always delay, in $W$ operator applications at most, the perception that the situation has changed. This is a tentative explanation for the fact that both Absolute and Normalized versions of the Instantaneous Credit Assignment scheme are clearly the best options for all the Operator Selection methods tried with them. This interpretation is supported by the output of the Racing process for each of the AOS combinations using the Average and the Extreme schemes: the best configuration retained for all of them has the same sliding window size $W = 10$, i.e., the lowest value in the range tried for this hyper-parameter (Table 6.4).

Between the rank-based Credit Assignment schemes, the AUC is the best option. The fact that a linear decay (i.e., $D = 1$) is retained as best configuration might also be related to the subtle overlap between the rewards received by the different operators: in such a case, there is no need of a strong decaying factor to differ between them.

On the Operator Selection side, the clear winner is DMAB, which, combined with the Absolute Instantaneous (AbsIns) Credit Assignment scheme, significantly outperforms all the other AOS combinations on all epochs, except for the same DMAB with Normalized Instantaneous (NormIns) Credit Assignment in three out of four cases. Indeed, for the longest epoch ($\Delta T = 2000$), the AbsIns-DMAB AOS technique obtains a TCR equivalent to 99.7% of the Oracle TCR, selecting the best operator in around 99.5% of the applications. This outstanding performance is explained by the well-calibrated PH test (as shown in Figures 6.2a and 6.2b). Although using a very sensitive hyper-parameter (see sensitivity analysis in Section 6.7), the restarted mechanism is very efficient on this kind of situation in which the qualities of the operators change abruptly. For faster dynamics (smaller $\Delta T$), the performance is gracefully degraded: shorter the steady-state epoch, less negligible becomes the price to pay for the extra exploration needed after each restart. But, still, DMAB remains the clear winner with respect to TCR and $p(\text{best})$ for all epoch lengths, as shown in Figure 6.1.

The standard MAB is the second best Operator Selection technique on this scenario, achieving a performance slightly (but significantly, given the small standard deviations) inferior than that of DMAB, for all epoch lengths. Its combination with either AbsIns or NormIns Credit Assignment schemes, although being the simplest bandit-based AOS combination considered, is able to achieve up to 97% of the Oracle TCR for $\Delta T = 2000$, selecting the best operator in around 92% of the times (Figure 6.2c); its well-tuned EvE balance enables the discovery of the new situation by paying the minimal price with respect to exploration. The SLMAB is also able to swiftly follow the dynamics of the scenario, employing the same level of exploration; but it achieves similar performance than that of MAB only because of the low error cost, as it is not able to differ well between the best and the second best operators in some of the epochs (Figure 6.2d); this explains why its $p(\text{best})$ is much lower than that of MAB, although obtaining a similar TCR. The AUC-RMAB combination with linear decay achieves very similar performance (Figure 6.2e) than that of MAB. Given the very low value retained for its scaling factor $C$, it becomes clear that all the adaptation is in fact done on the Credit Assignment side: an impact measure brought by the application of one operator might affect the whole ranking distribution,
Conversely, the baseline probability-based Adaptive Pursuit (AP) method does not follow the same trend, being significantly outperformed by all bandit-based schemes on the two longer epochs (see Figure 6.1). This limitation is blamed on an excess of exploration (lower bounded by the $p_{\text{min}}=0.05$ parameter), as shown in Figure 6.2f. It is worth noting that $p_{\text{min}}=0$ was also tried in the parameter tuning process (Table 6.4), but achieved lower performance than that of $p_{\text{min}}=0.05$. The value used in fact translates into up to 25% of exploration trials in this case, as $p_{\text{min}}$ refers to the minimal level of exploration for each operator. Finally, all AOS combinations succeed in performing significantly better than the Naive uniform baseline method.

Figure 6.1: Scaling of mean performance (TCR above, and $p(\text{best})$ below) in relation to the epoch length $\Delta T$, for each Operator Selection technique with its best Credit Assignment scheme, on the Uniform scenario.

consequently affecting the quality estimate of all operators, as discussed in Section 5.2.4.
6.3 On Artificial Scenarios

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>2077 ± 55</td>
<td>2171 ± 36</td>
<td>2120 ± 23</td>
<td>2103 ± 35</td>
</tr>
<tr>
<td></td>
<td>C10W50</td>
<td>C1G1W1</td>
<td>C5W1</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td>Pb: 65.6 ± 4.9</td>
<td>Pb: 66.9 ± 4.8</td>
<td>Pb: 63.5 ± 1.7</td>
<td>Pb: 58.9 ± 5.6</td>
<td></td>
</tr>
<tr>
<td>NormIns</td>
<td>1957 ± 109</td>
<td>2294 ± 30</td>
<td>2142 ± 61</td>
<td>2077 ± 43</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.01G.1W1</td>
<td>C5W1</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td>Pb: 44.8 ± 11.6</td>
<td>Pb: 75.6 ± 5.3</td>
<td>Pb: 55.8 ± 8.6</td>
<td>Pb: 56.6 ± 5.5</td>
<td></td>
</tr>
<tr>
<td>AbsAvg</td>
<td>1926 ± 103</td>
<td>2084 ± 49</td>
<td>1915 ± 40</td>
<td>2030 ± 70</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.5G1W10</td>
<td>C1W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td>Pb: 45.2 ± 10.7</td>
<td>Pb: 48.9 ± 6.0</td>
<td>Pb: 41.7 ± 4.6</td>
<td>Pb: 39.8 ± 6.4</td>
<td></td>
</tr>
<tr>
<td>NormAvg</td>
<td>1896 ± 102</td>
<td>1907 ± 52</td>
<td>1907 ± 52</td>
<td>1906 ± 100</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.5G1W10</td>
<td>C5W10</td>
<td>P.0.1B.9W10</td>
</tr>
<tr>
<td>Pb: 41.6 ± 11.3</td>
<td>Pb: 34.1 ± 5.3</td>
<td>Pb: 34.1 ± 5.3</td>
<td>Pb: 33.3 ± 8.0</td>
<td></td>
</tr>
<tr>
<td>AbsExt</td>
<td>1879 ± 97</td>
<td>2120 ± 33</td>
<td>1884 ± 46</td>
<td>1892 ± 43</td>
</tr>
<tr>
<td></td>
<td>C10W100</td>
<td>C.5G1W10</td>
<td>C1W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 38.8 ± 6.6</td>
<td>Pb: 70.0 ± 3.4</td>
<td>Pb: 40.8 ± 5.8</td>
<td>Pb: 46.7 ± 6.3</td>
<td></td>
</tr>
<tr>
<td>NormExt</td>
<td>1800 ± 105</td>
<td>1990 ± 54</td>
<td>1949 ± 39</td>
<td>1868 ± 41</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.01G.1W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 38.4 ± 9.3</td>
<td>Pb: 50.0 ± 7.2</td>
<td>Pb: 51.4 ± 3.3</td>
<td>Pb: 43.2 ± 5.9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>1952 ± 19</td>
<td>1947 ± 18</td>
<td>1927 ± 16</td>
<td>1937 ± 17</td>
</tr>
<tr>
<td></td>
<td>C.5D1W10</td>
<td>C.5D.75W10</td>
<td>C.5W10</td>
<td>C.5W10</td>
</tr>
<tr>
<td>Pb: 55.8 ± 1.3</td>
<td>Pb: 56.4 ± 1.5</td>
<td>Pb: 55.0 ± 1.8</td>
<td>Pb: 55.7 ± 1.6</td>
<td></td>
</tr>
</tbody>
</table>

(a) Results on the Uniform scenario, for ∆T=50 (Naive TCR: 1500, Optimal TCR: 2500)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>8530 ± 329</td>
<td>9773 ± 117</td>
<td>9145 ± 37</td>
<td>8826 ± 73</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.1G10W1</td>
<td>C5W1</td>
<td>P.05A.6B.9W1</td>
</tr>
<tr>
<td>Pb: 56.7 ± 11.8</td>
<td>Pb: 95.1 ± 1.5</td>
<td>Pb: 78.6 ± 0.7</td>
<td>Pb: 73.0 ± 2.5</td>
<td></td>
</tr>
<tr>
<td>NormIns</td>
<td>8610 ± 106</td>
<td>9699 ± 136</td>
<td>9156 ± 154</td>
<td>8806 ± 68</td>
</tr>
<tr>
<td></td>
<td>C5W500</td>
<td>C.01G.001W1</td>
<td>C5W1</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td>Pb: 71.3 ± 1.9</td>
<td>Pb: 89.7 ± 6.8</td>
<td>Pb: 70.4 ± 6.3</td>
<td>Pb: 71.2 ± 2.2</td>
<td></td>
</tr>
<tr>
<td>AbsAvg</td>
<td>8355 ± 346</td>
<td>9066 ± 108</td>
<td>8769 ± 64</td>
<td>8561 ± 300</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C1G1W10</td>
<td>C5W10</td>
<td>P.0A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 54.3 ± 12.2</td>
<td>Pb: 67.6 ± 5.1</td>
<td>Pb: 71.2 ± 1.7</td>
<td>Pb: 43.9 ± 9.3</td>
<td></td>
</tr>
<tr>
<td>NormAvg</td>
<td>7909 ± 427</td>
<td>8643 ± 138</td>
<td>8643 ± 138</td>
<td>8282 ± 121</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.5G1W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 36.6 ± 11.0</td>
<td>Pb: 49.5 ± 4.7</td>
<td>Pb: 49.5 ± 4.7</td>
<td>Pb: 50.7 ± 4.5</td>
<td></td>
</tr>
<tr>
<td>AbsExt</td>
<td>8128 ± 323</td>
<td>9526 ± 39</td>
<td>8867 ± 43</td>
<td>8626 ± 85</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.5G1W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 43.2 ± 10.6</td>
<td>Pb: 90.2 ± 0.8</td>
<td>Pb: 74.6 ± 0.8</td>
<td>Pb: 70.7 ± 2.7</td>
<td></td>
</tr>
<tr>
<td>NormExt</td>
<td>7668 ± 393</td>
<td>9236 ± 159</td>
<td>8908 ± 142</td>
<td>8593 ± 94</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.1G.001W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td>Pb: 34.3 ± 11.2</td>
<td>Pb: 77.8 ± 6.4</td>
<td>Pb: 69.8 ± 5.6</td>
<td>Pb: 69.6 ± 3.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>8755 ± 87</td>
<td>8469 ± 92</td>
<td>8469 ± 92</td>
<td>8469 ± 92</td>
</tr>
<tr>
<td></td>
<td>C.1D1W50</td>
<td>C.1D.5W50</td>
<td>C.1W50</td>
<td>C.1W50</td>
</tr>
<tr>
<td>Pb: 74.6 ± 2.5</td>
<td>Pb: 68.9 ± 3.7</td>
<td>Pb: 68.9 ± 3.7</td>
<td>Pb: 68.9 ± 3.7</td>
<td></td>
</tr>
</tbody>
</table>

(b) Results on the Uniform scenario, for ∆T=200 (Naive TCR: 6000, Optimal TCR: 10000)

Table 6.5: Results on the Uniform scenario for ∆T ∈ {50, 200}
### Chapter 6. Experimental Results

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>22948 ± 678</td>
<td>24785 ± 41</td>
<td>23572 ± 66</td>
<td>2270 ± 91</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.1G10W1</td>
<td>C5W1</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 73.2 ± 11.8</td>
<td>Pb: 98.1 ± 0.2</td>
<td>Pb: 85.5 ± 0.5</td>
<td>Pb: 76.3 ± 1.2</td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>21041 ± 637</td>
<td>24443 ± 400</td>
<td>23294 ± 377</td>
<td>22212 ± 97</td>
</tr>
<tr>
<td></td>
<td>C5W500</td>
<td>C.01G.001W1</td>
<td>C5W1</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 63.6 ± 6.6</td>
<td>Pb: 90.9 ± 8.0</td>
<td>Pb: 73.3 ± 7.1</td>
<td>Pb: 75.4 ± 1.1</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td>22659 ± 631</td>
<td>23494 ± 222</td>
<td>23199 ± 114</td>
<td>21778 ± 119</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C1G1W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 69.8 ± 11.0</td>
<td>Pb: 78.5 ± 4.0</td>
<td>Pb: 82.8 ± 1.0</td>
<td>Pb: 68.2 ± 1.9</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>20712 ± 1057</td>
<td>22777 ± 253</td>
<td>22502 ± 221</td>
<td>21671 ± 133</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.5G.01W10</td>
<td>C1W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 42.6 ± 12.8</td>
<td>Pb: 71.4 ± 3.9</td>
<td>Pb: 68.2 ± 3.4</td>
<td>Pb: 76.7 ± 2.0</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td>21578 ± 776</td>
<td>24509 ± 48</td>
<td>23275 ± 69</td>
<td>22085 ± 105</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.5G10W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 50.5 ± 14.1</td>
<td>Pb: 95.9 ± 0.4</td>
<td>Pb: 83.7 ± 0.5</td>
<td>Pb: 75.7 ± 1.3</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>20471 ± 1526</td>
<td>23882 ± 355</td>
<td>23109 ± 366</td>
<td>22063 ± 114</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.1G.001W10</td>
<td>C.5W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 43.0 ± 14.4</td>
<td>Pb: 84.4 ± 6.1</td>
<td>Pb: 75.7 ± 5.7</td>
<td>Pb: 75.5 ± 1.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMAB (∆F)</strong></td>
<td>23251 ± 117</td>
<td>22804 ± 134</td>
<td>22804 ± 134</td>
<td>22804 ± 134</td>
</tr>
<tr>
<td></td>
<td>C.1D1W50</td>
<td>C.1D5W50</td>
<td>C.1W50</td>
<td>C.1W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 86.0 ± 1.2</td>
<td>Pb: 81.8 ± 2.6</td>
<td>Pb: 81.8 ± 2.6</td>
<td>Pb: 81.8 ± 2.6</td>
</tr>
</tbody>
</table>

(a) Results on the Uniform scenario, for ∆T=500 (Naive TCR: 15000, Optimal TCR: 25000)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>95872 ± 2745</td>
<td>99781 ± 72</td>
<td>96954 ± 128</td>
<td>89617 ± 156</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C.1G10W1</td>
<td>C5W1</td>
<td>P.05A.3B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 83.3 ± 13.1</td>
<td>Pb: 99.5 ± 0.0</td>
<td>Pb: 92.2 ± 0.3</td>
<td>Pb: 78.5 ± 0.5</td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>83221 ± 6417</td>
<td>98731 ± 1583</td>
<td>95329 ± 916</td>
<td>89534 ± 159</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.01G.001W1</td>
<td>C1W1</td>
<td>P.05A.3B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 45.3 ± 17.2</td>
<td>Pb: 94.2 ± 7.9</td>
<td>Pb: 84.5 ± 3.3</td>
<td>Pb: 78.3 ± 0.5</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td>95135 ± 663</td>
<td>97625 ± 1105</td>
<td>96332 ± 330</td>
<td>89242 ± 174</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C1G10W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 86.2 ± 2.4</td>
<td>Pb: 89.8 ± 5.4</td>
<td>Pb: 91.1 ± 0.7</td>
<td>Pb: 76.6 ± 0.6</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>87364 ± 4064</td>
<td>95958 ± 718</td>
<td>93741 ± 912</td>
<td>89137 ± 192</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.5G.001W10</td>
<td>C1W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 51.0 ± 13.1</td>
<td>Pb: 84.5 ± 3.5</td>
<td>Pb: 77.4 ± 3.7</td>
<td>Pb: 76.7 ± 0.7</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td>94716 ± 613</td>
<td>99492 ± 79</td>
<td>96658 ± 99</td>
<td>89560 ± 154</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.5G10W10</td>
<td>C5W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 85.6 ± 1.9</td>
<td>Pb: 98.9 ± 0.1</td>
<td>Pb: 91.7 ± 0.2</td>
<td>Pb: 78.3 ± 0.4</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>85092 ± 6055</td>
<td>97103 ± 1444</td>
<td>95044 ± 656</td>
<td>89540 ± 161</td>
</tr>
<tr>
<td></td>
<td>C1W10</td>
<td>C.1G.001W10</td>
<td>C1W10</td>
<td>P.05A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 48.8 ± 15.1</td>
<td>Pb: 87.2 ± 7.3</td>
<td>Pb: 85.5 ± 2.1</td>
<td>Pb: 78.8 ± 0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMAB (∆F)</strong></td>
<td>96449 ± 231</td>
<td>95501 ± 234</td>
<td>95501 ± 234</td>
<td>95501 ± 234</td>
</tr>
<tr>
<td></td>
<td>C.1D1W100</td>
<td>C.1D5W100</td>
<td>C.1W100</td>
<td>C.1W100</td>
</tr>
<tr>
<td></td>
<td>Pb: 92.9 ± 0.6</td>
<td>Pb: 90.9 ± 1.1</td>
<td>Pb: 90.9 ± 1.1</td>
<td>Pb: 90.9 ± 1.1</td>
</tr>
</tbody>
</table>

(b) Results on the Uniform scenario, for ∆T=2000 (Naive TCR: 60000, Optimal TCR: 100000)

Table 6.6: Results on the Uniform scenario for ∆T ∈ {500, 2000}
6.3 On Artificial Scenarios

Figure 6.2: Behavior of DMAB, MAB, SLMAB, RMAB and AP, combined with their best Credit Assignment schemes, on the Uniform scenario with $\Delta T = 2000$. The outstanding performance of DMAB, using either AbsInst or AbsExt Credit Assignment schemes, is achieved by the fact that restarts are perfectly triggered by the PH test in the transitions, as indicated by the small peaks below the x-axis of the corresponding plots.
Boolean scenario

Differently from the Uniform scenario, in the Boolean scenario it is not the values of the rewards that inform which is the best operator, but rather the frequency with which each operator is rewarded. This makes this scenario a very difficult one for AOS, as discussed in Section 5.4.2. Tables 6.7 and 6.8 present, respectively, the results obtained on $\Delta T \in \{50, 200\}$ and $\Delta T \in \{500, 2000\}$, confirming this assumption.

For instance, with $\Delta T = 50$, only 47% of good choices are made by SLMAB, the best method in this case, and less than 40% by all others. When $\Delta T$ increases, the overall performance of all Operator Selection techniques (with their corresponding best Credit Assignment scheme) in relation to both TCR and $p(best)$ measures, gradually increases accordingly, as shown in Figure 6.3. Anyway, the maximum $p(best)$ attained in the longest epoch is 80%, while in the Uniform scenario rates up to 99.5% were found.

Concerning the Operator Selection techniques, for the shortest epoch, SLMAB achieves the best TCR, but its performance is not significantly different from all the others. Starting from $\Delta T = 200$, the DMAB takes the lead, with the gap between its performance and those of the others increasing, up to the longest epoch, in which the winner configuration for DMAB is significantly better than all the other techniques. It is important to note that, as the values of the rewards received are always the same (= 10), very few restarts are done by DMAB (see, e.g., Figures 6.4a and 6.4b); hence, the performance of the standard MAB is very similar to the DMAB performance, being significantly different only for the longest epoch. The other three Operator Selection techniques, namely, SLMAB, RMAB and the baseline AP, present equivalent but inferior performance for all epoch lengths, except for the longest one, in which SLMAB is significantly better than AP, but still equivalent to RMAB.

By analyzing these behavior plots for $\Delta T = 2000$ in Figure 6.4, it can be seen that, except for the overall winner NormIns-DMAB (Figure 6.4a) and the second best NormIns-MAB (Figure 6.4c), the performance of the other bandit-based AOS combinations is hindered by a lack of further exploitation of the best operator: in around 20% of the trials, a sub-optimal operator is applied (see Figures 6.4d and 6.4e, respectively, for the behavior of SLMAB and RMAB). For AP, although the $p_{min}$ value is set to 0 in this case, its low performance is explained by a failure in identifying the best operator in all cases where the best operator becomes the second best (Figure 6.4f), possibly due to the high inertia of the two-tiered update of empirical quality estimates employed by this method, as discussed in Section 4.4.2.

For the Credit Assignment, as in the Uniform scenario, the best scheme is the Instantaneous one, both Absolute and Normalized alternatives performing equivalently in almost all cases. The Average-based schemes perform almost as good; while the Extreme-based ones are outperformed by far: as the only values that are taken here are 0 or 10, it becomes difficult for the Extreme reward to distinguish among operators. Along the same lines, a tentative of interpretation for the inefficiency of the rank-based Credit Assignment schemes (no matter the decaying factor used), is that the only two possible values for the rewards do not enable enough granularity in the ranking distribution, consequently resulting in similar qualities to the operators.
6.3 On Artificial Scenarios

Figure 6.3: Scaling of mean performance (TCR above, and \( p(\text{best}) \) below) in relation to the epoch length \( \Delta T \), for each Operator Selection technique with its best Credit Assignment scheme, on the Boolean scenario.
### Chapter 6. Experimental Results

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>1922 ± 162 ⋆</td>
<td>1893 ± 136 ⋆</td>
<td>1893 ± 136 ⋆</td>
<td>1798 ± 148 ⋆</td>
</tr>
<tr>
<td></td>
<td><strong>C10W50</strong></td>
<td>C5G1000W1</td>
<td>C5W1</td>
<td>P0A.3B.9W1</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 47.3 ± 10.0</strong></td>
<td>Pb: 38.7 ± 7.7</td>
<td>Pb: 38.7 ± 7.7</td>
<td>Pb: 29.8 ± 9.2</td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>1922 ± 162 ▲</td>
<td>1890 ± 151 ▲</td>
<td>1890 ± 151 ▲</td>
<td>1852 ± 144 ▲</td>
</tr>
<tr>
<td></td>
<td><strong>C1W50</strong></td>
<td>C.5G1W01</td>
<td>C.5W10</td>
<td>P0A.1B.9W1</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 47.3 ± 10.0</strong></td>
<td>Pb: 38.9 ± 7.2</td>
<td>Pb: 38.9 ± 7.2</td>
<td>Pb: 26.3 ± 9.2</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td>1855 ± 182 ▲</td>
<td>1793 ± 132</td>
<td>1735 ± 132</td>
<td>1727 ± 116</td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td>C.5G1W10</td>
<td>C.5W10</td>
<td>P.05A.6B.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 37.4 ± 9.7</strong></td>
<td>Pb: 31.3 ± 7.0</td>
<td>Pb: 29.6 ± 7.3</td>
<td>Pb: 28.5 ± 6.7</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>1743 ± 169</td>
<td>1775 ± 153</td>
<td>1764 ± 127</td>
<td>1723 ± 130</td>
</tr>
<tr>
<td></td>
<td><strong>C1W10</strong></td>
<td>C.01G.1W10</td>
<td>C.1W10</td>
<td>P.05A.9B.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 31.4 ± 11.8</strong></td>
<td>Pb: 27.7 ± 9.3</td>
<td>Pb: 29.1 ± 6.5</td>
<td>Pb: 27.4 ± 8.2</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td>1639 ± 168</td>
<td>1679 ± 162</td>
<td>1674 ± 173</td>
<td>1656 ± 158</td>
</tr>
<tr>
<td></td>
<td><strong>C10W10</strong></td>
<td>C5G1000W10</td>
<td>C5W10</td>
<td>P0A.6B.6W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 23.6 ± 10.9</strong></td>
<td>Pb: 25.6 ± 6.2</td>
<td>Pb: 25.4 ± 5.9</td>
<td>Pb: 22.0 ± 7.1</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>1639 ± 168</td>
<td>1670 ± 173</td>
<td>1670 ± 173</td>
<td>1693 ± 128</td>
</tr>
<tr>
<td></td>
<td><strong>C1W10</strong></td>
<td>C.5G1000W10</td>
<td>C.5W10</td>
<td>P0A.3B.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 23.6 ± 10.9</strong></td>
<td>Pb: 25.6 ± 5.9</td>
<td>Pb: 25.6 ± 5.9</td>
<td>Pb: 23.4 ± 8.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>1798 ± 135 ⋆</td>
<td>1716 ± 119</td>
<td>1775 ± 127 ▲</td>
<td>1724 ± 116 ▲</td>
</tr>
<tr>
<td></td>
<td><strong>C5D1W50</strong></td>
<td>C1D.5W50</td>
<td>C1W500</td>
<td>C1W50</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 35.9 ± 6.6</strong></td>
<td>Pb: 30.8 ± 4.6</td>
<td>Pb: 33.7 ± 4.7</td>
<td>Pb: 31.2 ± 4.4</td>
</tr>
</tbody>
</table>

(a) Results on the Boolean scenario, for ΔT=50 (Naive TCR: 1500, Optimal TCR: 2500)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>8058 ± 427 ⋆</td>
<td>8154 ± 348 ⋆</td>
<td>8154 ± 348 ⋆</td>
<td>7966 ± 249</td>
</tr>
<tr>
<td></td>
<td><strong>C10W100</strong></td>
<td>C5G1000W1</td>
<td>C5W1</td>
<td>P0A.1B.6W1</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 55.7 ± 10.6</strong></td>
<td>Pb: 50.8 ± 8.6</td>
<td>Pb: 50.8 ± 8.6</td>
<td>Pb: 45.7 ± 6.1</td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>8058 ± 427 ▲</td>
<td><strong>8162 ± 356 ▲</strong></td>
<td>8152 ± 364 ▲</td>
<td>8160 ± 393 ▲</td>
</tr>
<tr>
<td></td>
<td><strong>C1W100</strong></td>
<td><strong>C.5G10W1</strong></td>
<td>C.5W1</td>
<td>P0A.1B.9W1</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 55.7 ± 10.6</strong></td>
<td><strong>Pb: 52.1 ± 8.9</strong></td>
<td>Pb: 51.5 ± 8.9</td>
<td>Pb: 45.0 ± 10.4</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td>7979 ± 380 ▲</td>
<td>8063 ± 360 ▲</td>
<td>7921 ± 313</td>
<td>7871 ± 271</td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td><strong>C1G10W10</strong></td>
<td>C.5W10</td>
<td>P0A.9B.6W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 46.5 ± 9.6</strong></td>
<td>Pb: 45.2 ± 8.6</td>
<td>Pb: 45.7 ± 6.2</td>
<td>Pb: 45.0 ± 5.0</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>7536 ± 494</td>
<td>8033 ± 416 ▲</td>
<td>7769 ± 376</td>
<td>7817 ± 249</td>
</tr>
<tr>
<td></td>
<td><strong>C1W10</strong></td>
<td><strong>C.01G.01W10</strong></td>
<td>C1W10</td>
<td>P0A.44.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 32.5 ± 11.0</strong></td>
<td>Pb: 43.3 ± 11.0</td>
<td>Pb: 39.2 ± 8.9</td>
<td>Pb: 44.4 ± 5.8</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td>7414 ± 420</td>
<td>7475 ± 520</td>
<td>7402 ± 429</td>
<td>7477 ± 526</td>
</tr>
<tr>
<td></td>
<td><strong>C10W10</strong></td>
<td><strong>C1G10W10</strong></td>
<td>C5W10</td>
<td>P0A.9B.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 30.5 ± 9.4</strong></td>
<td>Pb: 33.0 ± 10.7</td>
<td>Pb: 31.3 ± 5.5</td>
<td>Pb: 31.4 ± 9.7</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>7414 ± 420</td>
<td>7526 ± 480</td>
<td>7409 ± 437</td>
<td>7611 ± 358</td>
</tr>
<tr>
<td></td>
<td><strong>C1W10</strong></td>
<td><strong>C.01G1W10</strong></td>
<td>C.5W10</td>
<td>P0A.1B.9W10</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 30.5 ± 9.4</strong></td>
<td>Pb: 31.7 ± 11.4</td>
<td>Pb: 31.6 ± 5.7</td>
<td>Pb: 31.3 ± 8.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td>7841 ± 289 ⋆</td>
<td>7476 ± 361</td>
<td>7580 ± 246</td>
<td>7466 ± 267</td>
</tr>
<tr>
<td></td>
<td><strong>C5D1W100</strong></td>
<td>C.5D.75W50</td>
<td>C.5W100</td>
<td>C1W100</td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 47.4 ± 7.8</strong></td>
<td>Pb: 36.1 ± 8.9</td>
<td>Pb: 40.6 ± 3.2</td>
<td>Pb: 42.1 ± 5.0</td>
</tr>
</tbody>
</table>

(b) Results on the Boolean scenario, for ΔT=200 (Naive TCR: 6000, Optimal TCR: 10000)

Table 6.7: Results on the Boolean scenario for ΔT ∈ {50, 200}
6.3 On Artificial Scenarios

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>21234 ± 751 ▲</td>
<td><strong>21888 ± 611 ★</strong></td>
<td>21532 ± 834 ▲</td>
<td>20783 ± 1098 ▲</td>
</tr>
<tr>
<td>Pb: 56.5 ± 9.5</td>
<td>C5W10</td>
<td>C5G100W1</td>
<td>C5W1</td>
<td>P0A.3B.9W1</td>
</tr>
<tr>
<td>NormIns</td>
<td>21234 ± 751 ▲</td>
<td>21740 ± 658 ▲</td>
<td>21578 ± 817 ★</td>
<td>21245 ± 983 ★</td>
</tr>
<tr>
<td>Pb: 56.5 ± 9.5</td>
<td>C5W10</td>
<td>C5G100W1</td>
<td>C5W1</td>
<td>P0A.1B.9W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>21412 ± 531 ★</td>
<td>21588 ± 604 ▲</td>
<td>21106 ± 579</td>
<td>20406 ± 519</td>
</tr>
<tr>
<td>Pb: 59.3 ± 6.7</td>
<td>C5W10</td>
<td>C5G100W10</td>
<td>C10W10</td>
<td>P0A.5A.9W10</td>
</tr>
<tr>
<td>NormAvg</td>
<td>19883 ± 1003</td>
<td>21495 ± 591 ▲</td>
<td>20711 ± 752</td>
<td>20263 ± 445</td>
</tr>
<tr>
<td>Pb: 36.7 ± 11.6</td>
<td>C1W10</td>
<td>C0.01GW10</td>
<td>C1W10</td>
<td>P0A.5A.9W10</td>
</tr>
<tr>
<td>AbsExt</td>
<td>19852 ± 926</td>
<td>20515 ± 854</td>
<td>19649 ± 751</td>
<td>19728 ± 1144</td>
</tr>
<tr>
<td>Pb: 38.2 ± 10.8</td>
<td>C10W10</td>
<td>C1G10W10</td>
<td>C5W10</td>
<td>P0A.9B.6W10</td>
</tr>
<tr>
<td>NormExt</td>
<td>19852 ± 926</td>
<td>20492 ± 850</td>
<td>19692 ± 730</td>
<td>20074 ± 1188</td>
</tr>
<tr>
<td>Pb: 38.2 ± 10.8</td>
<td>C1W10</td>
<td>C0.01GW10</td>
<td>C5W10</td>
<td>P0A.1B.9W10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMAB (ΔF)</td>
<td>20942 ± 628 ★</td>
<td>20234 ± 615</td>
<td>19917 ± 515</td>
<td>20199 ± 657</td>
</tr>
<tr>
<td>C1D.9W50</td>
<td>19053 ± 2083 ▲</td>
<td>92361 ± 818</td>
<td>91954 ± 1044 ▲</td>
<td>85949 ± 4444</td>
</tr>
<tr>
<td>Pb: 69.6 ± 8.2</td>
<td>C5W10</td>
<td>C5G100W1</td>
<td>C10W10</td>
<td>P0A.3B.6W1</td>
</tr>
<tr>
<td>RMAB (∆F)</td>
<td>90844 ± 2083 ▲</td>
<td><strong>93735 ± 1291★</strong></td>
<td>91986 ± 1040 ★</td>
<td>88921 ± 3679 ★</td>
</tr>
<tr>
<td>C5W10</td>
<td>19053 ± 2083 ▲</td>
<td>92361 ± 818</td>
<td>91954 ± 1044 ▲</td>
<td>85949 ± 4444</td>
</tr>
<tr>
<td>Pb: 69.6 ± 8.2</td>
<td>C5W10</td>
<td>C5G100W10</td>
<td>C10W10</td>
<td>P0A.1B.9W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>91285 ± 2300</td>
<td>92650 ± 858</td>
<td>91479 ± 1037 ▲</td>
<td>85488 ± 812</td>
</tr>
<tr>
<td>Pb: 70.6 ± 8.2</td>
<td>C5W10</td>
<td>C5G100W10</td>
<td>C10W10</td>
<td>P0A.5B.6W50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMAB (ΔF)</td>
<td>90530 ± 1293 ★</td>
<td>87203 ± 1784</td>
<td>82875 ± 1501</td>
<td>87108 ± 1576</td>
</tr>
<tr>
<td>C1D.1W50</td>
<td>90530 ± 1293 ★</td>
<td>87203 ± 1784</td>
<td>82875 ± 1501</td>
<td>87108 ± 1576</td>
</tr>
<tr>
<td>Pb: 72.2 ± 4.8</td>
<td>19883 ± 1003</td>
<td>19053 ± 2083</td>
<td>19053 ± 2083</td>
<td>19053 ± 2083</td>
</tr>
</tbody>
</table>

(a) Results on the Boolean scenario, for ΔT=500 (Naive TCR: 15000, Optimal TCR: 25000)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>85793 ± 3017</td>
<td>91755 ± 2273</td>
<td>87394 ± 2652</td>
<td>85803 ± 1244</td>
</tr>
<tr>
<td>Pb: 50.6 ± 11.6</td>
<td>C1W10</td>
<td>C0.01GW10</td>
<td>C1W10</td>
<td>P0A.3B.9W1</td>
</tr>
<tr>
<td>NormAvg</td>
<td>85480 ± 3093</td>
<td>90258 ± 2535</td>
<td>83560 ± 3862</td>
<td>84695 ± 3298</td>
</tr>
<tr>
<td>Pb: 70.6 ± 8.2</td>
<td>C10W10</td>
<td>C0.01GW10</td>
<td>C5W10</td>
<td>P0A.9B.9W10</td>
</tr>
<tr>
<td>AbsExt</td>
<td>85468 ± 3070</td>
<td>90708 ± 2028</td>
<td>83620 ± 3806</td>
<td>85357 ± 4002</td>
</tr>
<tr>
<td>Pb: 48.8 ± 12.7</td>
<td>C1W10</td>
<td>C0.01GW10</td>
<td>C5W10</td>
<td>P0A.3B.9W10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMAB (ΔF)</td>
<td>90530 ± 1293 ★</td>
<td>87203 ± 1784</td>
<td>82875 ± 1501</td>
<td>87108 ± 1576</td>
</tr>
<tr>
<td>C1D.1W50</td>
<td>90530 ± 1293 ★</td>
<td>87203 ± 1784</td>
<td>82875 ± 1501</td>
<td>87108 ± 1576</td>
</tr>
<tr>
<td>Pb: 72.2 ± 4.8</td>
<td>85480 ± 3093</td>
<td>90258 ± 2535</td>
<td>83560 ± 3862</td>
<td>84695 ± 3298</td>
</tr>
</tbody>
</table>

(b) Results on the Boolean scenario, for ΔT=2000 (Naive TCR: 60000, Optimal TCR: 100000)

Table 6.8: Results on the Boolean scenario for ΔT ∈ {500, 2000}
Figure 6.4: Behavior of DMAB, MAB, SLMAB, RMAB and AP, combined with their best Credit Assignment schemes, on the Boolean scenario with $\Delta T = 2000$. DMAB is the overall winner again, combined with the NormIns Credit Assignment scheme: very few restarts are correctly triggered. Conversely, when combined with the AbsExt Credit Assignment, it triggers several misplaced restarts, while failing to correctly follow the changes (as does NormIns-AP).
Outlier scenario

The Outlier scenario is by far the most difficult between the three scenarios involving 5 artificial operators. As discussed in Section 5.4.2, although providing very informative rewards about the quality of each operator (such as in the Uniform case, but without any overlapping), the non-zero rewards are very rare (only 10% of the cases), resulting in a huge variance ($V = 225$ for the best operator, while $V = 25$ for the same best in the Boolean scenario, both with $E = 5$) that greatly complicates the job of the AOS schemes. As for the other two scenarios, empirical results on this scenario are presented in Tables 6.9 and 6.10, respectively, for $\Delta T \in \{50, 200\}$ and $\Delta T \in \{500, 2000\}$.

Accordingly, all techniques perform very poorly for small values of $\Delta T$. This is not a surprise, due to the small chance of seeing some outlier reward within 50 or even 200 time steps. For instance, the best TCR obtained over all techniques is 1722 for $\Delta T = 50$ (with a $p(\text{best})$ of only 28%), while the naive approach would do 1500; and 7560 for $\Delta T = 200$, versus 6000 for the naive strategy. This situation changes for some techniques, when the steady-state period between each switch in the rewards distribution is longer (i.e., bigger $\Delta T$, hence more chances of receiving informative rewards), as shown by the scaling of their performances with respect to $\Delta T$ in Figure 6.5.

For $\Delta T = 2000$ (Table 6.10b), MAB and RMAB attain, respectively, 92% and 89% of the maximum TCR, with a significant advantage to MAB with respect to RMAB and to all the other techniques: the standard MAB is able of efficiently recognizing and exploiting the best operator (Figure 6.6c). For $\Delta T = 500$, however, RMAB and AP are not significantly different from MAB, again the winner. Concerning DMAB, differently from the former two artificial scenarios, in this case its restarting mechanism is not able to provide good performance: given the high variance of the rewards received, it is very difficult to find a good value for the Page-Hinkley change-detection threshold $\gamma$, and this results in the triggering of several misplaced restarts, as shown in Figures 6.6a and 6.6b. The DMAB is only able to outperform SLMAB, in terms of both TCR and $p(\text{best})$ measures, for the two longest epochs, and AP (only with respect to TCR) for the longest epoch. While AP has its performance hindered again by $p_{\text{min}}=0.05$ (Figure 6.6f), SLMAB fails to efficiently recognize the best operator in the first epoch, takes a long time to adapt to new situations and, finally, it is not able to exploit the best operator more than 80% of the times even at the end of a steady-state epoch as long as 2000 time steps (Figure 6.6d).

Regarding the Credit Assignment schemes, here, the Absolute Extreme is clearly the best, as could be expected: when an outlier value is triggered, this scheme will maintain this operator in some top position longer than any other scheme. Interestingly, the Instantaneous reward is a complete disaster for AP, while maintaining a fair level of performance (at least similar to that of the Average reward) for the bandit-based techniques: some average actually takes place in the computation of $\hat{q}$ on the bandit-based approaches (see Equation 5.10), hence keeping some memory of the outlier value; while the benefit of such value vanishes more rapidly within the two-tiered mechanism of AP. It is also clear that Normalization does not work at all here, as it impacts on very few time steps (most rewards are 0), while hiding the outlier effect by bringing the extreme value of the reward back to 1. On the other hand, the rank-based Credit Assignment schemes show
Figure 6.5: Scaling of mean performance (TCR above, and \( p(\text{best}) \) below) in relation to the epoch length \( \Delta T \), for each Operator Selection technique with its best Credit Assignment scheme, on the Outlier scenario.

their value in this scenario: all the four different variants are not significantly different, being also equivalent with respect to the global best method in most epoch lengths; the only exception is the longest epoch, in which the different combinations with RMAB (Figure 6.6e) are still ranked second, but with a significant difference in relation to the best (AbsExt-MAB). Indeed, as discussed in Section 5.2.4, the use of the rank-based schemes with decay factor tries to mimic, in a smoother way, the intuition of the Extreme Credit Assignment; this explains their good performance on this scenario.
### 6.3 On Artificial Scenarios

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>1633 ± 284 ▲</td>
<td>1673 ± 244 ▲</td>
<td>1649 ± 283 ▲</td>
<td>1601 ± 249 ▲</td>
</tr>
<tr>
<td>Pb: 25.5 ± 8.0</td>
<td>C5W10</td>
<td>C10G100W1</td>
<td>C10W1</td>
<td>P.05A.1B.9W1</td>
</tr>
<tr>
<td>NormIns</td>
<td>1541 ± 254 ▲</td>
<td>1541 ± 254 ▲</td>
<td>1541 ± 254 ▲</td>
<td>1545 ± 255 ▲</td>
</tr>
<tr>
<td>Pb: 20.5 ± 3.6</td>
<td>C1W100</td>
<td>C.1G10W1</td>
<td>C.1W1</td>
<td>P.0A.3B.9W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>1677 ± 264 ▲</td>
<td>1650 ± 286 ▲</td>
<td>1647 ± 264 ▲</td>
<td>1607 ± 223 ▲</td>
</tr>
<tr>
<td>Pb: 26.8 ± 7.3</td>
<td>C10W10</td>
<td>C5G10W10</td>
<td>C10W10</td>
<td>P.1A.6B.3W10</td>
</tr>
<tr>
<td>NormAvg</td>
<td>1586 ± 217 ▲</td>
<td>1591 ± 205 ▲</td>
<td>1561 ± 284 ▲</td>
<td>1603 ± 278 ▲</td>
</tr>
<tr>
<td>Pb: 22.0 ± 3.4</td>
<td>C5W10</td>
<td>C1G1W10</td>
<td>C1W500</td>
<td>P.0A.9B.9W500</td>
</tr>
<tr>
<td>AbsExt</td>
<td>1722 ± 236 ★</td>
<td>1697 ± 255 ★</td>
<td>1697 ± 255 ★</td>
<td>1617 ± 232 ★</td>
</tr>
<tr>
<td>Pb: 28.3 ± 5.9</td>
<td>C100W10</td>
<td>C100G100W10</td>
<td>C100W10</td>
<td>P.1A.6B.3W10</td>
</tr>
<tr>
<td>NormExt</td>
<td>1575 ± 269 ▲</td>
<td>1568 ± 336 ▲</td>
<td>1550 ± 236 ▲</td>
<td>1616 ± 244 ▲</td>
</tr>
<tr>
<td>Pb: 22.1 ± 4.2</td>
<td>C5W10</td>
<td>C.01G1W10</td>
<td>C5W10</td>
<td>P.1A.9B.3W10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMAB (ΔF)</td>
<td>1634 ± 273 ▲</td>
<td>1640 ± 244 ▲</td>
<td>1599 ± 255 ▲</td>
<td>1619 ± 229 ▲</td>
</tr>
<tr>
<td>Pb: 26.6 ± 6.5</td>
<td>C1D.75W50</td>
<td>C.1D.25W500</td>
<td>C1W50</td>
<td>C1W50</td>
</tr>
</tbody>
</table>

(a) Results on the Outlier scenario, for ΔT=50 (Naive TCR: 1500, Optimal TCR: 2500)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>7210 ± 650 ▲</td>
<td>7085 ± 643 ▲</td>
<td>7085 ± 643 ▲</td>
<td>6758 ± 599</td>
</tr>
<tr>
<td>Pb: 33.8 ± 9.5</td>
<td>C10W50</td>
<td>C10G100W1</td>
<td>C10W1</td>
<td>P.05A.1B.6W1</td>
</tr>
<tr>
<td>NormIns</td>
<td>6150 ± 685</td>
<td>6090 ± 1003</td>
<td>6082 ± 1031</td>
<td>6203 ± 667</td>
</tr>
<tr>
<td>Pb: 20.8 ± 5.5</td>
<td>C.1W10</td>
<td>C.01G1W1</td>
<td>C.01W1</td>
<td>P.0A.1B.1W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>7270 ± 707 ▲</td>
<td>7231 ± 606 ▲</td>
<td>7231 ± 606 ▲</td>
<td>6896 ± 730</td>
</tr>
<tr>
<td>Pb: 38.3 ± 10.0</td>
<td>C10W50</td>
<td>C10G100W10</td>
<td>C10W10</td>
<td>P.0A.9B.9W50</td>
</tr>
<tr>
<td>NormAvg</td>
<td>6419 ± 673</td>
<td>6710 ± 659</td>
<td>6656 ± 803</td>
<td>6881 ± 604</td>
</tr>
<tr>
<td>Pb: 24.3 ± 6.9</td>
<td>C5W50</td>
<td>C1G1W50</td>
<td>C1W100</td>
<td>P.0A.6B.9W50</td>
</tr>
<tr>
<td>AbsExt</td>
<td>7288 ± 662 ★</td>
<td>7170 ± 527 ★</td>
<td>7329 ± 466 ★</td>
<td>7449 ± 641 ★</td>
</tr>
<tr>
<td>Pb: 41.3 ± 8.1</td>
<td>C100W50</td>
<td>C100G100W10</td>
<td>C100W50</td>
<td>P.0A.9B.6W50</td>
</tr>
<tr>
<td>NormExt</td>
<td>6801 ± 779 ▲</td>
<td>7086 ± 955 ▲</td>
<td>7086 ± 955 ▲</td>
<td>7429 ± 608 ▲</td>
</tr>
<tr>
<td>Pb: 27.7 ± 11.8</td>
<td>C1W50</td>
<td>C1G100W50</td>
<td>C1W50</td>
<td>P.0A.9B.9W50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMAB (ΔF)</td>
<td>7481 ± 633 ★</td>
<td>7223 ± 625 ▲</td>
<td>7190 ± 521 ▲</td>
<td>7275 ± 586 ▲</td>
</tr>
<tr>
<td>Pb: 42.0 ± 7.0</td>
<td>C1D.75W100</td>
<td>C.1D.5W100</td>
<td>C1W100</td>
<td>C1W100</td>
</tr>
</tbody>
</table>

(b) Results on the Outlier scenario, for ΔT=200 (Naive TCR: 6000, Optimal TCR: 10000)

Table 6.9: Results on the Outlier scenario for ΔT ∈ {50, 200}
Chapter 6. Experimental Results

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>19379 ± 1274</td>
<td>18896 ± 1270</td>
<td>18920 ± 1279</td>
<td>17391 ± 1145</td>
</tr>
<tr>
<td></td>
<td>Pb: 42.4 ± 9.8</td>
<td>Pb: 37.3 ± 7.5</td>
<td>Pb: 37.5 ± 7.7</td>
<td>Pb: 29.1 ± 4.6</td>
</tr>
<tr>
<td>NormIns</td>
<td>15418 ± 1705</td>
<td>15289 ± 1892</td>
<td>15135 ± 1851</td>
<td>15455 ± 829</td>
</tr>
<tr>
<td></td>
<td>C1W50</td>
<td>C1G1W1</td>
<td>C1W1</td>
<td>P.05A.1B.1W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 21.6 ± 9.1</td>
<td>Pb: 21.0 ± 5.8</td>
<td>Pb: 20.4 ± 5.6</td>
<td>Pb: 21.5 ± 2.1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>19156 ± 1227</td>
<td>19227 ± 1192</td>
<td>19222 ± 1194</td>
<td>18451 ± 1199</td>
</tr>
<tr>
<td></td>
<td>C5W50</td>
<td>C1G1000W10</td>
<td>C1W10</td>
<td>P.05A.9B.6W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 37.3 ± 10.7</td>
<td>Pb: 40.4 ± 8.3</td>
<td>Pb: 40.5 ± 8.4</td>
<td>Pb: 34.4 ± 6.6</td>
</tr>
<tr>
<td>NormAvg</td>
<td>17077 ± 1909</td>
<td>17950 ± 1327</td>
<td>17665 ± 1607</td>
<td>18500 ± 996</td>
</tr>
<tr>
<td></td>
<td>C1W50</td>
<td>C1G1W50</td>
<td>C1W100</td>
<td>P.05A.9B.1W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 26.7 ± 10.4</td>
<td>Pb: 31.7 ± 6.5</td>
<td>Pb: 28.3 ± 7.7</td>
<td>Pb: 35.0 ± 6.3</td>
</tr>
<tr>
<td>AbsExt</td>
<td>19191 ± 1164</td>
<td>19461 ± 1042</td>
<td>20734 ± 1052</td>
<td>20491 ± 1128</td>
</tr>
<tr>
<td></td>
<td>C100W50</td>
<td>C100G1000W50</td>
<td>C100W50</td>
<td>P.05A.9B.9W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 45.0 ± 6.5</td>
<td>Pb: 49.4 ± 3.3</td>
<td>Pb: 61.2 ± 4.6</td>
<td>Pb: 51.7 ± 5.4</td>
</tr>
<tr>
<td>NormExt</td>
<td>18130 ± 1990</td>
<td>19500 ± 1227</td>
<td>19038 ± 1747</td>
<td>20362 ± 1151</td>
</tr>
<tr>
<td></td>
<td>C1W50</td>
<td>C1G.1W50</td>
<td>C1W50</td>
<td>P.05A.9B.9W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 29.1 ± 11.7</td>
<td>Pb: 45.5 ± 7.7</td>
<td>Pb: 36.1 ± 11.3</td>
<td>Pb: 55.6 ± 5.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>19897 ± 982</td>
<td>20010 ± 993</td>
<td>20053 ± 981</td>
<td>20012 ± 1095</td>
</tr>
<tr>
<td></td>
<td>C1D.25W100</td>
<td>C1D.5W100</td>
<td>C1W100</td>
<td>C1W100</td>
</tr>
<tr>
<td></td>
<td>Pb: 53.0 ± 5.9</td>
<td>Pb: 53.4 ± 5.6</td>
<td>Pb: 54.4 ± 6.4</td>
<td>Pb: 53.4 ± 4.9</td>
</tr>
</tbody>
</table>

(a) Results on the Outlier scenario, for \( \Delta T=500 \) (Naive TCR: 15000, Optimal TCR: 25000)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>79428 ± 3113</td>
<td>82658 ± 2319</td>
<td>81842 ± 2705</td>
<td>69735 ± 2054</td>
</tr>
<tr>
<td></td>
<td>Pb: 44.6 ± 7.2</td>
<td>Pb: 51.4 ± 7.5</td>
<td>Pb: 50.4 ± 7.8</td>
<td>Pb: 29.5 ± 2.5</td>
</tr>
<tr>
<td>NormIns</td>
<td>62286 ± 6064</td>
<td>60675 ± 6308</td>
<td>60298 ± 6874</td>
<td>61442 ± 2398</td>
</tr>
<tr>
<td></td>
<td>Pb: 21.0 ± 9.1</td>
<td>Pb: 20.3 ± 6.7</td>
<td>Pb: 19.8 ± 6.8</td>
<td>Pb: 20.9 ± 2.1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>80723 ± 3381</td>
<td>83515 ± 2814</td>
<td>81949 ± 2759</td>
<td>79059 ± 2394</td>
</tr>
<tr>
<td></td>
<td>Pb: 41.7 ± 10.6</td>
<td>Pb: 53.9 ± 7.4</td>
<td>Pb: 51.3 ± 7.7</td>
<td>Pb: 44.5 ± 5.7</td>
</tr>
<tr>
<td>NormAvg</td>
<td>72584 ± 3463</td>
<td>80043 ± 3641</td>
<td>78659 ± 2492</td>
<td>79026 ± 2363</td>
</tr>
<tr>
<td></td>
<td>C5W50</td>
<td>C1G1W100</td>
<td>C5W100</td>
<td>P.05A.6B.1W100</td>
</tr>
<tr>
<td></td>
<td>Pb: 37.9 ± 7.6</td>
<td>Pb: 44.0 ± 8.0</td>
<td>Pb: 44.5 ± 5.3</td>
<td>Pb: 44.1 ± 5.8</td>
</tr>
<tr>
<td>AbsExt</td>
<td>86372 ± 2602</td>
<td>87699 ± 2108</td>
<td>92119 ± 1982</td>
<td>86595 ± 2035</td>
</tr>
<tr>
<td></td>
<td>Pb: 60.8 ± 9.1</td>
<td>Pb: 68.4 ± 2.6</td>
<td>Pb: 81.2 ± 2.4</td>
<td>Pb: 70.6 ± 3.1</td>
</tr>
<tr>
<td>NormExt</td>
<td>75437 ± 1877</td>
<td>87978 ± 3308</td>
<td>85638 ± 5590</td>
<td>86474 ± 2039</td>
</tr>
<tr>
<td></td>
<td>C5W50</td>
<td>C1G.1W50</td>
<td>C1W100</td>
<td>P.05A.9B.1W50</td>
</tr>
<tr>
<td></td>
<td>Pb: 37.9 ± 1.7</td>
<td>Pb: 64.2 ± 6.8</td>
<td>Pb: 53.9 ± 12.3</td>
<td>Pb: 70.1 ± 3.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OpSel/Credit</th>
<th>AUC (Decay)</th>
<th>SR (Decay)</th>
<th>AUC (NDCG)</th>
<th>SR (NDCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>89348 ± 2506</td>
<td>88547 ± 2801</td>
<td>89137 ± 3095</td>
<td>88785 ± 3214</td>
</tr>
<tr>
<td></td>
<td>C1D.5D.75/100</td>
<td>C1D.5D.5W100</td>
<td>C1W100</td>
<td>C1W100</td>
</tr>
<tr>
<td></td>
<td>Pb: 72.8 ± 4.4</td>
<td>Pb: 70.2 ± 6.9</td>
<td>Pb: 71.8 ± 6.8</td>
<td>Pb: 70.5 ± 7.5</td>
</tr>
</tbody>
</table>

(b) Results on the Outlier scenario, for \( \Delta T=2000 \) (Naive TCR: 60000, Optimal TCR: 100000)

Table 6.10: Results on the Outlier scenario for \( \Delta T \in \{500, 2000\} \)
Figure 6.6: Behavior of DMAB, MAB, SLMAB, RMAB and AP, combined with their best Credit Assignment schemes, on the Outlier scenario with $\Delta T = 2000$. Here, given the high variance of the rewards, the restarts are not helpful: the overall winner is AbsExt-MAB, followed by the robust Decay/AUC-RMAB.
6.3.3 Results on ART Scenarios

As described in Section 5.4.3, ART scenarios take into account 2 operators with rewards coming from 2 different TV distributions. It is important to remember that each distribution is defined by two parameters: the reward $R$, and the probability $p$ of getting reward $R$ (reward $r = 1$ otherwise); the resulting instance is referred to as $\text{ART}(p_1, R_1, p_2, R_2)$.

While many ART scenarios with different levels of difficulty were investigated, only the two most representative ones will be considered in the following. It is important to highlight that, for both instances, the two longest epochs ($\Delta T \in \{500, 2000\}$) are used to check how fast each AOS scheme can adapt to a new situation after a long period of stability; hence, only one permutation of rewards is done ($01 \rightarrow 10$) in these cases, i.e., only two epochs of length $\Delta T$ are considered.

**Low Average/High Variance vs. High Average/Low Variance scenario**

The $\text{ART}(0.01, 101, 0.5, 10)$ problem involves a low average/high variance distribution ($E_1 = 2$, $V_1 = 99$) and a high average/low variance distribution ($E_2 = 6$, $V_2 = 20.25$) operators. Detailed results are presented in Tables 6.11 and 6.12, respectively, for $\Delta T \in \{50, 200\}$, and for $\Delta T \in \{500, 2000\}$.

Indeed, the fact that the high-variance operator is also the one with lower reward expectation should make this operator to be easily discarded. Given this clearness in the reward distribution, the Absolute version of all the three kinds of Credit Assignment based on the raw values of fitness improvements are able to achieve good performance for all epoch lengths, with higher TCR values being attained by Operator Selection techniques employing the Absolute Extreme (AbsExt) Credit Assignment. Accordingly, the rank-based AUC Credit Assignment schemes also perform well, significantly better than the SR-based variants, with a small (although significant in most cases) difference between its linear (Decay with $D = 1$) and NDCG (equivalent to Decay with $D = 0.4$) variants. This confirms again the fact that, when there are only few possible reward values, the decay factor does not matter much in the distinction between the qualities of the operators.

Figure 6.7 shows how the performance of each Operator Selection technique, with its corresponding best Credit Assignment scheme, scales with respect to the epoch length $\Delta T$. As can be seen, the performance ranking of the Operator Selection techniques is very clear, in terms of both TCR and $p(\text{best})$ measures. The overall winner is again DMAB, combined with the AbsExt Credit Assignment, which precisely performs restarts every time a change occurs (Figures 6.8a and 6.8b), supported by its well-tuned PH change-detection test. Although having one hyper-parameter less, the SLMAB performs not significantly different from the winner DMAB in all cases, with its sliding update rule quickly adapting to the new situation (Figures 6.8c and 6.8d). RMAB comes in third place, combined with AUC with linear decay in all cases, one more time confirming that the adaptation being done on the Credit Assignment side is rather efficient (Figures 6.8e and 6.8f). It significantly outperforms the MAB and the AP methods for all epoch lengths, except for the longest epoch, in which the best configuration for AP becomes not significantly different with respect to all techniques, due to the very high standard deviation on its
6.3 On Artificial Scenarios

Figure 6.7: Scaling of mean performance (TCR above, and \( p(\text{best}) \) below) in relation to the epoch length \( \Delta T \), for each Operator Selection technique with its best Credit Assignment scheme, on the ART\((0.01, 101, 0.5, 10)\) scenario.

TCR performance (though with a lower mean). The standard MAB and the baseline AP Operator Selection techniques succeed in following the dynamics of the scenario, but their performances are greatly affected by the facts that their adaptation is slower than those of the others, and they are not able to exploit the best operator up to the maximal rate. By analyzing the behavior plots, it is also interesting to see the gain, in terms of speed of adaptation and exploitation efficiency, provided by the different bandit-based extensions with respect to the original MAB technique (Figures 6.8g and 6.8h).
### Chapter 6. Experimental Results

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td><strong>2559 ± 137</strong></td>
<td><strong>2461 ± 140</strong></td>
<td><strong>2423 ± 142</strong></td>
<td><strong>2341 ± 164</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td><strong>C5G10W1</strong></td>
<td><strong>C10W1</strong></td>
<td><strong>P.1A.9B.9W1</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Pb: 12.9 ± 3.8</strong></td>
<td><strong>Pb: 18.8 ± 2.3</strong></td>
<td><strong>Pb: 20.4 ± 4.6</strong></td>
<td><strong>Pb: 23.0 ± 5.3</strong></td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>2148 ± 145</td>
<td>2401 ± 162</td>
<td>2401 ± 162</td>
<td>2246 ± 194</td>
</tr>
<tr>
<td></td>
<td>C5W100</td>
<td>C1G10W1</td>
<td>C1W1</td>
<td>P.1A.9B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 36.2 ± 2.6</td>
<td>Pb: 22.0 ± 3.0</td>
<td>Pb: 22.0 ± 3.0</td>
<td>Pb: 27.8 ± 6.2</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td><strong>2513 ± 131</strong></td>
<td><strong>2447 ± 192</strong></td>
<td><strong>2304 ± 133</strong></td>
<td><strong>2277 ± 156</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td><strong>C1G10W10</strong></td>
<td><strong>C10W10</strong></td>
<td><strong>P.1A.9B.9W10</strong></td>
</tr>
<tr>
<td></td>
<td>Pb: 15.4 ± 4.5</td>
<td>Pb: 19.2 ± 7.0</td>
<td>Pb: 27.0 ± 4.9</td>
<td>Pb: 26.2 ± 5.3</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>2139 ± 251</td>
<td>2348 ± 156</td>
<td>2348 ± 156</td>
<td>2186 ± 168</td>
</tr>
<tr>
<td></td>
<td>C1W50</td>
<td>C1G10W10</td>
<td>C1W1</td>
<td>P.1A.9B.9W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 36.0 ± 10.0</td>
<td>Pb: 25.1 ± 3.5</td>
<td>Pb: 25.1 ± 3.5</td>
<td>Pb: 31.1 ± 6.2</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td><strong>2465 ± 114</strong></td>
<td><strong>2538 ± 148</strong></td>
<td><strong>2234 ± 205</strong></td>
<td><strong>2276 ± 149</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C10W10</strong></td>
<td><strong>C1G1W10</strong></td>
<td><strong>C10W10</strong></td>
<td><strong>P.1A.6B.9W10</strong></td>
</tr>
<tr>
<td></td>
<td>Pb: 18.0 ± 5.8</td>
<td>Pb: 14.0 ± 6.5</td>
<td>Pb: 30.1 ± 12.7</td>
<td>Pb: 26.3 ± 4.5</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>2141 ± 192</td>
<td>2335 ± 149</td>
<td>2332 ± 149</td>
<td>2207 ± 166</td>
</tr>
<tr>
<td></td>
<td>C5W10</td>
<td>C1G1W10</td>
<td>C1W1</td>
<td>P.1A.6B.9W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 35.1 ± 4.7</td>
<td>Pb: 25.4 ± 4.8</td>
<td>Pb: 25.4 ± 4.8</td>
<td>Pb: 29.9 ± 5.8</td>
</tr>
<tr>
<td><strong>OpSel/Credit</strong></td>
<td><strong>AUC (Decay)</strong></td>
<td><strong>SR (Decay)</strong></td>
<td><strong>AUC (NDCG)</strong></td>
<td><strong>SR (NDCG)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>2508 ± 136</strong></td>
<td><strong>2371 ± 153</strong></td>
<td><strong>2499 ± 129</strong></td>
<td><strong>2365 ± 131</strong></td>
</tr>
<tr>
<td></td>
<td>C.01D1W10</td>
<td>C1D.9W10</td>
<td>C.01W10</td>
<td>C1W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 15.1 ± 1.6</td>
<td>Pb: 23.3 ± 1.3</td>
<td>Pb: 15.3 ± 1.9</td>
<td>Pb: 22.7 ± 1.5</td>
</tr>
</tbody>
</table>

(a) Results $\mathcal{A}_{RT}(0.01, 101, 0.5, 10)$ scenario, for $\Delta T = 50$ (Naive TCR: 2000, Optimal TCR: 3000)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AbsIns</strong></td>
<td><strong>10656 ± 226</strong></td>
<td><strong>10348 ± 243</strong></td>
<td><strong>10182 ± 243</strong></td>
<td><strong>10204 ± 297</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td><strong>C5G10W1</strong></td>
<td><strong>C10W1</strong></td>
<td><strong>P.05A.6B.9W1</strong></td>
</tr>
<tr>
<td></td>
<td>Pb: 5.1 ± 1.4</td>
<td>Pb: 10.1 ± 2.7</td>
<td>Pb: 11.6 ± 1.9</td>
<td>Pb: 11.8 ± 2.8</td>
</tr>
<tr>
<td><strong>NormIns</strong></td>
<td>8635 ± 278</td>
<td>10012 ± 273</td>
<td>10012 ± 273</td>
<td>9921 ± 296</td>
</tr>
<tr>
<td></td>
<td>C10W500</td>
<td>C1G10W10</td>
<td>C1W1</td>
<td>P.1A.6B.9W1</td>
</tr>
<tr>
<td></td>
<td>Pb: 33.4 ± 0.8</td>
<td>Pb: 14.1 ± 1.8</td>
<td>Pb: 14.1 ± 1.8</td>
<td>Pb: 15.7 ± 2.9</td>
</tr>
<tr>
<td><strong>AbsAvg</strong></td>
<td><strong>10637 ± 229</strong></td>
<td><strong>10483 ± 262</strong></td>
<td><strong>10063 ± 260</strong></td>
<td><strong>10159 ± 303</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C5W10</strong></td>
<td><strong>C1G1W10</strong></td>
<td><strong>C10W10</strong></td>
<td><strong>P.05A.9B.9W10</strong></td>
</tr>
<tr>
<td></td>
<td>Pb: 5.2 ± 1.4</td>
<td>Pb: 7.5 ± 2.7</td>
<td>Pb: 13.5 ± 2.3</td>
<td>Pb: 12.3 ± 3.3</td>
</tr>
<tr>
<td><strong>NormAvg</strong></td>
<td>8777 ± 260</td>
<td>9899 ± 442</td>
<td>9899 ± 442</td>
<td>9875 ± 258</td>
</tr>
<tr>
<td></td>
<td>C5W50</td>
<td>C1G10W10</td>
<td>C1W1</td>
<td>P.1A.9B.9W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 32.3 ± 2.4</td>
<td>Pb: 15.7 ± 4.8</td>
<td>Pb: 15.7 ± 4.8</td>
<td>Pb: 16.4 ± 2.4</td>
</tr>
<tr>
<td><strong>AbsExt</strong></td>
<td><strong>10680 ± 216</strong></td>
<td><strong>10719 ± 221</strong></td>
<td><strong>9322 ± 1010</strong></td>
<td><strong>10149 ± 300</strong></td>
</tr>
<tr>
<td></td>
<td><strong>C10W10</strong></td>
<td><strong>C5G10W10</strong></td>
<td><strong>C10W10</strong></td>
<td><strong>P.05A.9B.9W10</strong></td>
</tr>
<tr>
<td></td>
<td>Pb: 4.6 ± 1.7</td>
<td><strong>Pb: 3.9 ± 0.4</strong></td>
<td>Pb: 23.9 ± 15.5</td>
<td>Pb: 12.4 ± 3.3</td>
</tr>
<tr>
<td><strong>NormExt</strong></td>
<td>8503 ± 278</td>
<td>9778 ± 399</td>
<td>9778 ± 399</td>
<td>9893 ± 270</td>
</tr>
<tr>
<td></td>
<td>C5W50</td>
<td>C1G10W10</td>
<td>C1W1</td>
<td>P.1A.6B.9W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 35.5 ± 3.6</td>
<td>Pb: 17.5 ± 5.1</td>
<td>Pb: 17.5 ± 5.1</td>
<td>Pb: 16.3 ± 2.5</td>
</tr>
<tr>
<td><strong>OpSel/Credit</strong></td>
<td><strong>AUC (Decay)</strong></td>
<td><strong>SR (Decay)</strong></td>
<td><strong>AUC (NDCG)</strong></td>
<td><strong>SR (NDCG)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>10419 ± 262</strong></td>
<td><strong>10052 ± 280</strong></td>
<td><strong>10160 ± 250</strong></td>
<td><strong>9977 ± 259</strong></td>
</tr>
<tr>
<td></td>
<td>C.5D1W100</td>
<td>C.5D9W10</td>
<td>C.01W10</td>
<td>C.5W10</td>
</tr>
<tr>
<td></td>
<td>Pb: 8.3 ± 1.5</td>
<td>Pb: 12.8 ± 1.4</td>
<td>Pb: 11.6 ± 0.6</td>
<td>Pb: 13.9 ± 2.1</td>
</tr>
</tbody>
</table>

(b) Results $\mathcal{A}_{RT}(0.01, 101, 0.5, 10)$ scenario, for $\Delta T = 200$ (Naive TCR: 8000, Optimal TCR: 12000)

Table 6.11: Results on $\mathcal{A}_{RT}(0.01, 101, 0.5, 10)$, 10 epochs for $\Delta T \in \{50, 200\}$
### 6.3 On Artificial Scenarios

![Image of a table with data](image)
Figure 6.8: Behavior of DMAB, SLMAB, RMAB and MAB, combined with their best Credit Assignment schemes, on the $\mathcal{ART}(0.01, 101, 0.5, 10)$ scenario, for $\Delta T = 200$ on the left column, and $\Delta T = 2000$ on the right column.
6.3 On Artificial Scenarios

High Average/High Variance vs. Low Average/Low Variance scenario

Oppositely to the previous ART scenario, the ART(0.1,39,0.5,3) problem has a high average/high variance (\(E_1 = 4.8, V_1 = 130\)) and a low average/low variance (\(E_2 = 2\) and \(V_2 = 1\)) operators. Detailed results are depicted in Tables 6.13 and 6.14, respectively, for \(\Delta T \in \{50,200\}\) and \(\Delta T \in \{500,2000\}\).

This scenario is much more complex than the previous one, as the regularity of the second operator might lead the AOS schemes to believe it is the best operator, while in fact the best is the first one. This kind of situation was the main motivation for the proposal of the Extreme-based Credit Assignment, as discussed in Section 5.2.2, which indeed performs significantly better than all the other schemes for the three longest epoch lengths, while also being the winner (but not significantly different from several others) for \(\Delta T = 50\). As expected, the Instantaneous schemes achieve a much lower performance: as they assign credit based on a single operator application, they need to be very lucky in order to catch the outlier reward that is received only in 10% of the cases for the first operator. This situation is alleviated by the Average Credit Assignment, but not enough to provide good performance. As in the Outlier scenario, the Normalized variants of these schemes do not work at all: the very different reward values of 3 and 39 given by each of the operators result in the same credit value of 1 in different moments of the search, as discussed in Section 5.2.4.

Surprisingly, the different AOS combinations involving the rank-based Credit Assignment schemes with the RMAB Operator Selection technique do not work well at all in this case, even when employing a strong decaying factor (what intuitively approaches it to the behavior of the Extreme Credit Assignment scheme, as discussed in Section 5.2.4). A tentative interpretation, based on the high variation of the instant selection rates depicted in Figures 6.10e and 6.10f, can be developed as follows. In the Extreme Credit Assignment, when the outlier operator receives the high reward (\(R = 39\) in this case), an equivalent credit of 39 is assigned to this operator at least \(W\) times, no matter how many bad rewards it receives during this period, and no matter how many times the other (regular) operator is applied, as there is a separate window for the rewards received by each operator. In the rank-based schemes, as there is only one sliding window for all operators, this dominance is greatly reduced by two factors: (i) the applications of the other operator, which, besides pushing the outlier reward out of the window, will always reduce the overall quality estimate of the outlier operator; and (ii) the receiving of bad rewards by the outlier operator (what happens in 90% of the times in this case), will also affect its quality estimate, consequently promoting the exploration of the other operator. And then, given the regularity of the other operator, one trial might be enough to start a long period of dominance.

Anyway, as for the other scenarios, the longer the steady-state epoch, the better the performance for all the techniques (even for the rank-based ones) with respect to the Naive approach, as summarized in Figure 6.9. Notably, both DMAB and SLMAB reach a quasi-perfect score when \(\Delta T = 2000\). The DMAB is once more the overall winner for all epoch lengths, but not significantly different from the second and third best techniques, namely, the SLMAB and the standard MAB, all of them combined with the AbsExt.
Figure 6.9: Scaling of mean performance (TCR above, and $p(\text{best})$ below) in relation to the epoch length $\Delta T$, for each Operator Selection technique with its best Credit Assignment scheme, on the $\text{ART}(0.1, 39, 0.5, 3)$ scenario.

Credit Assignment scheme. The baseline AP is significantly beaten by all the bandit-based approaches, with the exception of RMAB, which performs equally bad, for the reasons previously discussed. The behavior plots of all the winner configurations for the bandit-based approaches are presented in Figure 6.10.
### 6.3 On Artificial Scenarios

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>1843 ± 272 ▲</td>
<td>1864 ± 256</td>
<td>1862 ± 255 ▲</td>
<td>1737 ± 246 ▲</td>
</tr>
<tr>
<td>Pb: 60.1 ± 7.1</td>
<td>C10W10</td>
<td>C10G100W1</td>
<td>C10W1</td>
<td>P.1A.1B.3W1</td>
</tr>
<tr>
<td>NormIns</td>
<td>1720 ± 180</td>
<td>1714 ± 177</td>
<td>1687 ± 157</td>
<td>1718 ± 160 ▲</td>
</tr>
<tr>
<td>Pb: 49.7 ± 1.8</td>
<td>C100W10</td>
<td>C10G.01W1</td>
<td>C10W10</td>
<td>P0A.9B.1W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>1927 ± 269 ▲</td>
<td>1946 ± 247 ▲</td>
<td>1894 ± 228 ▲</td>
<td>1761 ± 229 ▲</td>
</tr>
<tr>
<td>Pb: 65.2 ± 7.1</td>
<td>C10W10</td>
<td>C10G100W10</td>
<td>C10W10</td>
<td>P.1A.6B.6W50</td>
</tr>
<tr>
<td>NormAvg</td>
<td>1782 ± 202</td>
<td>1757 ± 194</td>
<td>1757 ± 194</td>
<td>1768 ± 213 ▲</td>
</tr>
<tr>
<td>Pb: 54.4 ± 3.3</td>
<td>C5W50</td>
<td>C10G100W10</td>
<td>C10W10</td>
<td>P.1A.9B.6W50</td>
</tr>
<tr>
<td>AbsExt</td>
<td>2012 ± 301 ★</td>
<td>2040 ± 238 ★</td>
<td>2036 ± 239 ★</td>
<td>1793 ± 252 ★</td>
</tr>
<tr>
<td>Pb: 71.1 ± 5.9</td>
<td>C100W10</td>
<td>C100G1000W10</td>
<td>C100W10</td>
<td>P.1A.1B.6W10</td>
</tr>
<tr>
<td>NormExt</td>
<td>1772 ± 196</td>
<td>1758 ± 218</td>
<td>1758 ± 218</td>
<td>1743 ± 240 ▲</td>
</tr>
<tr>
<td>Pb: 55.1 ± 5.2</td>
<td>C5W50</td>
<td>C5G100W10</td>
<td>C5W10</td>
<td>P.05A.1B.1W10</td>
</tr>
<tr>
<td>OpSel/Credit</td>
<td>AUC (Decay)</td>
<td>SR (Decay)</td>
<td>AUC (NDCG)</td>
<td>SR (NDCG)</td>
</tr>
<tr>
<td>RMAB (ΔF)</td>
<td>1934 ± 245 ★</td>
<td>1860 ± 199 ★</td>
<td>1844 ± 241 ★</td>
<td>1869 ± 227 ★</td>
</tr>
</tbody>
</table>

(a) Results ART(0.1, 0.39, 0.5, 3) scenario, for ΔT = 50 (Naive TCR: 1700, Optimal TCR: 2400)

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsIns</td>
<td>8337 ± 537</td>
<td>8232 ± 475</td>
<td>7996 ± 572</td>
<td>7277 ± 639</td>
</tr>
<tr>
<td>Pb: 77.5 ± 3.8</td>
<td>C10W50</td>
<td>C10G100W1</td>
<td>C10W1</td>
<td>P.2A.1B.1W1</td>
</tr>
<tr>
<td>NormIns</td>
<td>6961 ± 628</td>
<td>6830 ± 361</td>
<td>6792 ± 340</td>
<td>6820 ± 379</td>
</tr>
<tr>
<td>Pb: 51.7 ± 8.8</td>
<td>C1W50</td>
<td>C.1G.01W1</td>
<td>C.01W1</td>
<td>P0A.9B.6W1</td>
</tr>
<tr>
<td>AbsAvg</td>
<td>8522 ± 497 ▲</td>
<td>8344 ± 538</td>
<td>8160 ± 466</td>
<td>7795 ± 741 ▲</td>
</tr>
<tr>
<td>Pb: 79.4 ± 3.3</td>
<td>C10W50</td>
<td>C10G100W10</td>
<td>C10W10</td>
<td>P.1A.6B.3W50</td>
</tr>
<tr>
<td>NormAvg</td>
<td>7700 ± 450</td>
<td>8110 ± 581</td>
<td>7871 ± 790</td>
<td>7780 ± 526 ▲</td>
</tr>
<tr>
<td>Pb: 65.6 ± 2.6</td>
<td>C5W100</td>
<td>C1G.1W50</td>
<td>C1W50</td>
<td>P.2A.9B.6W50</td>
</tr>
<tr>
<td>AbsExt</td>
<td>8746 ± 511 ★</td>
<td>8830 ± 478 ★</td>
<td>8555 ± 530 ★</td>
<td>7931 ± 606 ★</td>
</tr>
<tr>
<td>Pb: 84.3 ± 2.9</td>
<td>C100W50</td>
<td>C10G1W50</td>
<td>C10W10</td>
<td>P.1A.3B.6W50</td>
</tr>
<tr>
<td>NormExt</td>
<td>7944 ± 376</td>
<td>8046 ± 634</td>
<td>8046 ± 634</td>
<td>7830 ± 486 ▲</td>
</tr>
<tr>
<td>Pb: 71.1 ± 2.1</td>
<td>C5W50</td>
<td>C1G100W50</td>
<td>C1W50</td>
<td>P.2A.9B.6W50</td>
</tr>
<tr>
<td>OpSel/Credit</td>
<td>AUC (Decay)</td>
<td>SR (Decay)</td>
<td>AUC (NDCG)</td>
<td>SR (NDCG)</td>
</tr>
<tr>
<td>RMAB (ΔF)</td>
<td>8066 ± 667 ★</td>
<td>7887 ± 433 ★</td>
<td>7848 ± 701 ★</td>
<td>7819 ± 428 ★</td>
</tr>
</tbody>
</table>

(b) Results ART(0.1, 0.39, 0.5, 3) scenario, for ΔT = 200 (Naive TCR: 6800, Optimal TCR: 9600)

Table 6.13: Results ART(0.1, 0.39, 0.5, 3), 10 epochs for ΔT ∈ {50, 200}

129
Table 6.14: Results $\alphaRT(0.1, 0.39, 0.5, 3)$, 2 epochs for $\Delta T \in \{500, 2000\}$
Figure 6.10: Behavior of DMAB, SLMAB, RMAB and MAB, combined with their best Credit Assignment schemes, on the $\mathcal{ART}(0.1, 0.39, 0.5, 0.3)$ scenario, for $\Delta T = 200$ on the left column, and $\Delta T = 2000$ on the right column.
6.3.4 Discussion

The problems used in this Section provided conditions that are very artificial, with abrupt changes happening every $\Delta T$ time steps. In real optimization problems, these abrupt changes in the rewards distribution might also occur (e.g., when escaping from a local optima and reaching a new region of the search space); but globally, the dynamics of the operator qualities tend to be much more complex and usually unpredictable (except for the simple benchmark problems, such as the OneMax; see Section 6.4.2). However, the initial motivation for using these artificial scenarios is confirmed by the experimental findings: they enable the detailed analysis of the behavior of the AOS methods, and the verification of their characteristics in practice. The results presented in this Section can be summarized as follows.

The baseline probability-based AP method, although systematically (and significantly) outperforming the original Probability Matching (PM) method (results for PM are not shown here), still provides a slow adaptation when compared to all proposed bandit-based approaches. The main reason for this, in most cases, is the limitation provided by the enforced minimal level of exploration $p_{\text{min}}$. But, even when $p_{\text{min}}$ is set to zero, its two-tiered update mechanism needs some time in order to start to efficiently exploit the new best operator.

As for the bandit-based approaches, the standard MAB Operator Selection technique is able to follow the dynamics in an efficient way when the changes happen smoothly, i.e., when the magnitude of the adaptation to be done is small (e.g., in the Uniform scenario, when the second best operator becomes the best, or vice-versa). Whenever faster dynamics are considered, the DMAB succeeds in adapting very quickly to new situations, supported by its change-detection mechanism. Moreover, as originally expected, the SLMAB is able to perform as efficiently as the DMAB in most cases, due to its parameterless window-based relaxation mechanism.

The RMAB, using any of the proposed rank-based Credit Assignment schemes, outperforms the baseline AP and performs equivalently to the standard MAB. But the main benefit brought by these rank-based schemes (as well as by the Normalized versions of the Instantaneous, Average and Extreme schemes) is a higher robustness with respect to: (i) different values of rewards gathered in different stages of the search; and (ii) different (unknown) fitness ranges provided by different problems. Both issues were not assessed in this experimental set. Conversely, for each scenario, the same range of reward values was considered during all the optimization process. Besides, each scenario/epoch length was independently tackled after a preliminary off-line tuning phase, what explains why the very sensitive and problem-dependent (assumed after discussion in Sections 5.2.3 and 5.3.3, and confirmed by the sensitivity analysis that will be presented in Section 6.7) AbsExt-DMAB combination was found to be the winner in almost all the cases.

Indeed, the gain in robustness provided by RMAB results in a gain in performance, in relation to the other AOS methods, when several problems are tackled using the same hyper-parameter setting. This situation will be further stressed in the experiments that will be analyzed in the following, specially in the hyper-parameter sensitivity and robustness analysis considering different optimization benchmark problems (Section 6.7).
6.4 On Boolean Benchmark Problems

Some EA boolean benchmark problems were also used to empirically compare the AOS schemes, in situ, i.e., combined with an EA and selecting between actual evolutionary operators on some (still artificial) fitness landscapes with different complexities and levels of difficulty with respect to AOS. Needless to say, in these cases, the dynamics of the performance of the operators are not deterministically switched after every epoch, but rather depending on the search trajectory and on the fitness landscape.

Three different problems were considered, namely: the eternal OneMax problem, and two harder problems, the Long \( k \)-Path and the Royal Road. The experimental settings, in complement to the general settings presented in Section 6.2, will be described in Section 6.4.1. The problems will be presented, and the empirical results will be analyzed, in Sections 6.4.2, 6.4.3 and 6.4.4, respectively, for the OneMax, Long \( k \)-Path, and Royal Road. Finally, Section 6.4.5 will conclude this analysis with a discussion about the highlights of these experiments. The results that will be analyzed here were partially published in [Fialho et al., 2008; Fialho et al., 2009a; Fialho et al., 2010c].

6.4.1 Experimental Settings

The performance of the AOS schemes embedded within real EAs is assessed by the number of generations needed to achieve the optimal solution, the lower the better. The resulting total number of fitness evaluations can be roughly measured as the number of generations times the size of the offspring population. Besides the presentation of the detailed Tables with the average and standard deviation of the performance achieved by each of the considered AOS methods (e.g., Table 6.15a), the ECDFs are also used to compare the complete performance distribution for each of the winner techniques (e.g., Figure 6.15b).

The stopping criteria are: optimal solution found or maximum number of generations attained. For this latter, a value of 15,000 is used for the OneMax and Long \( k \)-Path problems, while 25,000 is used for the Royal Road problem. For the first two problems, the unique solution maintained in the population is initialized to \((0, \ldots, 0)\), while for the Royal Road, the population is uniformly initialized.

In addition to the AP Operator Selection technique, combined with all the Credit Assignment schemes based on raw values of fitness improvements, the optimal Oracle, the Naive uniform, and the off-line tuned Static operator selection strategies are used as baseline for both, OneMax and Long \( k \)-Path problems. For the Royal Road problem, all of them are used as well, except for the Oracle strategy: as the fitness landscape of this problem includes many paths toward the optimal solution, the Oracle can not be easily assessed.

More specific experimental settings, such as the definition of the EA used, as well as the set of operators automatically controlled by the AOS schemes, will be described in the following, together with the presentation of each problem.
6.4.2 The OneMax Problem

The OneMax problem involves an unimodal fitness function that simply counts the number of “1”s in the binary bit-string that represents the individual solution. The only difficulty comes from the size of the problem; in the presented experiments, the size $\ell$ of the bit-string is 10,000. Given its simplicity, it is often used to preliminary evaluate new empirical or theoretical methods, being for this reason considered as the “Drosophila of EC”.

In order to be assessed on this problem, the AOS schemes are combined with a standard $(1 + \lambda)$-EA ($\lambda$ offspring are created from the current parent; next parent is the best among the current offspring and parent). Different values for $\lambda$ were analyzed in [Fialho et al., 2008], achieving similar conclusions; $\lambda = 50$ is used here. The objective is to automatically select between some mutation operators, namely, the standard bit-flip operator (every bit is flipped with probability $1/\ell$), and a set of $b$-bit mutations (flipping exactly $b$ randomly chosen bits) with $b \in \{1, 3, 5\}$.

In many respects, the considered setting is still far from being realistic evolutionarily speaking: applying a $(1 + \lambda)$-EA, with $\lambda > 1$ and $b$-bit mutation operators, is meaningless on the OneMax problem (though it might make more sense on multi-core architectures). It nevertheless confronts the proposed and baseline approaches with the actual difficulties of taming a dynamic system, where the decisions made at a given moment govern the expected benefits of further decisions (the selected operators determine the position of the population and hence the improvement expectation of the operators at further stages of the search), as opposed to the artificial scenarios tackled in Section 6.3.

One main advantage of this kind of “sterile EC-like” environment is to enable the assessment of the AOS approaches by comparison with the performance of the known optimal operator selection. The optimal baseline is provided by the optimal behavior of all operators (computed by means of Monte-Carlo simulations). Figure 6.11a depicts the operator landscape from the perspective of a $(1 + 50)$-EA: for each fitness value of the unique parental individual, we report the fitness gain for the best out of 50 offsprings generated by each of the considered mutation operators, averaged over 100 runs. As can be seen, the trajectory of evolution involves distinct phases with respect to operator dynamics. In stable phases, the optimal operator remains the same (though its performance might decrease). For instance, the 5-bit mutation dominates all other operators until around $F(x) = 6579$, although its performance starts to gradually decrease after $F(x) = 5300$. In transition phases, the established best operator becomes dominated by another one; the 3-bit mutation outperforms the 5-bit after $F(x) = 6579$, and the 1-bit mutation outperforms the 3-bit after $F(x) = 8601$. The last phase is a desert, where hardly any operator brings any improvement; by being less disruptive, the 1-bit has higher chances of fine-tuning the solution towards the optimum, being thus the preferred operator at this phase.

A clearer view of the optimal behavior with respect to Operator Selection is presented in Figure 6.11b: at each stage of the search, according to the current fitness value of the parent, one of the operators is the best one, and should thus be applied at a rate of 100% (until the situation changes). This illustration indeed represents the behavior of the Oracle strategy that was used to achieve the empirical results for the optimal baseline.

Although being a rather simplistic scenario, this operator landscape provided by the
OneMax problem enables one to assess the basic skills of an AOS mechanism: the abilities (i) to pick up the best operator and stick to it in stability phases; (ii) to swiftly switch to the next best operator in transition phases; and (iii) to remain efficient during the desert phases. An empirical analysis on this scenario will now be presented.

**Empirical Results**

The detailed results for the OneMax problem, with the average number of generations (plus the standard deviation) to achieve the optimum, and the winner hyper-parameter configuration for each AOS combination, are presented in Table 6.15a. Complementarily,
Table 6.15b depicts the complete distribution of results for each Operator Selection technique, with its best Credit Assignment scheme and hyper-parameter configuration, in the form of Empirical Cumulative Distribution Functions (ECDFs).

The complete Naive strategy, that uniformly selects between the four available mutation operators, is able to find the optimum in 7955 generations in average. Another common approach is to tune off-line the application rates of each operator: the best Static strategy applies the 5-bit mutation operator at a rate of 20%, and the 1-bit at a rate of 80%, achieving the optimum in roughly 6206 generations. The Oracle strategy (depicted in Figure 6.11b), which represents the complete knowledge about the operator landscape, finds the optimum in 5136 generations in average.

The ECDF plot (Table 6.15b) is bounded on the right by the average performance of the Naive uniform approach. As can be seen in this Figure, 100% of the runs of all the winner AOS combinations achieve the optimum many generations before the Uniform does, all being significantly better than it in average. Besides, with a few exceptions (the combinations involving MAB and SLMAB), all the winner configurations for each Operator Selection technique are also able to significantly outperform the baseline method that employs off-line tuned Static probabilities.

Interestingly, several AOS methods are able to achieve a performance not significantly different from the Oracle strategy, namely: AP with both Normalized and Absolute versions of the Extreme Credit Assignment; and RMAB with any of the rank-based Credit Assignment methods based on ranks over the fitness improvements. The only exception is the NDCG/AUC, which also performs very well, but significantly worse, due to the tight standard deviations. In fact, as can be seen in the operator quality landscape depicted in Figure 6.11a, the 5-bit, 3-bit and 1/\ell bit-flip mutation operators are rather equivalent, starting from fitness 7000 up to around 9000. This explains why these AOS methods are able to achieve optimal performance by controlling the operator applications in very different ways for the fitness values within the mentioned interval. The Normalized Extreme (NormExt)-AP selects different operators in three very well-defined phases (Figure 6.12a), instead of four in the case of the Oracle, efficiently exploiting the 5-bit, then the 3-bit, and finally the 1-bit. The Decay/AUC-RMAB (Figure 6.12e) achieves basically the same performance by exploiting only the 5-bit mutation operator up to 100% at the initial stages of the search, and the 1-bit at the final stages, while for the mentioned fitness interval all the operators are equally explored to some extent. The best configuration for DMAB, implementing the AbsExt Credit Assignment, also achieves good performance, although significantly worse than the overall winner. As shown in Figure 6.12b, it also exploits the 5-bit in the initial stages, and then the 5-bit, 3-bit and bit-flip operators in the middle stages, by means of restarts; however, its performance is degraded by the fact that it is not able to maintain an optimal level of exploitation for the 1-bit operator during the final stages of the search. The MAB with NormExt is also able to efficiently follow the changes; but, in the same way as the DMAB, it gets lost during the final desert phase. A big deception in this scenario is the performance of all the AOS combinations considering the SLMAB Operator Selection technique, which shows rather poor performance: a tentative explanation is that it is designed to react very quickly to abrupt changes with respect to the operator qualities (as empirically verified in Section 6.3), while in the OneMax
problem the operator qualities tend to gradually decrease as the search goes on (Figure 6.11a).

It is also worth noticing that the best results are attained by AOS combinations using rather the NormExt or one of the rank-based Credit Assignment schemes. This empirical finding clearly confirms that, even in such a simplistic scenario as the OneMax problem, a robust Credit Assignment is important in order to achieve good performance: as discussed in Section 5.2.3, it prevents the AOS mechanism from the need of tackling an extra problem, the gradual reduction of the magnitude of the credits, that might greatly affect the performance of the AOS schemes (not to mention its robustness with respect to its hyper-parameters, that will be separately analyzed in Section 6.7). However, the comparison-based Credit Assignment schemes, i.e., the rank-based methods that assign ranks over the fitness values (F) instead of fitness improvements (Section 5.2.5), which are expected to be the most robust schemes over all, achieve a rather regular performance in this experimental scenario, although still significantly outperforming both Naive and Static baseline approaches. Figure 6.12f depicts the behavior of RMAB with the FAUC scheme. As can be seen, surprisingly, there are big variations (in both senses) in the operator selection rates; a tentative interpretation for this very noisy behavior goes as follows. Firstly, the fact that a (1 + 50)-EA setting is being used implies that even an improvement of 1 bit will generate a fitness value higher than all the values attained during the previous generation: if this happens in the beginning of the generation, such fitness value will be top-ranked in the Credit Assignment sliding window (what does not happen when considering fitness improvements), consequently leading to further trials for the operator used to generate it, being it really the best operator or not. Given the dynamics of the underlying algorithm, this situation, i.e., the exploration of a sub-optimal operator in the first trials of a new generation, could be considered to be rare. The problem in this case lies in the fact that the RMAB, in the way it is conceived (Section 5.3.4), enforces at least one application of each operator every \( W \) trials\(^1\). Hence, all \( k \) operators are explored in the initial \( k \) steps and once every \( W \) steps: as \( W = 100 \) and the offspring population size \( \lambda = 50 \), the initial \( k \) steps of every 2 generations will always be (coincidentally) exploration trials in this case, consequently leading to the noisy variations presented in the behavior plot roughly every 100 steps.

Finally, this experimental setting was also used to empirically compare the current version of the AUC method with the preliminary one (referred to as AUCv1). As discussed in Section 5.2.4, AUCv1 has normalization issues when considering several operators, and this results in a degraded performance, as presented in the last line of Table 6.15a; while the current version achieves optimal performance, as previously discussed. The behavior plots of RMAB with AUCv1 and FAUCv1 are presented, respectively, in Figures 6.12g and 6.12h: the former erroneously exploits the \( 1/\ell \) bit-flip during its desert phase, while the latter is totally lost with respect to the Operator Selection task.

\(^1\)RMAB does not consider the \( n \) term in the MAB formula (Equation 5.11) as the total number of times each operator was applied since the beginning of the search, but rather as the number of times each operator appears in the current sliding window of the Credit Assignment scheme. Hence, by doing so, the exploration term will always ensure that there is at least one application of each operator every \( W \) trials.
Table 6.15: Results on the 10k-bits OneMax problem: objective is to minimize number of generations to achieve the optimum, selecting between 1-bit, 3-bit, 5-bit and 1/ℓ bit-flip mutation operators within a (1+50)-EA. Baseline performances: Optimal (5134 ± 291), Best Static (1-bit 80% + 5-bit 20% : 6206 ± 326), Naive (7955 ± 634). For the sake of comparison, the performance of the preliminary version of AUC (AUCv1) is also presented.
Figure 6.12: Behavior of AP, DMAB, SLMAB, MAB and RMAB, combined with their best Credit Assignment schemes, on the 10k-bits OneMax problem. For the sake of comparison, the behavior of the preliminary version of AUC (AUCv1) is also plotted.
6.4.3 The Long $k$-Path Problem

Proposed by [Horn et al., 1994], Long Paths are unimodal problems designed to challenge local search algorithms. The optimum can be found by following a path in the fitness landscape, the length of which increases exponentially with respect to the bit-string length $\ell$. Accordingly, solving the Long Path using the 1-bit mutation requires a computational time that increases exponentially with $\ell$; efficient optimization relies on taking shortcuts on this path.

A generalization of Long Path problems was proposed by [Rudolph, 1997], referred to as Long $k$-Path, where $k$ is the minimal number of bits to be simultaneously flipped in order to take a shortcut on the path. Formally, the Long $k$-Path can be described as follows [Garnier and Kallel, 2000]:

- The path starts at point $0, \ldots, 0$, with fitness $\ell$; the fitness of any point not on the path is the number of its 0 bits;
- Any point on the path has exactly 2 neighbors with Hamming distance 1 on the path; consequently, two consecutive points on the path have a fitness difference of 1;
- Mutating $i < k$ bits of a point on the path leads to a point which is either off the path (hence with a very low fitness), or on the path but only $i$ positions away from the parent point;
- A shortcut is found by mutating the correct $k$ bits (or more), thus with probability at most $p^k(1-p)^{\ell-k}$.
- The length of the path is calculated as $(k+1)2^{(\ell-1)/k} - k + 1$;

Long $k$-Path problems are defined by recurrence on $\ell$. Starting from the problem $P(k, \ell)$, the path associated to problem $P(k, \ell + k)$ is built as the sequence of $(x_i, 0_k)$, where $x_i$ belongs to $P(k, \ell)$ and $0_k$ is the $k$-length vector made of 0s. This initial sequence is then linked by a “bridge” to the sequence $(x_{L-i}, 1_k)$, where $x_{L-i}$ ranges in inverse order in $P(k, \ell)$ and $1_k$ is the $k$-length vector made of 1s. The bridge is the sequence of $(x_L, y_z)$ where $x_L$ is the last point of path $P(k, \ell)$ and $y_z$ is the $k$-length vector made of $z$ 0s followed by $k - z$ 1s. To exemplify, the construction of the path $P(3, 10)$ is done as follows. Starting from $P(3, 1) = \{0, 1\}$, the sequence $P(3, 4)$ is created:

\[
\begin{array}{c}
0000_1; 0001_2; 0011_3; 0111_4; 1111_5; 1110_6 \\
S_0 \hspace{1cm} Bridge \hspace{1cm} S_1
\end{array}
\]

from which the path $P(3, 7)$ is constructed (Table 6.16a); and finally we arrive to $P(3, 10)$ (Table 6.16b). Going on three more steps with these recursive construction, we arrive to $P(3, 19)$, represented in Figure 6.16c as the number of ones (unitation) versus fitness (red part of the path represents $S_0$, blue is the Bridge, and green is the final sequence $S_1$).

It turns out that the path length decreases as $k$ increases (the original Long Path corresponds to $k = 2$). Nevertheless, the probability of finding a shortcut decreases exponentially with $k$, and the fastest strategy for $k > \sqrt{\ell}$ is to simply follow the path.
6.4 On Boolean Benchmark Problems

Table 6.16: Examples of Long 3-Paths of different length.

<table>
<thead>
<tr>
<th>00000001</th>
<th>00111107</th>
<th>11111109</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000012</td>
<td>01111108</td>
<td>11111110</td>
</tr>
<tr>
<td>00000113</td>
<td></td>
<td>11101111</td>
</tr>
<tr>
<td>00011114</td>
<td></td>
<td>11100112</td>
</tr>
<tr>
<td>00011115</td>
<td></td>
<td>11100013</td>
</tr>
<tr>
<td>00011116</td>
<td></td>
<td>11100001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>00000000000</th>
<th>00111100000</th>
<th>11111100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000000011</td>
<td>01111100000</td>
<td>11111100000</td>
</tr>
<tr>
<td>0000000111</td>
<td>1111100111</td>
<td>1111110111</td>
</tr>
<tr>
<td>0000011111</td>
<td>1111111111</td>
<td>1111111111</td>
</tr>
<tr>
<td>0000001110</td>
<td>1111111010</td>
<td>1111111010</td>
</tr>
<tr>
<td>0000111110</td>
<td>1110011101</td>
<td>1110011101</td>
</tr>
<tr>
<td>0001111110</td>
<td>1110001110</td>
<td>1110001110</td>
</tr>
<tr>
<td>0011111110</td>
<td>1110000111</td>
<td>1110000111</td>
</tr>
<tr>
<td>0011111111</td>
<td>1110000011</td>
<td>1110000011</td>
</tr>
<tr>
<td>0011110000</td>
<td>1110000001</td>
<td>1110000001</td>
</tr>
<tr>
<td>001111000014</td>
<td></td>
<td>1110000000</td>
</tr>
</tbody>
</table>

(a) Path P(3, 7)

<table>
<thead>
<tr>
<th>00000000</th>
<th>00111100000</th>
<th>11111100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0</td>
<td>Bridge</td>
<td>1</td>
</tr>
<tr>
<td>S1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(b) Path P(3, 10)

<table>
<thead>
<tr>
<th>0000000000000</th>
<th>0011110000000</th>
<th>1111110000000</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0</td>
<td>Bridge</td>
<td>1</td>
</tr>
<tr>
<td>S1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(c) Path P(3, 19)

Otherwise \((k \leq \sqrt{l})\), optimization should provably strive to find the shortcuts; in such cases, \emph{exceptional properties of operators are more relevant to EAs behavior than their average properties} [Garnier and Kallel, 2000].

Along the same lines than the ART instances analyzed in Section 6.3.3, thus, Long \(k\)-Path problems can be seen as yet another case in which one operator constantly gives a small reward (when the parent individual belongs to the path, the 1-bit mutation improves the fitness by 1 with probability \(1/\ell\)), while all other mutation operators will fail to improve the fitness in most cases, but possibly achieving very high outlier fitness improvements (shortcuts in the path) with a very small probability. This possibility of having outlier fitness improvements was the main motivation for the use of such a scenario in the assessment of AOS schemes by the time we proposed the Extreme Credit Assignment [Fialho et al., 2009a; Fialho et al., 2009b] (see Section 5.2.2), which indeed showed to perform better than the Average one for most of the Operator Selection techniques considered. These results will be presented in the following, together with the most recently proposed AOS schemes.

The reported experiments consider \(k = 3\) with \(\ell = 49\). Other problem sizes were also tried, with \(\ell \in \{19, 31, 43, 55, 61\}\); they are omitted here, mainly because the conclusions...
attained on all of them were roughly the same (except for $\ell = 61$, in which none of the methods was found to be effective, due to the very low probability of finding shortcuts). A $(1 + 50)$-EA is used here again, with the AOS schemes selecting between some mutation operators. The same operator set used in the OneMax problem (1-bit, 3-bit, 5-bit and $1/\ell$ bit-flip) is considered here, with one additional mutation operator, the $k/\ell$ bit-flip (flipping each bit with probability $k/\ell = 3/49$ in this case), which is the best operator in this scenario according to theoretical studies [Garnier and Kallel, 2000].

In the same way as for the OneMax problem, the benefit of using this benchmark setting is that it enables the identification of the optimal operator at each point of the path, by means of intensive Monte-Carlo simulations, in order to further compare the AOS approaches with the resulting optimal Oracle strategy. Figure 6.13 shows the average fitness improvement achieved by each of the considered operators, starting from each fitness point on the $P(3, 49)$ path, calculated as for the OneMax problem (best gain out of 50 trials for each operator, averaged over 100 runs).

![Figure 6.13: Average fitness gain of mutation operators with respect to the fitness of the parent, within a $(1 + 50)$-EA applied to the Long 3-Path problem with $\ell = 49$, averaged over 100 trials. The optimum fitness value in this case is $F = 262142$.](image)

From this Figure, it can be seen that the 3-bit (or $k$-bit) operator is the one that receives the highest gains during almost all the path, as it deterministically flips $k$ bits every time it is applied, thus having higher chances of taking a shortcut. For the same reason, the $k/n$ bit-flip comes next, flipping $k$ bits in average. The $k$-bit operator, however, is able to achieve the optimum just in case it succeeds in taking a shortcut, while the $k/n$ bit-flip can manage to succeed in either cases: this empirically confirms the theoretical findings presented in [Garnier and Kallel, 2000]. It is also important to note that there are two factors controlling the variance of the gains brought by taking shortcuts: the distance to the optimum, and the distance to some transition points found in the path. The Oracle
6.4 On Boolean Benchmark Problems

Operator Selection strategy was implemented following the results presented in this Figure: the 3-bit is mostly applied, and the 1-bit is used in a few short transition phases and in the final fine-tuning phase, while the other operators are also very occasionally applied.

Empirical Results

By construction, some runs on the Long $k$-Path problem can be “lucky” and discover shortcuts in the path, thus yielding large standard deviations in the performance, as shown in the detailed results presented in Table 6.17a. This is true even for the Oracle strategy, which achieves the optimum in 2821 generations in average, but with a standard deviation of 2496. In this case, the ECDFs, shown in Table 6.17b, are much more informative for the analysis of the empirical comparison. As can be seen, for all the considered techniques, there are runs reaching the optimum in the very early steps (close to 0), while many others are not able to achieve the optimum before the average performance of the Naive uniform strategy, which bounds the plot at 5815 generations. With such a big variance, the behavior plots become meaningless: the instant selection rates are averaged over 50 runs, but the good runs attain the optimum very early by taking shortcuts; thus, a behavior plot would be averaging only the longer (hence bad) runs, what does not correspond to the mean behavior of the method.

Interestingly, the off-line tuned approach using Static probabilities, which applies the 1-bit at 20% of the trials and the 3-bit at a rate of 80%, is able to outperform the Oracle strategy. The Oracle explores mostly the same operators, but in a fixed manner: the 3-bit is used for some fitness ranges due to its high probability of finding a shortcut, while the 1-bit is used only in transition phases where no shortcut is possible. In practice, however, it seems that using the 1-bit at a fixed small rate is more beneficial: although providing very small improvements (1 by 1 in fact), its probability to improve the fitness is high (1/ℓ) when compared to the probability of taking outlier shortcuts in the path.

The winner AOS combination in this case is the MAB, which, in the ECDF plot (Table 6.17b) is the only method able to follow the performance of both Oracle and Static baseline methods. As previously discussed, the MAB (as well as the other bandit-based approaches) does some averaging on the update of the empirical quality estimates (Equation 5.12) for each operator. Thus, even when using the AbsIns Credit Assignment, it takes into account some history of the operator performance, consequently not forgetting it very quickly in such a noisy environment. Additionally, it is known that MAB is the slowest Operator Selection technique with respect to adaptation between the bandit-based methods considered here (as verified in Section 6.3); this seems to be a beneficial characteristic in this case: the 3-bit should continue to be exploited as much as possible even if some other operator appears to be very good from time to time, because it has a much higher probability of taking shortcuts in a Long $k$-Path with $k = 3$, as previously pointed out. The RMAB with FAUC (fitness values, comparison-based) follows the same trend in around 40% of the runs, but its global picture is rather similar to both AbsExt-DMAB and AbsIns-SLMAB. Lastly, the AP is not able to cope well with Long $k$-Paths: its best results are obtained with $p_{\text{min}} = 0.2$, what is equivalent to the Naive uniform selection of operators in this case, as 5 operators are considered.
### Table 6.17: Results on the Long 3-Path ($\ell = 49$) problem: objective is to minimize number of generations to achieve the optimum, selecting between 1-bit, 3-bit, 5-bit, 1/$\ell$-bit-flip and 3/$\ell$-bit-flip mutation operators within a (1+50)-EA. Baseline performances: Oracle (2821 ± 2496), Naive (5815 ± 3884), Best Static (1-bit 20% + 3-bit 80% : 2354 ± 1400)
6.4 On Boolean Benchmark Problems

6.4.4 The Royal Road Problem

The Royal Road (RR) is an optimization problem that was intentionally created to be easy for GAs [Mitchell et al., 1992] (with the crossover operators exploring the “building blocks” of the function), while being difficult for hill-climbing algorithms. Due to unexpected difficulties (the so-called hitch-hiking phenomenon), a revised version was later proposed in [Holland, 1993] and analyzed in [Jones, 1994]. The revised version is the one considered in this work.

The solutions are represented as bit-strings. Each bit-string is composed of $2^k$ regions, referred to as lower-level (or level 0) schemata. Higher level schemata are formed by combining pairs of lower-level ones, as shown in Table 6.18.

Table 6.18: Example of constructions of higher order schemata from lower order ones on the Royal Road problem

| Level 0: | \{B_0\}, \{B_1\}, \{B_2\}, \{B_3\}, \{B_4\}, \{B_5\}, \ldots, \{B_{12}\}, \{B_{13}\}, \{B_{14}\}, \{B_{15}\} |
| Level 1: | \{B_0, B_1\}, \{B_2, B_3\}, \{B_4, B_5\}, \ldots, \{B_{10}, B_{11}\}, \{B_{12}, B_{13}\}, \{B_{14}, B_{15}\} |
| Level 2: | \{B_0, B_1, B_2, B_3\}, \{B_4, B_5, B_6, B_7\}, \ldots, \{B_{12}, B_{13}, B_{14}, B_{15}\} |
| Level 3: | \{B_0, B_1, \ldots, B_7\}, \{B_8, B_9, \ldots, B_{15}\} |
| Level 4: | \{B_0, B_1, \ldots, B_{15}\} |

Formally, a higher level $L$ has $2^{k-L}$ schemata composed by $2^L$ first-level ones (the building-blocks, supposedly defining a crossover-friendly landscape). Each first-level schema is further divided into a block and a gap string, of respective lengths $b$ and $g$. A bit-string is thus represented by $2^k \cdot (b + g)$ bits.

For the calculation of the fitness of a candidate solution (bit-string), each first-level schema is independently evaluated, with the fitness resulting in the sum of the evaluations of all the schemata. Only the block region of each low level schema is considered, the gap region is completely ignored. The fitness is measured by the PART function or by the BONUS function, as follows. The PART function computes the number $z$ of correct bits in the $b$-length block, resulting in a function value of $(z \cdot v)$ if $(z < m)$ and $((b - z) \cdot v)$ for $(m < z < b)$, where $m$ is a threshold that tunes the level of local deception in the function (see Figure 6.14). The completed blocks in the bit-string, i.e., the ones that have $z = b$, are evaluated by the BONUS function instead, which accounts a score of $u^*$ for the first block to be completed, and $u$ for the additional ones.

The Royal Road was found to be another interesting scenario to empirically analyze the AOS combinations within a real evolutionary algorithm as, by considering common crossover operators, the following can be stated (intuitively, and confirmed in [Quick et al., 1996]): the uniform crossover is the best operator during the initial evolution stages (exploration), 1-point crossover is the best in the final stages (exploitation), while the 4-point crossover is the best in-between. This operator set was used for the experiments that will be presented in the following, with the addition of two other operators, the 2-point crossover, and a disruptive bit-flip mutation operator that flips 8 bits on average (and hence possibly one block). After every crossover application, a mutation operator was also systematically applied, flipping each bit with a probability of 1%. These
operators were applied within a (100,100)-GA with weak elitism, i.e., at every generation, the entire population of 100 individuals is completely replaced by the newly generated 100 offspring, with the possible exception of the best parent, which is maintained (and the worst offspring is removed) if better than the best offspring. The parental selection mechanism used was the tournament one, with size 2.

The problem function was defined using the default parameter values proposed in [Holland, 1993]: $k = 4$, $b = 8$, $g = 7$, $m = 4$, $v = 0.02$, $u^* = 1.0$ and $u = 0.3$. The parameter $m = 4$ defines a medium level of deception; the fully deceptive case ($m = 1$) and the not deceptive one ($m = 7$) were also investigated, but the former was found too difficult to be solved within the given budget of 25,000 generations, while the latter was too easy, thus not enabling any distinction to be made between the AOS schemes. Figure 6.14 illustrates a comparison of these 3 different levels of deceptiveness for the first 30 bits of this problem setting, on a unitation (number of 1s) versus fitness plot. With $2^k$ regions involving $(b + g)$ bits, the total dimension of the considered search space accounts to 240 bits.

![Figure 6.14: Different levels of deceptiveness on the Royal Road problem, varying $m$ and using the default values for the other parameters.](image)

**Empirical Results**

The behavior of each operator is very difficult to be guessed on the Royal Road problem. Despite the 1/30 mutation operator, the other 4 crossover operators (specially the $x$-points ones) tend to present a similar behavior, all exploring the building blocks of the intentionally designed search space. Besides, there is no “fine-tuning operator” between them: even the 1-point crossover substantially modifies the solution to which it is applied.
6.4 On Boolean Benchmark Problems

to, making it easier to miss the target, thus explaining the high variance of the detailed performance results shown in Table 6.19a. As can be seen, almost all AOS combinations are not significantly different with respect to the best. For this reason, as for the Long \( k \)-Path problem, ECDFs (Table 6.19b) are used in order to have a more complete view of the performance distribution for each \textit{Operator Selection} technique with its better \textit{Credit Assignment} scheme.

Despite the big variance, all methods achieve the optimum faster than the average performance of the Naive uniform selection strategy in at least 80\% of the cases. Notably, the best Static strategy found for this problem is the use of a single operator at a rate of 100\%, the 4-point crossover, which achieves the optimum in 6244 generations in average. Several other configurations using different combinations of the 1-point, 2-point and 4-point operators are also able to achieve equivalent performance; for instance, 1-point at 20\%, 2-point at 20\% and 4-point at 60\% achieves the optimum in 6679 ± 4278 generations. Hence, in order to achieve reasonable performance on this experimental setting, an AOS method should “simply” be capable of discarding the 1/30 mutation and the uniform crossover operators; the way the other three operators are used does not matter much. To confirm this assumption, additional experiments were done for the Naive uniform strategy considering only these 3 operators: the optimum is found in 7066 ± 4215 generations, a performance better than (although not significantly different from) those obtained by most of the AOS methods.

The only AOS combination able to closely follow the performance of the Static baseline up to 100\% of the trials is the DMAB with, surprisingly, the Normalized Average (NormAvg) \textit{Credit Assignment}. It is worth noting that here, again, the Normalized outperform the Absolute for the different kinds of \textit{Credit Assignment} in most cases, with the rank-based schemes also presenting reasonable performance. This is also the first case in which the RMAB with the rank-based SR outperforms (but not significantly) its combination with the AUC \textit{Credit Assignment}, with both versions based on fitness improvements and on fitness values (FSR) achieving similar performance.

Concerning specifically the \textit{Operator Selection} techniques, the DMAB and RMAB, with their corresponding best \textit{Credit Assignment} schemes, greatly outperform in terms of average performance (but not significantly due to the mentioned high variance) all the combinations involving MAB, SLMAB and AP. It is important to note that for most of the winner configurations of the bandit-based approaches, a very small value is used for the scaling factor \( C \); accordingly, very small values are used for the adaptation and learning rates of AP. This indicates that these techniques are rarely adapting to exploit other operators, mostly exploiting the first operator that they found to be the best. As previously discussed, if this operator is one of the \( x \)-point crossover operators, this choice will not greatly affect their performances.
Table 6.19: Results on the Royal Road \((m = 4)\) problem: objective is to minimize number of generations to achieve the optimum, selecting between 1-point, 2-point, 4-point and uniform crossover operators, and 1/30 bit-flip mutation operator, within a (100,100)-EA with weak elitism. Baseline performances: Oracle not available, Naive \((15940 \pm 6928)\), Best Static \((4\text{-point }100\%: 6244 \pm 3037)\).
6.4.5 Discussion

The benchmark optimization problems considered in this Section enabled a preliminary analysis of the AOS methods in practice, selecting between real evolutionary operators, based on the feedback given by real (although still artificially created) fitness landscapes and by the trajectory taken by the EA in the search space. The OneMax is a very simple problem, but its use in this experimental setting was of great value, as it provided a very detailed and complete behavioral analysis of each AOS method. The empirical analyses on the Long $k$-Path and Royal Road problems did not provide the same level of information, but they challenged the AOS methods in different situations that might happen in real cases. In the Long $k$-Path case, only one operator should be mostly exploited at any time, in order to possibly take very rewarding shortcuts in the path; while in the Royal Road two out of five operators should be discarded, the other three being equally beneficial.

For each of these problems, the AOS methods were compared to three non-adaptive baseline approaches, namely, the Naive, the Oracle, and the Static strategies, defining three different levels of available knowledge. The Naive strategy, as its name says, represents the situation in which nothing is known about the performance of the operators on the problem at hand; the Oracle represents the complete detailed information about their performances with respect to each fitness value. While the former strategy is a rather straightforward choice when there is no time for a deeper analysis, the latter strategy can be precisely assessed only in simple problems such as the OneMax one; hence, it is not a valid choice in the real world. It is also worth noting that the Oracle strategy is defined based on statistics over several runs of the underlying EA with its corresponding operators, i.e., based on results of stochastic nature; this explains why sometimes the adaptive methods were able to outperform it. In the middle of these two approaches, there is the Static strategy, which is the approach most commonly used since the very early days of applied research in the area; it requires a reasonable level of knowledge that can be gathered whenever a few runs are affordable, in order to find the best off-line tuned static approach.

Although requiring a preliminary off-line tuning of their hyper-parameters, compared to these baseline approaches, the best AOS methods were able to achieve performance equivalent to the Oracle behavior on both OneMax and Long $k$-Path problems. They also showed to be able to significantly outperform the Naive approach on all the problems, although the high variance of the results found for the Long $k$-Path and the Royal Road problems. And concerning the Static strategy, it was also significantly outperformed on the OneMax scenario; but in the Long $k$-Path, the Static strategy surprisingly outperformed the Oracle strategy, significantly outperforming all the AOS methods by using only two out of the four available operators; while for the Royal Road, the best Static strategy found uses only one out of five operators, with some AOS methods being able to show similar performance.

In what concerns the Operator Selection techniques, the DMAB and RMAB showed to efficiently and consistently improve over the standard MAB approach, except for the outlier Long $k$-Path scenario, in which the standard MAB is surprisingly the winner: its slower adaptation seems to be beneficial in such a noisy scenario. The SLMAB, however,
was not able to outperform the standard MAB in any of the cases, while in the ART scenarios it was always between the overall winners. A tentative of explanation for this deception is that its update mechanism is designed to adapt quickly to very abrupt changes in the operator qualities, what is not the case in these experimental settings. The last Operator Selection technique, the baseline AP, is top-ranked only on the OneMax scenario; on the Long k-Path its best performance is equivalent to the Naive approach, and on the Royal Road it is also outperformed by all bandit-based approaches.

Finally, in these problems, differently from the ART scenarios, the benefits brought by the use of more robust Credit Assignment schemes could be better highlighted. On the OneMax and Royal Road problems, the Normalized versions of the Credit Assignment schemes based on the raw values of fitness improvements outperformed the Absolute versions in most cases. On the Long k-Path, the situation is the opposite, as in this scenario the magnitude of the outlier improvements achieved are very important. In either cases, most of the several available options for the rank-based Credit Assignment schemes were top-ranked, performing not significantly different from the winner configurations. Still, as for the ART problems, such analysis compared the performance of each AOS combination with its best hyper-parameter setting found by a preliminary off-line tuning procedure for each problem. Thus, this gain in performance provided by the rank-based and normalized schemes were attained only by the minimization of one of the issues that motivated their proposal, that of providing rewards at the same value range during all the search process, as discussed in Section 5.2.3. The second and consequent benefit, that of providing a robust behavior with respect to the hyper-parameters of the AOS method, will be separately analyzed in Section 6.7.

6.5 Collaboration On Satisfiability Problems

The preliminary complete AOS combination proposed in our work involves the Absolute Extreme (AbsExt) Credit Assignment scheme (Section 5.2.2) with the Dynamic Multi-Armed Bandit (DMAB) Operator Selection technique (Section 5.3.2), which will be simply referred to as Ex-DMAB in this Section, for the sake of brevity. While working on its further assessment on different benchmark scenarios, we established a collaboration with Université d’Angers, France, published in [Maturana et al., 2009a; Maturana et al., 2010a], whose results will be surveyed in this Section.

The combination of Ex-DMAB with the Compass Credit Assignment (Section 4.3.4) will be described in Section 6.5.1. This AOS combination was assessed in the light of Boolean Satisfiability (SAT) problems, which will be presented in Section 6.5.2. Sections 6.5.3 and 6.5.4 will describe, respectively, the specific experimental settings and the off-line tuning procedure used in this work. Finally, the empirical results will be presented in Section 6.5.5, with a concluding discussion in Section 6.5.6.

6.5.1 Compass + Ex-DMAB = ExCoDyMAB

The Compass-based AOS technique, proposed in [Maturana and Saubion, 2008a], is in fact the combination of an engineered Credit Assignment mechanism referred to as Compass,
which measures the effect of operators application taking into account fitness, diversity and CPU time, with a rather simple Operator Selection mechanism, the Probability Matching (PM). At the same time, the Ex-DMAB AOS technique is the combination of a simple Credit Assignment scheme, the Extreme value out of the recent fitness improvements achieved by the operator, with an efficient Operator Selection mechanism, the DMAB.

From this brief review, it becomes clear that both AOS approaches have complementary strengths and weaknesses: Compass might enable DMAB to be efficiently applied to multi-modal problems, while Ex-DMAB might provide to Compass a more efficient Operator Selection mechanism, while also improving it by the use of the Extreme paradigm. However, even though merging both modules can be done in a straightforward manner, some important issues need to be further explored:

- Compass uses sliding windows in the “impact evaluation” stage (see Figure 4.1), outputting a unique value; while Ex-DMAB keeps a sliding window in the Credit Assignment stage, from which it extracts the maximum or Extreme values. Should we keep both windows, or would it degrade or disappear with the interesting characteristics provided by Compass? And if only one of these windows is kept, which one should it be? From here on, these two windows will be respectively referred to as W1 and W2.

- Another issue concerning the sliding windows is that of what should be their output. Originally, the output of Compass W1 is the Average over the impacts measured after the most recent applications [Maturana and Saubion, 2008a]; a simpler approach would be the Instantaneous value, i.e., no window at all. Ex-DMAB uses the Extreme Credit Assignment, which was successfully validated in the scope of different artificial (Section 6.3) and unimodal (Section 6.4) benchmark problems. But would these results also hold in such a completely different setting?

- The last issue concerns the other hyper-parameters. Besides the size and type of W1 and W2, we need to tune the values of the angle Θ in Compass, and the scaling factor C and change detection threshold γ in DMAB. Since the idea is not to simply replace some parameters (the operator application probabilities) by other ones, even if at a higher level of abstraction, we need to better understand their effects. One way to do so is to experimentally study their influence on the performance of the AOS in situation, and to propose some robust default values.

The resulting combination of Ex-DMAB and Compass is referred to as Extreme Compass - DMAB (ExCoDyMAB). An empirical analysis of the discussed issues will be presented in the following.

### 6.5.2 Boolean Satisfiability Problems

The ExCoDyMAB AOS method has been assessed within an EA applied to the well-known combinatorial Boolean Satisfiability (SAT) problem [Cook, 1971], which consists in assigning values to binary variables in order to satisfy a Boolean formula.
Chapter 6. Experimental Results

Formally, an instance of the SAT problem is defined by a set of Boolean variables \( X = \{x_1, \ldots, x_n\} \) and a Boolean formula \( F : \{0, 1\}^n \rightarrow \{0, 1\} \). The formula is said to be satisfiable if there exists an assignment \( v : X \rightarrow \{0, 1\}^n \) satisfying \( F \), unsatisfiable otherwise. Instances are classically formulated in conjunctive normal form (conjunctions of clauses) and one thus has to satisfy all these clauses. Given that SAT was the first problem to be proved NP-complete, and also due to its very general boolean (bit-string) representation, many different problems from both real-world and theoretical background have been expressed as SAT instances. So, by tackling such problem, we can deal with a very diverse set of fitness landscapes with different characteristics.

Table 6.20 shows the instances used here, extracted from the SATLIB [Hoos and Stützle, 2000] and from the SAT-Race 2006 [Sinz et al., 2006]. It also points out whether the instances are satisfiable or not, their family, and the number of variables and clauses they involve.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Sat?</th>
<th># Vars.</th>
<th># Clauses</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>4blocks</td>
<td>Yes</td>
<td>758</td>
<td>47820</td>
<td>Blocks World Problem</td>
</tr>
<tr>
<td>aim</td>
<td>Yes</td>
<td>200</td>
<td>320</td>
<td>Random-3-SAT</td>
</tr>
<tr>
<td>f1000</td>
<td>Yes</td>
<td>1000</td>
<td>4250</td>
<td>Random-3-SAT</td>
</tr>
<tr>
<td>CBS</td>
<td>Yes</td>
<td>100</td>
<td>449</td>
<td>Controlled Backbone</td>
</tr>
<tr>
<td>flat200</td>
<td>Yes</td>
<td>600</td>
<td>2237</td>
<td>Flat Graph Coloring</td>
</tr>
<tr>
<td>logistics</td>
<td>Yes</td>
<td>828</td>
<td>6718</td>
<td>Logistics Planning</td>
</tr>
<tr>
<td>medium</td>
<td>Yes</td>
<td>116</td>
<td>953</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>par16</td>
<td>Yes</td>
<td>1015</td>
<td>3310</td>
<td>Parity Learning Problem</td>
</tr>
<tr>
<td>sw100-p0</td>
<td>Yes</td>
<td>500</td>
<td>3100</td>
<td>Morphed Graph Coloring</td>
</tr>
<tr>
<td>sw100-p1</td>
<td>Yes</td>
<td>500</td>
<td>3100</td>
<td>Morphed Graph Coloring</td>
</tr>
<tr>
<td>uf250</td>
<td>Yes</td>
<td>250</td>
<td>1065</td>
<td>Phase Transition Region</td>
</tr>
<tr>
<td>uuf250</td>
<td>No</td>
<td>250</td>
<td>1065</td>
<td>Phase Transition Region</td>
</tr>
<tr>
<td>Color*</td>
<td>No</td>
<td>1444</td>
<td>119491</td>
<td>Chessboard Coloring</td>
</tr>
<tr>
<td>G125*</td>
<td>Yes</td>
<td>2125</td>
<td>66272</td>
<td>Graph Coloring</td>
</tr>
<tr>
<td>Goldb-heqc*</td>
<td>No</td>
<td>5980</td>
<td>35229</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>Grieu-vmpc</td>
<td>Yes</td>
<td>729</td>
<td>96849</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>Hoons-vbmc*</td>
<td>No</td>
<td>8503</td>
<td>25116</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>Schup</td>
<td>No</td>
<td>14809</td>
<td>48483</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>Simon*</td>
<td>No</td>
<td>2424</td>
<td>14812</td>
<td>Randomly Generated</td>
</tr>
<tr>
<td>Manol-pipe</td>
<td>Yes</td>
<td>14052</td>
<td>41596</td>
<td>Pipelined Machine Verification</td>
</tr>
<tr>
<td>Velev-eng*</td>
<td>No</td>
<td>6944</td>
<td>66654</td>
<td>Pipelined Machine Verification</td>
</tr>
<tr>
<td>Velev-sss*</td>
<td>No</td>
<td>1453</td>
<td>12531</td>
<td>Pipelined Machine Verification</td>
</tr>
</tbody>
</table>

Table 6.20: SAT instances used in the empirical assessment of ExCoDyMAB

### 6.5.3 Experimental Settings

The ExCoDyMAB is applied to an EA that uses a standard binary representation (one bit per boolean variable) to represent each solution. As in [Maturana and Saubion, 2008a],
the purpose here is not to use state-of-the-art SAT operators, but rather to manage a set of completely unknown operators, as a naive user would do when facing a new problem. Desirably, the AOS mechanism should then be able to autonomously discriminate good from bad operators at any given time of the search, further exploiting the best operator. The very heterogeneous operator set is constituted by the following operators:

- **1-point Crossover** randomly chooses two individuals and a random position, and exchanges their first and second parts.

- **Contagion** randomly chooses two individuals and sets the variables in all false clauses of the worst individual to the values they have in the best one.

- **Hill Climbing** checks all neighbors at Hamming distance 1 and moves to the best one, repeating the process as long as it improves the fitness. It is important to note that this is a local search operator, which has been included here for the sake of diversity of variation operators.

- **Tunneling** swaps variables without decreasing the number of true clauses, according to a tabu list of length equal to \( \frac{1}{4} \) of the number of variables (it can be seen, again, as a local search operator).

- **Bad Swap** swaps all variables that appear in false clauses, whatever their values are.

- **Wave** swaps the values of the variable that appears in the highest number of false clauses and in the minimum number of clauses only supported by it; the process is repeated at most \( \frac{1}{2} \) times the number of variables, while improvements can be found.

The parental selection mechanism is the steady-state, defined in Section 2.3.3: after the generation of each offspring, the worst individual in the population is immediately replaced (except when the 1-point Crossover operator is applied, in which case the best out of the two generated offspring replaces the worst parent). The population size (3) and maximum number of generations (5000 – the only stopping criterion) were arbitrarily fixed.

### 6.5.4 Architecture definition and tuning of hyper-parameters

In order to efficiently integrate Compass and Ex-DMAB, some open issues were discussed in Section 6.5.1. The definition of these components, as well as the off-line tuning of the other hyper-parameters, will be now analyzed in turn.

The first decision concerns whether to include or not the sliding windows \( W_1 \) and/or \( W_2 \), and which should be their outputs. The following possible output policies were tried for each window (note that the Normalized versions were not considered in this work):

- Instantaneous (I) value, i.e., no sliding window.

- Average (A) value from stored measures;
Chapter 6. Experimental Results

- Extreme (E) value (maximum) from stored values (except for the execution time, kept in \( W_1 \), as the extreme of this measure would not make sense);

Besides, the following hyper-parameters also need to be analyzed and tuned:

- The size of Compass window of impact measures \( W_1 \), and the size of the sliding window of the outputs of the Credit Assignment scheme \( W_2 \).
- The Compass angle \( \Theta \), that defines the tradeoff between the exploration and exploitation at the Credit Assignment level.
- The DMAB scaling \( C \) parameter, that defines the tradeoff between exploration and exploitation at the Operator Selection level.
- The DMAB \( \gamma \) parameter, the threshold of the change detection test that triggers the restarts.

The range of values tried for the different hyper-parameters are defined as follows: \( C \in \{5, 7, 10\} \); \( \gamma \in \{1, 3, 5\} \); and the windows type(size) combinations \( \in \{A(10), A(50), E(10), E(50), I(1)\} \) for both \( W_1 \) and \( W_2 \). Thus, the initial number of possible configurations is 225.

The angle \( \Theta \) for Compass was set to \( \pi/4 \), as preliminary experiments have shown that a different value causes a positive-feedback phenomenon\(^2\), that moves the EA to an extreme behavior. For instance, Figure 6.15 shows the curve of the best fitness found, with respect to the number of time steps elapsed, when using different values of \( \Theta \) for the original Compass AOS combination applied to the par16-1 instance, averaged over 50 runs. Note that all values below 0.25\( \pi \) tend to produce a similar erratic behavior, while values above 0.25 have a poor improvement rate.

The same F-Race off-line tuning procedure was used for the tuning of these hyper-parameters. The stopping criteria for the Racing was set to 80 runs over all the instances, with eliminations taking place after each run, starting from the 11\(^{th} \). All 22 SAT instances listed in Table 6.20 have been considered for the final empirical comparison, but only 7 of them were taken into account for this off-line tuning phase. This sub-set, marked with an asterisk in Table 6.20, was chosen among the hardest instances with short enough running times, reducing the experimental cost for the platform definition. Tuning the hyper-parameters on a small set of instances and testing them further on “unseen” instances witnesses the generality of the tuned parameters.

6.5.5 Empirical Results

At the end of the Racing, 4 configurations were still active in the process, which are presented in Table 6.21. These results clearly indicate that the most important sliding window is \( W_1 \), \( i.e., \) the Compass window for the impact measures, and it should be used in its Extreme configuration with a size of 10 \( (i.e., \) taking as Compass inputs the maximum

\(^2\)In systems theory, positive feedback is a process in which a system responds to a perturbation in the same sense of the perturbation, thus distancing the system from its original state.
6.5 Collaboration On Satisfiability Problems

Figure 6.15: Curve of the best fitness found in relation to the number of time steps elapsed, for different values of $\Theta$ on Compass applied to the par16-1 instance, averaged over 50 runs.

<table>
<thead>
<tr>
<th>Name</th>
<th>$W_1$ type, size</th>
<th>$W_2$ type, size</th>
<th>$C$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Extreme, 10</td>
<td>Instantaneous</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>Extreme, 10</td>
<td>Average, 10</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>Extreme, 10</td>
<td>Average, 50</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>Extreme, 10</td>
<td>Extreme, 10</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 6.21: Racing survivors after ExCoDyMAB hyper-parameters tuning

of the last 10 assessed impact measures), not matter which kind/size of $W_2$ is used. This fact emphasizes the need to identify rare-but-good improvements, greatly supporting the idea raised by the proposal of the Extreme Credit Assignment, described in Section 5.2.2. Besides, the size of 10 for $W_1$ could be interpreted by the following reasoning. With the Extreme policy, a larger $W_1$ would produce a long perdurability of the extreme values, even when the behavior of the operator has already changed. In the other hand, a shorter value, up to $W_1 = 1$ (i.e., the same as choosing the Instantaneous policy) would quickly forget these “rare-but-good” cases. One could suppose that an optimal size for $W_1$ depends on the fitness landscape and the operators used - further research is needed to better understand the setting of this hyper-parameter.

To check the generality of those parameters, 50 runs were performed on the 22 SAT instances with each of the 4 configurations, promoting an empirical comparison between them, and also verifying their performance in relation to the baseline methods: the original combinations of Compass and Ex-DMAB (including a Racing phase for Ex-DMAB similar

155
to that of ExCoDyMAB), and the Naive uniform selection of operators. The results of this comparison are summarized in Table 6.22. Each cell value represents the number of problems in which one architecture is significantly better than the other (using a Student T-test with 95% confidence). For example, in the lower left corner, “18-2” means that D outperformed Compass on 18 instances, while the opposite happened only 2 times. Finally, the rightmost column shows the number of times that an architecture wins, minus the times that it loses, as a global measure of comparative quality.

<table>
<thead>
<tr>
<th></th>
<th>Compass</th>
<th>Ex-DMAB</th>
<th>Naive</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>∑ dom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compass</td>
<td>–</td>
<td>9-9</td>
<td>22-0</td>
<td>4-18</td>
<td>2-17</td>
<td>2-18</td>
<td>2-18</td>
<td>-39</td>
</tr>
<tr>
<td>Ex-DMAB</td>
<td>9-9</td>
<td>–</td>
<td>22-0</td>
<td>0-18</td>
<td>0-21</td>
<td>0-21</td>
<td>0-21</td>
<td>-59</td>
</tr>
<tr>
<td>Naive</td>
<td>0-22</td>
<td>0-22</td>
<td>–</td>
<td>0-22</td>
<td>0-22</td>
<td>0-22</td>
<td>0-22</td>
<td>-132</td>
</tr>
<tr>
<td>A</td>
<td>18-4</td>
<td>18-0</td>
<td>22-0</td>
<td>–</td>
<td>0-1</td>
<td>0-5</td>
<td>0-2</td>
<td>46</td>
</tr>
<tr>
<td>B</td>
<td>17-2</td>
<td>21-0</td>
<td>22-0</td>
<td>1-0</td>
<td>–</td>
<td>0-2</td>
<td>3-1</td>
<td>59</td>
</tr>
<tr>
<td>C</td>
<td>18-2</td>
<td>21-0</td>
<td>22-0</td>
<td>5-0</td>
<td>2-0</td>
<td>–</td>
<td>4-0</td>
<td>70</td>
</tr>
<tr>
<td>D</td>
<td>18-2</td>
<td>21-0</td>
<td>22-0</td>
<td>2-0</td>
<td>1-3</td>
<td>0-4</td>
<td>–</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 6.22: Comparative results on the 22 SAT instances: each cell indicates the number of times the row-algorithm is better than the column-algorithm according to a Student T-test with 95% confidence.

After this analysis, between all the four survivors of the Racing procedure, the configuration “C” was found to be the best for ExCoDyMAB, and was thus used for further empirical comparison with the baseline techniques, namely, the original Compass-PM and Ex-DMAB AOS combinations, and the Naive uniform choice. The results are presented in Table 6.23. The columns show the mean number of false clauses after 5000 function evaluations, averaged over 50 runs, and the standard deviation between parentheses. The best results for each instance are highlighted in **bold-face**. As can be seen, ExCoDyMAB outperforms the other techniques in the vast majority of the cases. These results will be further discussed in the following.

### 6.5.6 Discussion

The dominance of ExCoDyMAB is overwhelming, and confirms the hypothesis that motivated the combination of both Compass and DMAB approaches. These latter approaches alone, within their respective original combinations, present a performance roughly equivalent between each other on this experimental setting, and clearly inferior to the newly combined one, ExCoDyMAB – though still outperforming in turn the Naive uniform selection policy.

Another interesting point is the seemingly good generalization capacity of ExCoDyMAB with respect to its hyper-parameters: the best configurations found by F-Race on the 7 “training” instances showed to perform also very well when solving the other 15 unseen instances. Moreover, the credits assigned by Compass are normalized by construction, and this might result into a more robust technique with respect to the
configuration of its hyper-parameters. But this deserves further analysis; in the meantime, the main drawback of this combination is still that of needing to count with a preliminary expensive off-line tuning phase in order to achieve reasonable performance, specially in what concerns the DMAB hyper-parameters.

Note that this work was done before the proposal by us of the more robust rank-based AOS approaches. In the same way, the Compass authors have come up with more efficient Credit Assignment schemes that also integrate both impact measures, based on the Pareto Front paradigm [Maturana et al., 2010b]. As a further work, a combination of these newly proposed components will be analyzed on this scenario, hopefully achieving better results while showing to be more robust with respect to its hyper-parameters (or at least cheaper in relation to their off-line tuning).

It is also important to remember, as previously mentioned, that the purpose of this work was not to build an overwhelming SAT solver, but rather to experiment and validate the ExCoDyMAB as an AOS technique with an EA solving a general difficult and highly multi-modal combinatorial problem. The main interesting result is that this set of benchmarks was difficult enough to highlight the benefits of using the proposed combination of

<table>
<thead>
<tr>
<th>Method</th>
<th>ExCoDyMAB</th>
<th>Compass</th>
<th>Ex-DMAB</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>4blocks</td>
<td>2.8 (0.9)</td>
<td>6 (0.9)</td>
<td>6.2 (0.9)</td>
<td>13.4 (0.6)</td>
</tr>
<tr>
<td>aim</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1.2 (0.3)</td>
<td>3.6 (1.8)</td>
</tr>
<tr>
<td>f1000</td>
<td>10.3 (2.3)</td>
<td>30.9 (6.2)</td>
<td>16.4 (2.6)</td>
<td>55.8 (8.6)</td>
</tr>
<tr>
<td>CBS</td>
<td>0.6 (0.6)</td>
<td>0.4 (0.5)</td>
<td>1 (0.9)</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>flat200</td>
<td>7.2 (1.7)</td>
<td>10.6 (2.1)</td>
<td>10.7 (2.2)</td>
<td>37.7 (5.5)</td>
</tr>
<tr>
<td>logistics</td>
<td>6.5 (1.3)</td>
<td>7.6 (0.5)</td>
<td>8.8 (1.5)</td>
<td>17.9 (4.1)</td>
</tr>
<tr>
<td>medium</td>
<td>1.5 (1.5)</td>
<td>0 (0)</td>
<td>1.8 (1.6)</td>
<td>8.8 (3.4)</td>
</tr>
<tr>
<td>par16</td>
<td>15.2 (3.1)</td>
<td>64 (10.2)</td>
<td>24.1 (5.7)</td>
<td>131.1 (14.5)</td>
</tr>
<tr>
<td>sw100-p0</td>
<td>9.2 (1.2)</td>
<td>12.8 (1.4)</td>
<td>12.5 (1.7)</td>
<td>25.9 (3.4)</td>
</tr>
<tr>
<td>sw100-p1</td>
<td>0 (0)</td>
<td>0.5 (0.6)</td>
<td>1.1 (0.8)</td>
<td>11.3 (3.5)</td>
</tr>
<tr>
<td>uf250</td>
<td>0.9 (0.7)</td>
<td>1.8 (0.9)</td>
<td>1.7 (0.8)</td>
<td>9.1 (3.3)</td>
</tr>
<tr>
<td>uuf250</td>
<td>2.5 (1)</td>
<td>4.5 (1.2)</td>
<td>3.1 (1.1)</td>
<td>12.7 (3.2)</td>
</tr>
<tr>
<td>Color</td>
<td>48 (2.5)</td>
<td>61.3 (2.2)</td>
<td>49.3 (3.4)</td>
<td>80.4 (6.6)</td>
</tr>
<tr>
<td>G125</td>
<td>8.8 (1.3)</td>
<td>20.6 (2)</td>
<td>13.5 (1.7)</td>
<td>28.8 (4.6)</td>
</tr>
<tr>
<td>Goldb-heqc</td>
<td>72.9 (8.5)</td>
<td>112.2 (15.2)</td>
<td>133.2 (15.9)</td>
<td>609.7 (96.2)</td>
</tr>
<tr>
<td>Grieu-vmpe</td>
<td>16.7 (1.7)</td>
<td>15.2 (1.7)</td>
<td>19.6 (1.8)</td>
<td>24.1 (3.3)</td>
</tr>
<tr>
<td>Hoons-vbmc</td>
<td>69.7 (14.5)</td>
<td>268.1 (44.6)</td>
<td>248.3 (24.1)</td>
<td>784.5 (91.9)</td>
</tr>
<tr>
<td>Manol-pipe</td>
<td>163 (18.9)</td>
<td>389.6 (37.2)</td>
<td>321 (38.1)</td>
<td>1482.4 (181.5)</td>
</tr>
<tr>
<td>Schup</td>
<td>306.6 (26.9)</td>
<td>807.9 (81.8)</td>
<td>623.7 (48.5)</td>
<td>1639.5 (169.9)</td>
</tr>
<tr>
<td>Simon</td>
<td>29.6 (3.3)</td>
<td>43.5 (2.7)</td>
<td>35.3 (6.3)</td>
<td>72.6 (11.3)</td>
</tr>
<tr>
<td>Velef-eng</td>
<td>18.3 (5.2)</td>
<td>29.5 (7.3)</td>
<td>118 (37.1)</td>
<td>394 (75.8)</td>
</tr>
<tr>
<td>Velef-sss</td>
<td>2 (0.6)</td>
<td>4.6 (1)</td>
<td>5.9 (3.9)</td>
<td>62.7 (25.2)</td>
</tr>
</tbody>
</table>

Table 6.23: Comparative results on the 22 SAT instances: average (std dev.) number of false clauses (over 50 runs)
Compass and Ex-DMAB rather than either separately – or than the naive blind choice. The deliberate choice of several non-specialized operators was also an important point to validate the control ability of ExCoDyMAB when facing variation operators of very different efficiencies.

Finally, although the results presented in Table 6.23 show that a basic EA using rather naive operators can indeed solve some instances, competing for SAT Race implies using highly specialized operators, and possibly problem-dependent knowledge, as done in [Wei et al., 2008]. We are currently working on this in collaboration with University of British Columbia, the AOS schemes choosing between state-of-the-art heuristics for variable selection; but by the time this manuscript is being written, there are no conclusive results yet.

6.6 On Continuous Benchmark Problems

The experimental results surveyed so far in this Chapter considered the AOS schemes coupled with a Genetic Algorithm, and applied to artificial, boolean benchmark, and SAT problems. In order to analyze the applicability of such methods in a totally different context, we will present in this Section an empirical analysis of AOS schemes selecting between some mutation strategies within a different EA, the Differential Evolution (DE) algorithm (described in Section 2.4.5), applied to continuous optimization problems. In this context, AOS is also sometimes referred to as Adaptive Strategy Selection (AdapSS) [Gong et al., 2010; Fialho et al., 2010b].

The experimental framework used in these experiments will be introduced in Section 6.6.1. The specific experimental settings will be presented in Section 6.6.2. Section 6.6.3 will describe the PM-AdapSS-DE, a different AOS method used as baseline for comparison. The empirical results will be depicted and analyzed in Section 6.6.4. Finally, Section 6.6.5 will discuss the findings and point out possible directions for further work. The results that will be analyzed here were partially published in [Fialho et al., 2010b; Fialho and Ros, 2010].

6.6.1 Black-Box Optimization Benchmarking

The Black-Box Optimization Benchmarking (BBOB) is a workshop that was held during the 2009 and 2010 editions of the ACM Genetic and Evolutionary Computation Conference (GECCO) [Hansen et al., 2010b]. Partly organized by some members of the Project-team TAO, INRIA Saclay - Ile-de-France, the main objective of this workshop was to present and discuss empirical comparisons of different optimization algorithms in the continuous domain, using a common experimental framework.

As a result of this initiative, important contributions have been made to the research field of empirical analysis of continuous optimizers. Firstly, the BBOB framework provides: two well-defined and documented sets of benchmark functions, a noiseless and a noisy one; an experimental set-up [Hansen et al., 2010a] for analyzing the algorithms in several dimensions and function classes; and some post-processing scripts to generate graphs and tables to assist the user into the analysis of the performance data. Thus, in case one wants
to empirically assess a given optimization algorithm, this task is greatly facilitated by the use of the BBOB framework: the user only needs to interface his optimization algorithm with the framework, allocate some CPU-time, launch some runs, and finally do the post-processing with the aid of the available scripts. Accordingly, this experimental framework can be (and should be) seen as a standard for the empirical analysis of optimization algorithms: by using it, newly proposed algorithms, or new improvements to existing algorithms, can be easily compared to state-of-the-art methods in a rigorous scientific manner.

The empirical analysis that will be presented in this Section has greatly benefited from the use of this experimental framework. More details will be given in the following.

### 6.6.2 Experimental Settings

The goal of the experiments that will be presented here is to assess the comparative performances of the AOS schemes when coupled with the standard version of the Differential Evolution algorithm [Storn and Price, 1997]. The only difference between the considered methods regards the way in which they control the use of the mutation strategies. As described in Section 2.4.5, the DE algorithm is governed by three parameters: \( NP \), \( F \) and \( CR \), respectively denoting the population size, the mutation scaling factor and the crossover rate. It must be emphasized that our goal is not to compete with state-of-the-art continuous optimizers, but rather to provide another proof-of-concept of the possible benefits brought by the AOS paradigm, in a totally different context in relation to both, the underlying EA (just GAs were considered in the previous empirical analyses) and problem domain (continuous in lieu of boolean/combinatorial). Hence, no specific effort was put on tuning the DE parameters with respect to the problem at hand. Population size \( NP \) is set to \( 10 \cdot d \) (as recommended by [Storn and Price, 2008]), where \( d \) denotes the dimension of the search space; mutation scaling factor \( F \) is set to .5, and crossover rate \( CR \) is set to 1. This latter choice provides to DE the invariance property with respect to rotation, while stressing the impact of the application of the mutation strategies (although being counter-intuitive, \( CR = 1 \) means no crossover at all, only mutation strategies are applied), consequently emphasizing the gain brought by each AOS scheme on their control.

The set of variation operators is composed of four standard mutation strategies, retaining the same set as in [Gong et al., 2010] for the sake of comparative evaluation:

1. “DE/rand/1”: \( v_i = x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) \)
2. “DE/rand/2”: \( v_i = x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) + F \cdot (x_{r_4} - x_{r_5}) \)
3. “DE/rand-to-best/2”: \( v_i = x_i + F \cdot (x_{best} - x_{i}) + F \cdot (x_{r_2} - x_{r_3}) + F \cdot (x_{r_4} - x_{r_5}) \)
4. “DE/current-to-rand/1”: \( v_i = x_i + F \cdot (x_{r_1} - x_i) + F \cdot (x_{r_2} - x_{r_3}) \)

where \( x_i \) is the current (or target) individual, \( x_{best} \) is the current best one, and \( x_{r_1}, x_{r_2}, x_{r_3}, x_{r_4}, x_{r_5} \) are individuals uniformly drawn from the population.

As mentioned before, two benchmark sets of single-objective continuous functions are available in the BBOB framework, a noiseless [Hansen et al., 2009a] and a noisy
Chapter 6. Experimental Results

[Hansen et al., 2009b] one. Only the noiseless test-bed will be considered here. It involves 24 functions divided into 5 classes, according to their most relevant characteristics:

- 5 separable functions;
- 4 functions with low or moderate conditioning;
- 5 unimodal functions with high conditioning;
- 5 multi-modal functions with adequate global structure;
- and 5 multi-modal functions with weak global structure.

Additionally, for each of these 24 functions, there are 15 instances defined by different translation and rotation transformations over the original function. The noiseless test-bed, described in detail in [Hansen et al., 2009a], totalizes thus 360 different function instances.

The BBOB framework enables experimentation on different dimensions (although the post-processing scripts, by default, consider only \( d \in \{2, 3, 5, 10, 20, 40\} \)). As a representative set, experiments were done for \( d \in \{5, 20\} \). But for the shorter dimension, the results attained are much less interesting: the problems are too quickly solved; consequently, not much significant difference can be observed between the performance of the different AOS schemes. For this reason, only the results for \( d = 20 \) will be reported here; the results on \( d = 5 \) can be found in [Fialho and Ros, 2010]. The stopping conditions of each optimization run are: the achievement of the optimum solution \( f_{opt} \) (with a tolerance of \( 10^{-8} \)), or the maximum number of function evaluations attained, this latter being fixed to \( 10^5 \cdot d \).

An informative measure of performance used in this experimental framework is the so-called Expected Running Time (ERT), which can be defined as follows: given a target function value, ERT is the empirical expected number of function evaluations for achieving a fitness value below the target. Formally, it is measured as the ratio of the number of function evaluations for reaching the target value over successful trials, plus the maximum number of evaluations for unsuccessful trials, divided by the number of successful trials. In addition to the standard ECDF plots used throughout this Chapter, a different kind of plot will be used here, the ECDF-ratio, which clearly depicts the speed-up ratio of one technique with respect to the others.

All the combinations involving rank-based and Extreme-based Credit Assignment schemes were tried on this experimental setting and will be compared in the following. Besides, the PM Operator Selection technique (Section 4.4.1) will also be considered here, but combined with a different Credit Assignment scheme, the average of relative fitness improvements. This combination, proposed in [Gong et al., 2010] and referred to as PM-AdapSS-DE, is a preliminary outcome of our on-going collaboration with the China University of Geosciences, which later motivated the assessment of the bandit-based approaches on this domain [Fialho et al., 2010b]. For the sake of self-containedness, the PM-AdapSS-DE method will be reminded in Section 6.6.3.

As done for the other empirical analyses presented in this Chapter, each AOS combination had its hyper-parameters tuned off-line prior to the experiments used to gather the
comparative results. The same tuning procedure defined in Section 6.2.2 was used, independently for each dimension. The only difference is that each elimination round happens after one run over all the function instances, what in fact corresponds to 360 performance results for each of the AOS schemes under comparison, up to 11 runs over all instances or one configuration left. The best hyper-parameter configuration found on dimension 20 for each AOS combination is presented in Table 6.24.

<table>
<thead>
<tr>
<th>Credit/OpSel</th>
<th>SLMAB</th>
<th>DMAB</th>
<th>MAB</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbsExt</td>
<td>C100W500</td>
<td>C100G.1W10</td>
<td>C100W500</td>
<td>P.2A.1B.3W500</td>
</tr>
<tr>
<td>OpSel/Credit</td>
<td>AUC (Decay)</td>
<td>SR (Decay)</td>
<td>AUC (NDCG)</td>
<td>SR (NDCG)</td>
</tr>
<tr>
<td>RMAB ($\Delta F$)</td>
<td>C.5D1W100</td>
<td>C.5D.75W50</td>
<td>C.5W50</td>
<td>C.5W50</td>
</tr>
<tr>
<td>RMAB ($F$)</td>
<td>C.5D.9W50</td>
<td>C.5D.5W50</td>
<td>C.5W50</td>
<td>C.5W50</td>
</tr>
</tbody>
</table>

Table 6.24: Hyper-parameter configurations used on BBOB dimension 20

Besides the comparison between the different AOS methods proposed here, the best AOS combination will be further compared with some different baseline approaches, namely, the Naive uniform selection between the same four strategies, and four variants of DE, each one applying only one of the considered strategies.

Besides, two different off-line tuned Static strategies will also be considered: the “global” one, which was tuned according to the performance over all problems; and the “local” one, which was independently tuned for each of the different function classes. The respective configurations found after the off-line tuning phase are presented in Table 6.25.

<table>
<thead>
<tr>
<th>tuning scenario</th>
<th>DE1 rand/1</th>
<th>DE2 rand/2</th>
<th>DE3 rand-to-best</th>
<th>DE4 current-to-rand</th>
</tr>
</thead>
<tbody>
<tr>
<td>separable</td>
<td>0</td>
<td>0</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>moderate</td>
<td>0</td>
<td>0</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>ill-condition</td>
<td>0</td>
<td>0</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>multi-modal</td>
<td>20%</td>
<td>0</td>
<td>20%</td>
<td>60%</td>
</tr>
<tr>
<td>weak-struct.</td>
<td>0</td>
<td>0</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td>all functions</td>
<td>0</td>
<td>0</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 6.25: Off-line tuned application rates for each mutation strategy on different tuning scenarios for BBOB with dimension 20.

Finally, a kind of optimal baseline is defined by a state-of-the-art continuous optimizer, the CMA-ES with an Increasing POPulation size restart strategy (IPOP-CMA-ES) [Auger and Hansen, 2005]. This algorithm was tested with the same parameter tuning as used in [Hansen, 2009a].

It is important to note that, differently from the boolean benchmark problems, in this experimental setting the fitness function should be minimized.
6.6.3 The PM-AdapSS-DE Method

The PM-AdapSS-DE AOS method uses as Operator Selection mechanism the Probability Matching (PM), described in Section 4.4.1. Equally motivated by the need of a higher robustness in order to be efficient in a variety of different problems with the same hyperparameter configuration, its Credit Assignment employs a different kind of normalization scheme, that takes place on the impact measurement level; while our Normalized schemes (Section 5.2.3) do so in the Credit Assignment output level.

The relative fitness improvement $\eta_i$, proposed in [Ong and Keane, 2004], measures the impact of an operator application as:

$$\eta_i = \frac{\delta}{c_f} \cdot |p_f - c_f|$$ (6.1)

where $i = \{1, \cdots, NP\}$ refers to each individual of the population of size $NP$, $\delta$ is the fitness of the best-so-far solution in the population, and $p_f$ and $c_f$ represent, respectively, the fitnesses of the target parent and of the generated offspring. In case of no improvement (i.e., the offspring is worse than or equal to its target parent), the impact is assessed as being null. Finally, the credit assigned to each strategy is the Absolute Average value of these impact measures.

Besides the relative measure, a main difference of this AOS method with respect to the other methods tried in this thesis is that it assigns credit to the operators and updates their empirical estimates only once per generation, based on their production during the given generation. Hence, there is no hyper-parameter $W$ on the Credit Assignment side; the only hyper-parameters that remain to be tuned are the ones from PM: the minimal probability of selecting each operator $p_{min}$, and the adaptation rate $\alpha$. These hyper-parameters were also off-line tuned, in the same way as for the other methods, using the ranges of values defined for AP in Table 6.4. The best configuration found by F-Race uses $p_{min} = 0$ and $\alpha = 0.6$.

Although being a quite simple method, a good performance is achieved mainly due to its robust Credit Assignment. However, as discussed in Section 5.2.4, it is still based on the raw values of the fitness improvements, thus not being as robust as our proposed rank-based Credit Assignment schemes, as shown in the empirical comparison that will be presented in the following.

6.6.4 Empirical Results

As previously mentioned, a complete set of experiments on this scenario involves 1 run on each of the 15 instances for each of the 24 functions, thus summing up to 360 results for each technique. Given this huge quantity of numerical data, it would be meaningless to present the detailed results for each function in the form of Tables, as done for the previous benchmark scenarios. Standard ECDFs and ECDF speed-up ratio plots are used instead, summarizing the results for the functions altogether, and for each function class.

The main objective of these experiments is to confirm the expectation that, based on the very robust rank-based Credit Assignment and on the optimal EvE balance provided by the bandit-based Operator Selection, the combinations involving RMAB and the different
6.6 On Continuous Benchmark Problems

Versions of AUC and SR should perform better than all the other considered methods on such an heterogeneous scenario. In this way, thus, both the performance and the robustness are jointly assessed, as all methods will use a single hyper-parameter configuration over all functions.

This empirical comparison will be performed in four different steps, as follows. Firstly, a representative configuration for the rank-based AOS schemes will be chosen; then it will be compared with the variants of DE using a single mutation strategy; also with other AOS schemes considered in this thesis; and finally with further baseline approaches. Results for each of these steps are summarized, respectively, in Figures 6.16, 6.17, 6.18, and 6.19; they will be now discussed in turn.

Selection of a Representative for the Rank-based Methods

Different alternatives for rank-based Credit Assignment were proposed in Sections 5.2.4 and 5.2.5, namely, Area-Under-Curve (AUC) and Sum-of-Ranks (SR), assigning ranks over fitness improvements, and Fitness-based Area-Under-Curve (FAUC) and Fitness-based Sum-of-Ranks (FSR), which are comparison-based methods that use ranks over fitness values. For each of them, there are still two options for the decaying mechanism, the parameterized one, referred to as Decay, and the parameter-less NDCG, which is equivalent to Decay with \( d = 0.4 \). The total number of possibilities proposed and tried in this thesis, always in combination with the RMAB Operator Selection technique, thus sums up to 8. All of them were compared with one another. The ECDF for each of them, for all results over all functions, and separately for each function class, are presented in Figure 6.16. As can be seen from these aggregated results, their behavior is rather the same; the choice of the method to be compared with the other baseline approaches is thus determined by convenience, as follows.

Firstly, the comparison-based feature provided by the methods that use ranks over the fitness values (FAUC, FSR) can not be precisely assessed on this scenario, as there are no instances being defined by monotonous transformations over the original function [Hansen et al., 2009a]. Therefore, it becomes interesting to note that, although providing such extra level of robustness, these methods still perform equivalently to the other versions that use ranks over fitness improvements (AUC and SR), while in the previous experimental scenarios the latter frequently outperformed the former. Hence, it is preferred to keep the comparison-based ones. Then, between FAUC and FSR, the first is chosen, by having shown a better performance on a couple of scenarios, although not being significantly different in most cases. Finally, the NDCG version is preferred, by being parameter-less. The chosen rank-based AOS combination is thus NDCG/FAUC-RMAB, which will be referred to as FAUC-B in the following, for the sake of brevity.

Comparison of FAUC-B with DE using a Single Strategy

FAUC-B is firstly compared with the standard DE algorithm using a single strategy, with four variants (\( DE1 \ldots DE4 \)), each one using one of the four mutation strategies considered by the AOS schemes, as defined in Section 6.6.2.
Figure 6.16: Standard ECDF plots of the distribution of ERT over dimension for combinations of RMAB with all the rank-based credit assignment schemes, for all functions and sub-groups with $d = 20$ and target $10^{-8}$.
6.6 On Continuous Benchmark Problems

Figure 6.17: ECDFs of the speed-up ratios in dimension $d = 20$ for the FAUC-B compared with the DE using only one out of the four available mutation strategies. The speed-up ratios are the pairwise ratios of the number of function evaluations for FAUC-B to surpass the target function value $10^{-8}$, over the one of the baseline techniques, over all trials for each function. Pairs where both trials failed are disregarded, pairs where one trial failed are visible in the limits being $> 0$ or $< 1$ (for this reason, some lines are not visible, as they coincide with the axes). The legends also indicate the number of functions that were solved in at least one trial (FAUC-B first).
The global picture (Figure 6.17a) shows in the legends that FAUC-B, DE1 and DE3 are able to solve at least one instance for 15 functions out of 24, while DE2 solves 12. DE4 shows unable to solve any of the functions on this dimension; interestingly, it is the mutation strategy mostly used by the off-line tuned Static strategies (see Table 6.25). A tentative explanation for this is that DE4 is a local search operator (current-to-rand; the base vector is the current individual), which tends to get trapped in local optima more easily; but it becomes efficient when mixed with a diversification operator, such as DE3 (rand-to-best; the base vector is a random one). The DE4 strategy will hence be neglected in the remainder of this analysis.

Compared with \textit{DE1}, FAUC-B shows to be around 3 times faster in approximately 90\% of all the considered cases. \textit{DE2} is around 20 times slower than FAUC-B on around 65\%, 50\%, 80\% and 40\% of the trials, respectively, for the separable, moderate, ill-conditioned and weak-structure function classes, while not solving any instance for the multi-modal class. \textit{DE3} is the best one out of the single strategies, performing roughly 10 times faster than \textit{DE2}; overall, it is around 2 times slower than FAUC-B.

It is worth noting that all the schemes failed on most multi-modal functions. Two out of the five separable functions are multi-modal and were not solved by any of the schemes. In the multi-modal class, only two out of five functions were solved by some of the schemes. And in the weak-structure class, only one out of five functions were solved in at least one trial.

\textbf{Comparison of FAUC-B with other AOS Schemes}

The second series of experiments compare FAUC-B with the other bandit-based approaches proposed in this thesis, namely, MAB, DMAB, and SLMAB, and with the baseline AP, all using the Absolute Extreme Credit Assignment. It is also compared with PM-AdapSS-DE, the method that uses PM being fed by Absolute Average of relative fitness improvements, detailed in Section 6.6.3.

Globally speaking (Figure 6.18a), FAUC-B is approximately 1.5 times faster than most of the other AOS schemes in around 90\% of the trials, except for the PM-AdapSS-DE, which is outperformed in approximately half of the trials, performing faster than FAUC-B in around 15\% of the function trials. More specifically, FAUC-B is up to 10 times faster than the standard MAB in half of the trials, while being 2 times faster than DMAB and SLMAB in around 75\% of the trials.

This global assessment corresponds roughly to the speed-up ratios attained on the separable, moderate and ill-conditioned function classes. Again, none of the schemes are able to perform well on multi-modal functions, with only 2 functions being solved by all schemes in the multi-modal class, and only 1 in the harder weak-structure class. Although the low number of successful trials on these latter classes, FAUC-B still outperforms all the schemes in most trials.
6.6 On Continuous Benchmark Problems

Figure 6.18: ECDFs of the speed-up ratios in dimension $d = 20$ for the FAUC-B compared with other AOS combinations. The speed-up ratios are the pairwise ratios of the number of function evaluations for FAUC-B to surpass the target function value $10^{-8}$, over the one of the baseline techniques, over all trials for each function. The legends indicate the number of functions that were solved in at least one trial (FAUC-B first).
Comparison of FAUC-B with further Baseline Approaches

The last empirical comparison, whose results are illustrated in Figure 6.19, considers three different situations. On one side, there is the Naive uniform operator selection. FAUC-B shows to be around 1.5 times faster than Naive-DE on around 80% of the trials for all unimodal functions in the separable, moderate and ill-conditioned function classes. It attains the same speed-up ratio on approximately 65% and 55% of the trials, respectively, for the multi-modal and weak-structure function classes.

On the other side, there is the state-of-the-art continuous optimizer IPOP-CMA-ES [Auger and Hansen, 2005], which significantly outperforms FAUC-B in around 90% of the trials for all cases. Besides, it succeeds in solving all the 5 functions for the multi-modal class, while FAUC-B and all the other schemes previously considered solve only 2; and it also solves 2 functions for the weak structure class, one more than all the other schemes.

In the middle, there are the two off-line tuned Static strategies. The best static set of application rates for each case is presented in Table 6.25. As can be seen, both, the globally (StAll) and the locally (StEach) tuned variants, perform rather equivalently to FAUC-B in all cases.

6.6.5 Discussion

Differently from the other scenarios previously analyzed in this Chapter, the use of the BBOB experimental framework enabled a much broader and realistic assessment of the proposed AOS schemes. By the evaluation of the methods on the many functions with different characteristics and levels of difficulty, treated as black-boxes, it showed to be an ideal scenario in order to depict the gain in robustness brought by the use of rank-based Credit Assignment schemes.

And indeed, the NDCG/FAUC-RMAB (simply referred to as FAUC-B) AOS method, a representative of all the proposed rank-based methods, succeeded in outperforming the baseline approaches in the vast majority of the cases: the standard DE using each of the considered mutation strategies, the Naive uniform strategy selection, all the other bandit-based approaches, as well as two other baseline adaptive schemes, Adaptive Pursuit and PM-AdapSS-DE. This performance improvement of the FAUC-B with respect to the others, in terms of Expected Running Time (ERT) to achieve a given function target value, is attributed mostly to: (i) the use of a rank-based Credit Assignment, which is robust with respect to the very different situations tackled within this benchmark suite, while being able to efficiently follow the changes in the qualities of the strategies; and to (ii) the use of a bandit-based Operator Selection, which has already shown to be very efficient in the other experimental settings. FAUC-B, as well as the other methods using ranks over the fitness values instead of fitness improvements, still provide an extra layer of robustness that was not challenged in this scenario, that of being comparison-based: in case there were instances derived by monotonous transformations over the original functions, the performance gap could be even bigger, possibly with the comparison-based methods (FAUC and FSR) significantly outperforming the simply rank-based methods (AUC and SR). This situation should be further addressed in the near future.
6.6 On Continuous Benchmark Problems

Figure 6.19: ECDFs of the speed-up ratios in dimension $d = 20$ for the FAUC-B compared with the Naive uniform operator selection, the off-line tuned Static strategies (stAll tuned over all functions, stEach tuned over each function class), and the state-of-the-art IPOP-CMA-ES optimizer. The speed-up ratios are the pairwise ratios of the number of function evaluations for FAUC-B to surpass the target function value $10^{-8}$, over the one of baseline techniques, over all trials for each function. The legends also indicate the number of functions that were solved in at least one trial (FAUC-B first).
Moreover, PM-AdapSS-DE [Gong et al., 2010] (Section 6.6.3), showed to be the best out of the four other adaptive schemes, what confirms the gain in robustness achieved by the use of a relative instead of a raw reward. It is also worth noticing that, as for the other simpler benchmark scenarios, DMAB greatly outperformed the standard MAB by the use of its change-detection test, while SLMAB was able to closely follow its performance, both of them performing very similarly to AP. But, although improving over MAB, both DMAB, SLMAB and AP techniques performed rather equivalently to the Naive uniform strategy, and this puts into question their efficiency in this experimental setting.

Compared to the Naive uniform approach, it is true that the efficiency gain presented by FAUC-B seems to be moderate in relation to the price to pay for an adaptive scheme in terms of computational complexity. Nevertheless, it should be observed that the Naive uniform strategy considers here a small number of strategies, most of which performing well: $DE_1$ and $DE_3$ perform quite well; although much slower, $DE_2$ still reaches the target; the only inefficient strategy (when applied alone) is $DE_4$. Along the same line, the performance of FAUC-B was found to be comparable to the off-line tuned Static strategies, which indeed perform very well by mixing two very complementary operators: the local search $DE_4$ (current-to-rand), which performs random variations over the current vector (what explains its poor performance when applied alone), and the $DE_3$ (rand-to-best) operator, which provides enough level of diversity, while still moving towards the best solution. In the general case, however, the performance and the characteristics of each strategy alone, and of all the possible static combinations, are unknown; they were assessed here through extensive experiments, while FAUC-B used only the information of the current run to adapt its behavior while solving the problem. And even if the computational effort spent on their off-line tuning is considered, FAUC-B remains a superior choice when compared to the given Static strategies: in this case, 28 configurations were tried for the AOS method (although much fewer could have been considered) while 56 configurations were tried for the off-line tuning of the Static strategies. The use of an AOS scheme remains thus relevant in the general case.

Additionally, though much improved over the results of all mentioned approaches used within DE, the best results of the FAUC-B +DE algorithm remains far below those of the state-of-the-art optimizer IPOP-CMA-ES. But the DE algorithm that FAUC-B has been applied to is the very standard one; several improvements have been recently proposed to it, e.g., adding adaptive parameter control for $F$ and $CR$ [Qin et al., 2009]. The applicability of FAUC-B in DE framework opens the path for fruitful research using the numerous recent DE variants.

Another further work is to address the multi-modality issue: all tested algorithms failed on most multi-modal functions (40% of the separable class, plus all the functions for the multi-modal and weak structure classes). In the same way as for the work on SAT problems (Section 6.5), in order to efficiently tackle multi-modal problems, the maintenance of some level of diversity in the population should also be accounted somehow for the rewarding of operator applications, as discussed in Section 4.5.2.
6.7 Hyper-Parameters Analysis

As discussed in Section 6.2.1, the AOS mechanisms analyzed in this work have their own hyper-parameters (Table 6.2), whose values might critically impact their performances. This issue has been addressed using the F-Race, and the best hyper-parameter configuration for each considered AOS has been determined and further used on each of the different empirical comparisons presented in this Chapter. It is, however, important to study the hyper-parameters, in order to identify which of them should receive more attention during the tuning process when addressing a new problem. An analysis of their sensitivity will be presented in Section 6.7.1.

Another important aspect concerning the hyper-parameter configuration of the AOS schemes is how robust they are when tackling different problems. Ideally, every time a new problem needs to be tackled, the AOS method under consideration should be able to deliver reasonable performance without requiring off-line tuning. The methods based on the raw values of fitness improvements are expected to be very problem-dependent, as discussed in Section 5.2.3, while the rank-based methods are expected to be more robust to different situations. An analysis of their robustness will be presented in Section 6.7.2.

The results that will be analyzed here were partially published in [Fialho et al., 2010a; Fialho et al., 2010c; Fialho and Ros, 2010].

6.7.1 On the Sensitivity of the Hyper-Parameters

When only 1 or 2 hyper-parameters are concerned, a 3-D plot of the response surface of these parameters gives a clear picture (as has been done in [Da Costa et al., 2008] for AP and DMAB on some artificial scenarios, for instance). However, for each AOS technique, there are the hyper-parameters of both Operator Selection and Credit Assignment schemes: only MAB and SLMAB combined with the Extreme or Average Credit Assignment schemes have 2 hyper-parameters; DMAB and AP, combined with the same Credit Assignment schemes, have respectively 3 and 4 hyper-parameters; and RMAB with any of the rank-based schemes has 3 (see the list in Table 6.2). This is why ECDF plots will also be used for this analysis, as described in the following. Another alternative could have been to use some parameter setting procedure like REVAC [Nannen and Eiben, 2007] (Section 3.3.2), that gives an idea of the sensitivity of all parameters it optimizes while finding its optimal value.

ECDF Sensitivity Plots

Though ECDF sensitivity plots have been generated for all AOS schemes and all scenarios previously presented, only a few series of plots offering typical behavior will be depicted and analyzed in detail here. All the non rank-based methods are considered with the AbsExt Credit Assignment. This scheme was chosen because it performs best in most cases; for a few exceptions, the Instantaneous performed slightly better, but the corresponding ECDF plots looked rather similar, though involving one less hyper-parameter, the size $W$ of the rewards window (expect for SLMAB, which needs $W$ anyway). For the rank-based
methods, the RMAB with the Decay/AUC Credit Assignment is considered. Accordingly, the other alternatives could have been tried, but they present rather similar performance in most scenarios; between Decay and NDCG, the former is preferred so as to be able to also analyze the decay factor $D$.

For each of these AOS combinations, its sensitivity with respect to its hyper-parameters will be analyzed on four benchmark problems, a kind of representative set of all the different problems considered in this thesis: the Uniform scenario with $\Delta T = 500$, the ART $(0.1, 39, 0.5, 3)$ with $\Delta T = 200$, and the OneMax and Royal Road boolean benchmark problems.

All figures display a number of ECDF curves representing the results of the same AOS method on the same scenario. For the artificial scenarios, the x-axis represents the TCR, ranging from the value of the optimal strategy (i.e., the highest possible value on average for any AOS) down to the value that would, on average, be gathered by the Naive uniform strategy (hopefully the lowest possible value on average for any AOS). For the OneMax and Royal Road cases, the x axis represents the number of generations to achieve the optimum, starting from 4500 for the former and zero for the latter problem, up to the average number of generations taken by the Naive uniform choice. The y-axis shows the proportion of runs that reached the corresponding x value, out of the total number of runs considered.

For each benchmark problem, a large amount of experimental results have been gathered during the Racing procedure, and at least 11 runs have been performed for all hyper-parameter combinations of the factorial design sketched by the values in Table 6.4. The results over these 11 runs will be used for the analysis on the OneMax and Royal Road problems; this explains possible divergences between the best configurations considered here and the ones presented in the tables of results corresponding to each scenario. Experiments on the Uniform and ART artificial scenarios are significantly cheaper, in terms of computational time; hence, for these scenarios, even the parameter configurations that did not make it through the end of the Racing procedure have been run 50 times for the sake of the sensitivity analysis.

Two lines are given as references on each plot: the top-left-most line (continuous, and green on color printouts), labeled with a “***” and referred to as “Best/All”, represents the overall best results, in terms of average TCR (or number of generations to the optimum in the OneMax and Royal Road cases), obtained on this scenario by a single hyper-parameter configuration, between all the AOS combinations, after the number of runs considered for each scenario. The next line going down/right, labeled with a “**” and referred to as “Best/This”, represents the results obtained by the best configuration of hyper-parameters for the AOS scheme named on the caption of the sub-figure, under the plot. The discrepancy between both lines shows in detail how different are the performances of the considered AOS and of that of the method that performed best on this scenario.

At the other extreme of each sub-figure, the bottom-right-most line (solid, red on color printouts) represents the ECDF of all runs for the particular AOS, i.e., for all hyper-parameter configurations ever tried (the average performance of the complete factorial design from values given on Table 6.4). The lines in-between represent partial aggregations.
of these runs. More precisely, if the best results have been obtained for a given set of hyperparameters (recalled in the legend), each line represents the ECDF obtained when one hyper-parameter is varied over all its values used in the Racing procedure, while all others are kept to the optimal value found. If a line corresponding to a given parameter is close to the line of the best configuration, it is a clear indication that the results are not very sensitive to this hyper-parameter. Oppositely, a high difference indicates a high sensitivity. Besides the ECDF curve, the average and standard deviation of the performance are also presented in the legends of the plots for each of the different aggregations considered.

Sensitivity Analysis

In order to facilitate the comparison of the impact of the hyper-parameters on the performance of each AOS technique on the different benchmark problems, and possibly find common sensitivity hints for their tuning, the plots are grouped by technique. Figure 6.20 presents the ECDF sensitivity plots for the baseline AbsExt-AP method, for the sake of comparison. Figure 6.21 depicts in each column the plots for both AbsExt-MAB and AbsExt-SLMAB, while Figure 6.22 shows the plots for AbsExt-DMAB on the right and Decay/AUC-RMAB on the left column.

Starting with the combination of Adaptive Pursuit (AP) Operator Selection and Absolute Extreme (AbsExt) Credit Assignment, the global picture with respect to sensitivity is rather clear. The adaptation rate $\alpha$ and the learning rate $\beta$ are very robust indeed: their aggregated ECDF plots are very close to the curves representing the best configuration for this technique on all problems. The minimal probability $p_{\text{min}}$ is a much more sensitive parameter, as expected (and as discussed in Section 4.4.2). But its sensitivity clearly (and intuitively) depends on the number of operators and on the maximum value tried for it: on the ART scenario it seems to be as insensitive as $\alpha$ and $\beta$ (Figure 6.20b); however, only two operators are considered in this case, thus the total exploration when using the maximum value tried for $p_{\text{min}} (= 0.2)$ sums up to “only” 40% of exploration; while on the Royal Road scenario, which has 5 operators, $p_{\text{min}}=0.2$ refers to a complete Naive uniform behavior. The window size $W$ is the most sensitive parameter on all cases. Altogether, AP seems to be quite robust with respect to its hyper-parameters, but its global performance, when compared to the bandit-based approaches on most benchmark scenarios considered in this thesis, make it a poor choice anyway.

The scaling factor $C$ and the sliding window size $W$ are common hyper-parameters for all bandit-based AOS combinations. Starting with $C$, for MAB, SLMAB and DMAB, it is definitely a very sensitive parameter. As discussed in Section 5.3.4, when using Credit Assignment schemes based on the raw values of fitness improvements, $C$ has a double role on the bandit-based approaches: besides controlling the Exploration versus Exploitation balance of the Operator Selection, it needs to account for the different ranges of fitness improvements (which tend to vary as the search goes on, and according to the problem at hand). For RMAB, $C$ still seems to be a sensitive hyper-parameter; however, as can be seen in the captions of the respective plots, the winner configurations for the different problems all use $C \leq 1$, while values up to 100 are being considered in the corresponding (dark blue) curve, consequently greatly degrading its aggregated performance. From this
alternative analysis, thus, it can be said that the use of a rank-based scheme fulfills its original motivation, that of providing an invariant range of rewards during all the search process on a given problem, and over different problems, what reflects in similar \( C \) values being used by the winner configurations. Conversely, the winner configurations of the other bandit-based approaches use very different values for \( C \) on the different problems. For instance, SLMAB uses \( C = 0.01 \) on the OneMax problem, and \( C = 100 \) on the ART problem. More on this will be discussed in the robustness analysis presented in Section 6.7.2.

The window size \( W \) also seems to be very sensitive for the bandit-based approaches, specially on the Uniform and ART scenarios, in which there is a strong link between \( W \) and \( \Delta T \): for values of \( W \) larger than the epoch size, the frequency of the changes will result in using too old information. This issue becomes even more important for the approaches other than RMAB, in which there is one window for each operator: in the worst case, a window might contain rewards as old as \( K \cdot W \) iterations ago, \( K \) being the number of operators. For instance, the “steps” shown in the corresponding ECDF curves for MAB,
6.7 Hyper-Parameters Analysis

(a) AbsExt-MAB on Uniform $\Delta T = 500$

(b) AbsExt-SLMAB on Uniform $\Delta T = 500$

(c) AbsExt-MAB on $\mathcal{A}RT(0.1, 0.3, 0.5, 3)$

(d) AbsExt-SLMAB on $\mathcal{A}RT(0.1, 0.3, 0.5, 3)$

(e) AbsExt-MAB on OneMax

(f) AbsExt-SLMAB on OneMax

(g) AbsExt-MAB on Royal Road

(h) AbsExt-SLMAB on Royal Road

Figure 6.21: ECDF sensitivity plots for AbsExt-MAB and AbsExt-SLMAB
Chapter 6. Experimental Results

DMAB and RMAB on the Uniform scenario (respectively, on Figures 6.21a, 6.22a and 6.22b), clearly depict how large values for $W$ are hindering the overall performance of the aggregated distribution in this case. This hyper-parameter becomes less sensitive in the more realistic OneMax and Royal Road scenarios, in which there is no clear relation between operators qualities and time. For SLMAB, the $W$ hyper-parameter has a double role: it controls both the size of the sliding window and the size of the intrinsic memory of the relaxation update rule used on the Operator Selection side; for this reason, it seems to be much more sensitive to this hyper-parameter than the other bandit-based approaches on three out of the four cases (except for the ART, in which it is as sensitive as for the other approaches). More specifically, in the case of RMAB, as there is only one window for all operators, (intuitively) larger window sizes are preferred.

The DMAB change-detection threshold $\gamma$ shows not to be sensitive on the Uniform and ART scenarios. A tentative interpretation goes as follows. As the range of rewards for the best operator are the same during the whole search process, on the Uniform scenario it seems that $3/4$ of values tried for DMAB $\gamma$ perform well (see again the “step” on around 75% of the cyan-colored distribution in Figure 6.22a). Indeed, in this case, from the values explored, only $\lambda \in \{100, 1000\}$ achieve a TCR much worse than the others, as these thresholds are too big in relation to the actual changes in the reward distributions, thus not triggering any restart. Along the same lines, on the ART scenario, whenever there is an exchange in the reward distributions, i.e., whenever an outlier reward (of value 39 in this case) is received, as it is much higher than the expectation of the previous distribution ($= 3$), most values tried will succeed in detecting the change. For the OneMax and Royal Road problems, the best operator might present very different expected reward during the search process, thus there is no value for $\gamma$ that works optimally during all the search; for this reason, in these cases, the $\gamma$ threshold appears to be as sensitive as the very sensitive scaling factor $C$. In more realistic scenarios, having different reward ranges during the search is very likely to be the case. Hence, this hyper-parameter tends to hinder the performance of DMAB (although still outperforming the standard MAB in most cases), while requiring a considerable amount of computational time for its off-line tuning when compared to the other methods.

Accordingly, the RMAB ranking decay factor $D$ is also much less sensitive on the Uniform scenario: as the rewards received are always already ranked somehow according to the quality of the operator (despite some overlap between subsequent operators), the operator qualities are already ranked by construction, thus not needing an extra decaying factor in order to efficiently differ between them. Anyway, even for the other scenarios, it does not seem to be as sensitive as the other hyper-parameters; expect for the Royal Road case, in which it is surprisingly more sensitive than $W$, although much less sensitive than the scaling factor $C$. Another empirical proof of the non-sensitivity of this hyper-parameter is related to the fact that, in most empirical performance comparisons presented in this Chapter, the Decay version using some high value for $D$ showed equivalent performance to the $NDCG$ version, which is basically the same as using Decay with $D = 0.4$, as shown in Figure 5.1. It is also worth noticing that the fact that this hyper-parameter is always bounded between 0 and 1 results in a much easier and cheaper off-line tuning than those of the scaling factor $C$ and the change-detection threshold $\gamma$, which have no known bounds.
6.7 Hyper-Parameters Analysis

Figure 6.22: ECDF sensitivity plots for AbsExt-DMAB and Decay/AUC-RMAB.
Chapter 6. Experimental Results

6.7.2 On the Robustness of the Hyper-Parameters

The sensitivity of a given hyper-parameter might be alleviated by the robustness of the method with respect to many different problems, *i.e.*, even if a lot of effort is needed for the preliminary off-line tuning of this parameter, in case the configuration found by the off-line tuning performs reasonably well for several different problems, the computational budget spent on this initial effort might become worth the expense.

In this Section, the robustness of the AOS schemes with respect to their hyper-parameters is analyzed on two experimental scenarios: firstly, on the OneMax problem and on 3 other problems defined by monotonous transformations of OneMax; then, on the very heterogeneous set of functions provided by BBOB.

Robustness on Transformations over the OneMax Problem

A first series of experiments was conducted, based on the OneMax problem, to analyze the expected gain in robustness provided by the use of the rank-based *Credit Assignment* schemes and, specially, the invariance with respect to monotonous transformations (leading to the so-called comparison-based property) featured by the schemes using ranks over the fitness values instead of ranks over the fitness improvements.

As presented in Section 6.4.2 (and as done for every performance benchmarking scenario considered in this Chapter), the F-Race procedure was first used to tune the hyper-parameters of all the AOS schemes prior to the comparison of their average empirical performance on the OneMax problem. For the robustness analysis that will be presented here, this same hyper-parameter configuration found to be the best on the OneMax problem for each AOS method, was used for assessing it on functions defined by monotonous non-linear transformations over the original OneMax function \( F \), namely, \( \log(F) \), \( \exp(F) \), and \( F^2 \).

The complete results, gathered with the same experimental setting used in the original performance comparison presented in Section 6.4.2, can be found in Table 6.26. The average performance (number of generations to optimum) on the original and on the three transformed functions are presented for each AOS scheme. These performance measures are ranked independently for each function, and the first column of the Table summarizes these ranks, by presenting their sum. The Table is sorted by this column: the more robust technique is the one with lowest values for \( \sum r \). Additionally, the second column presents the gap between the worst and the best performance achieved by each AOS scheme over all four functions, what can be seen as a complementary view of their robustness with respect to this experimental scenario. Finally, as done in the previously analyzed performance comparisons, the best result for each function is highlighted with bold-face and grey background, like *this*, and the results which are not significantly different (according to the same statistical tests) are displayed with a grey background.

These results empirically confirm several expectations. Firstly, the rank-based methods that use the rank of the fitness improvements, AUC-RMAB and SR-RMAB, achieve an overall better performance, while showing to be quite robust with respect to the monotonous transformations. Although the comparison-based counterparts FAUC-RMAB
### Table 6.26: Results of each AOS scheme on the original OneMax function and on three other functions defined by monotonous transformations over it. The schemes were ranked on each function, and the first column presents the sum of their ranks, which defines the order of their presentation in the table. The second column depicts the difference between the highest and the lowest average performance of each AOS scheme over all the four functions.

<table>
<thead>
<tr>
<th>$\sum r$</th>
<th>$F = \sum b_i$</th>
<th>$\log(F)$</th>
<th>$\exp(F)$</th>
<th>$(F^2)$</th>
<th>AOS technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>485</td>
<td>5103 427</td>
<td>5195 430</td>
<td>5562 950</td>
<td>5588 950</td>
</tr>
<tr>
<td>13</td>
<td>244</td>
<td>5231 503</td>
<td>5421 524</td>
<td>5475 422</td>
<td>5475 422</td>
</tr>
<tr>
<td>15</td>
<td>362</td>
<td>5215 374</td>
<td>5347 547</td>
<td>5577 634</td>
<td>5575 534</td>
</tr>
<tr>
<td>18</td>
<td>1321</td>
<td>5097 230</td>
<td>6295 1176</td>
<td>4974 201</td>
<td>4984 184</td>
</tr>
<tr>
<td>19</td>
<td>108</td>
<td>5444 252</td>
<td>5458 382</td>
<td>5511 347</td>
<td>5403 332</td>
</tr>
<tr>
<td>24</td>
<td>300</td>
<td>5366 478</td>
<td>5434 596</td>
<td>5630 887</td>
<td>5666 881</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>5652 644</td>
<td>5652 644</td>
<td>5652 644</td>
<td>5652 644</td>
</tr>
<tr>
<td>31</td>
<td>807</td>
<td>5123 218</td>
<td>5431 223</td>
<td>5930 334</td>
<td>5930 334</td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>5667 729</td>
<td>5667 729</td>
<td>5667 729</td>
<td>5667 729</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>5726 399</td>
<td>5726 399</td>
<td>5726 399</td>
<td>5726 399</td>
</tr>
<tr>
<td>46</td>
<td>54</td>
<td>5728 204</td>
<td>5767 312</td>
<td>5780 329</td>
<td>5769 263</td>
</tr>
<tr>
<td>50</td>
<td>172</td>
<td>5718 239</td>
<td>5805 279</td>
<td>5748 227</td>
<td>5890 386</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>5796 420</td>
<td>5796 420</td>
<td>5796 420</td>
<td>5796 420</td>
</tr>
<tr>
<td>54</td>
<td>186</td>
<td>5759 254</td>
<td>5782 251</td>
<td>5936 277</td>
<td>5875 287</td>
</tr>
<tr>
<td>62</td>
<td>355</td>
<td>5790 226</td>
<td>5910 271</td>
<td>6048 337</td>
<td>6145 319</td>
</tr>
<tr>
<td>65</td>
<td>2591</td>
<td>5376 285</td>
<td>7967 718</td>
<td>7722 2151</td>
<td>6138 516</td>
</tr>
<tr>
<td>74</td>
<td>706</td>
<td>6427 597</td>
<td>6956 784</td>
<td>7133 866</td>
<td>6808 691</td>
</tr>
<tr>
<td>77</td>
<td>3310</td>
<td>5508 823</td>
<td>8818 3653</td>
<td>7173 3288</td>
<td>6865 2861</td>
</tr>
<tr>
<td>80</td>
<td>2673</td>
<td>5480 276</td>
<td>8079 743</td>
<td>7961 653</td>
<td>8153 684</td>
</tr>
<tr>
<td>80</td>
<td>502</td>
<td>7193 1614</td>
<td>6890 1294</td>
<td>7370 1033</td>
<td>7392 1117</td>
</tr>
<tr>
<td>81</td>
<td>2285</td>
<td>5997 503</td>
<td>6345 644</td>
<td>8282 2044</td>
<td>7639 1807</td>
</tr>
<tr>
<td>83</td>
<td>1905</td>
<td>6662 961</td>
<td>8277 1553</td>
<td>6507 839</td>
<td>6372 786</td>
</tr>
<tr>
<td>92</td>
<td>225</td>
<td>8013 671</td>
<td>7893 681</td>
<td>8021 582</td>
<td>8118 738</td>
</tr>
<tr>
<td>96</td>
<td>365</td>
<td>7903 638</td>
<td>8023 716</td>
<td>8268 949</td>
<td>8008 874</td>
</tr>
<tr>
<td>97</td>
<td>631</td>
<td>7494 611</td>
<td>8122 724</td>
<td>8125 1159</td>
<td>8075 740</td>
</tr>
<tr>
<td>108</td>
<td>6971</td>
<td>6059 667</td>
<td>8863 694</td>
<td>13030 3053</td>
<td>12136 949</td>
</tr>
<tr>
<td>111</td>
<td>7089</td>
<td>8198 683</td>
<td>7910 549</td>
<td>14999 0</td>
<td>14999 0</td>
</tr>
<tr>
<td>111</td>
<td>8423</td>
<td>6576 705</td>
<td>8721 695</td>
<td>14999 0</td>
<td>9838 1430</td>
</tr>
<tr>
<td>112</td>
<td>182</td>
<td>8463 818</td>
<td>8593 753</td>
<td>8577 737</td>
<td>8645 862</td>
</tr>
<tr>
<td>112</td>
<td>7083</td>
<td>8369 891</td>
<td>7916 635</td>
<td>14999 0</td>
<td>14999 0</td>
</tr>
<tr>
<td>113</td>
<td>7052</td>
<td>9044 840</td>
<td>7947 1267</td>
<td>14999 0</td>
<td>14999 0</td>
</tr>
<tr>
<td>114</td>
<td>845</td>
<td>8347 596</td>
<td>8899 808</td>
<td>9192 862</td>
<td>8395 721</td>
</tr>
</tbody>
</table>
Chapter 6. Experimental Results

and FSR-RMAB are less competitive, their invariance property is verified (exactly the same performance on all the functions, \((h - l) = 0\)); this might show to be more beneficial in a bigger and more difficult class of problems. For all the other approaches, the best results in terms of robustness are achieved using Normalized versions of the Extreme (or Instantaneous for DMAB) Credit Assignment schemes; this also confirms the expectations after discussion in Section 5.2.3.

Between the Operator Selection techniques other than RMAB, the baseline probability-based AP approach is the most robust: all its combinations are ranked in the first half of the Table. Its best combination in terms of performance, the NormExt-AP, achieves the best result in three out of four functions; however, it does not cope well with such a simple transformation as the logarithmic one. This result clearly demonstrates that the lack of invariance under simple nonlinear transformation could eventually cause some serious loss of efficiency for more difficult problems.

In what concerns the other bandit-based approaches using Credit Assignment schemes based on the raw values of fitness improvements, only the NormIns-DMAB is surprisingly able to perform well on this experimental scenario. A huge variance is shown in the results of the other combinations involving DMAB and MAB. Although presenting a smaller variance, SLMAB is still a bad choice, due to its non-competitive performance, populating mostly the bottom of the ranked Table, together with MAB.

Robustness on the BBOB Functions

An alternative analysis of the robustness of each AOS technique with respect to its hyper-parameters was done in the context of the BBOB benchmark scenario. Here, instead of using the same hyper-parameter configuration over different problems, the opposite approach is taken: different off-line tuning procedures were done, considering different groups of functions, and the best hyper-parameter configurations found for each of them are compared. A robust technique should present similar best hyper-parameter setting on most of the different cases, while still presenting competitive performance.

In the same way as for the empirical performance analysis presented in Section 6.6, the NDCG/FAUC-RMAB is used here as a representative of all the rank-based AOS schemes, being simply referred to as FAUC-B. MAB and SLMAB are disregarded here, due to their poor performance. The robustness of FAUC-B is compared to those of PM-AdapSS-DE, AP and DMAB, these two latter being coupled with the AbsExt Credit Assignment scheme.

Six different tuning procedures were performed for each dimension \(d \in \{5, 20\}\), which are: tuning considering independently each of the 5 function classes; and tuning considering all functions. The best configuration found for each technique on each of the analyzed cases is presented in Table 6.27.

This tuning experiment confirms again the higher robustness of FAUC-B: \(\{C = .5, W = 50\}\) is always the best configuration; except for the multi-modal and weak-structure class functions, in which none of the techniques was able to perform well, as discussed in Section 6.6. For the PM-AdapSS-DE, the benefits of using a relative instead of a raw reward are also shown on dimension 20, with always a very low value for \(p_{min}\), and a high one for the
Table 6.27: Robustness analysis: best hyper-parameters configuration found for each technique on the BBOB benchmark set for dimensions 5 and 20, off-line tuned under different conditions.

<table>
<thead>
<tr>
<th>$d = 5$</th>
<th>FAUC-B</th>
<th>PM-AdapSS-DE</th>
<th>AbsExt-AP</th>
<th>AbsExt-DMAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>separable</td>
<td>C.5W50</td>
<td>P.05α.9</td>
<td>P.2α.6/3.3W50</td>
<td>C10γ.01W100</td>
</tr>
<tr>
<td>moderate</td>
<td>C.5W50</td>
<td>P.05α.3</td>
<td>P.2α.9/3.3W50</td>
<td>C01G100W50</td>
</tr>
<tr>
<td>ill-conditioned</td>
<td>C.5W50</td>
<td>P.05α.9</td>
<td>P.2α.6/3.3W50</td>
<td>C100G1000W500</td>
</tr>
<tr>
<td>multi-modal</td>
<td>C.5W50</td>
<td>P.05α.9</td>
<td>P.2α.9/3.3W50</td>
<td>C100G1W50</td>
</tr>
<tr>
<td>weak-structure</td>
<td>C1W500</td>
<td>P.05α.1</td>
<td>P.2α.9/3.6W500</td>
<td>C100γ.1W50</td>
</tr>
<tr>
<td>all functions</td>
<td>C.5W50</td>
<td>P.0α.9</td>
<td>P.2α.3/3.6W500</td>
<td>C100γ.1W50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d = 20$</th>
<th>FAUC-B</th>
<th>PM-AdapSS-DE</th>
<th>AbsExt-AP</th>
<th>AbsExt-DMAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>separable</td>
<td>C.5W50</td>
<td>P.0α.9</td>
<td>P.2α.3/3.1W500</td>
<td>C100γ.01W50</td>
</tr>
<tr>
<td>moderate</td>
<td>C.5W50</td>
<td>P.0α.9</td>
<td>P.2α.9/3.3W500</td>
<td>C100γ.01W50</td>
</tr>
<tr>
<td>ill-conditioned</td>
<td>C.5W50</td>
<td>P.0α.9</td>
<td>P.2α.6/3.3W500</td>
<td>C100γ.01W50</td>
</tr>
<tr>
<td>multi-modal</td>
<td>C1W50</td>
<td>P.0α.3</td>
<td>P.2α.3/3.1W100</td>
<td>C100γ.01W50</td>
</tr>
<tr>
<td>weak-structure</td>
<td>C1W50</td>
<td>P.0α.9</td>
<td>P.2α.3/3.3W100</td>
<td>C100γ.01W50</td>
</tr>
<tr>
<td>all functions</td>
<td>C.5W50</td>
<td>P.0α.6</td>
<td>P.2α.1/3.1W500</td>
<td>C100γ.1W50</td>
</tr>
</tbody>
</table>

adaptation rate $\alpha$.

For AP, however, several configurations reached the end of the Racing process, all of them sharing $P.2$ and $W.500$, but presenting all possible combinations for the adaptation rate $\alpha$ and the learning rate $\beta$. This could be seen as a good sign, possibly showing the robustness of this AOS combination. However, the use of $p_{\min}=0.2$ in fact confirms that the method is presenting a behavior very close to the Naive uniform approach: as 4 mutation strategies are considered in this experimental scenario, the completely uniform behavior would be achieved with $p_{\min}=0.25$, regardless the other parameters.

The same kind of conclusions can be drawn for the tuning experiments of DMAB. The configurations found for the different situations were all quite similar, with $C100$, $\gamma \leq .1$ and $W.50$. However, a very high scaling factor $C$ was found to be the best; this means that much more weight was given to the exploration term of the UCB formula, *i.e.*, although knowing which is the current best strategy, the algorithm prefers to explore the others. Besides, a very low value for the Page-Hinkley change detection threshold $\gamma$ was chosen in most cases; this means that the probability of having restarts during the search is really high, also favoring the exploration, consequently dramatically degrading the performance of the method.

After all, the single fact that the same hyper-parameter tuning is found to be the best on different situations is not sufficient to conclude that a given technique is robust. Intuitively, if the final performance is as good as the uniform one, the configurations found are meaningless. The FAUC-B and the PM-AdapSS-DE, while presenting similar configurations for different situations, also perform very well, as shown in the empirical comparison presented in Section 6.6.
6.8 General Discussion

The use of so many and distinct benchmark scenarios was motivated by the possibility of analyzing different properties of the AOS schemes. On the artificial scenarios (Section 6.3), for instance, the agility to adapt to completely different situations was assessed under different conditions with respect to the definition of the artificial reward distributions and to the level of informativeness of the received rewards. On the other hand, the results on the boolean benchmark problems (Section 6.4) represent the preliminary experiments in which the AOS schemes were analyzed in situ, selecting between different evolutionary operators within a real EA, applied to fitness landscapes with very different characteristics and levels of complexity. Additionally, in Section 6.5, the DMAB Operator Selection technique was evaluated in the light of the hard combinatorial Boolean Satisfiability (SAT) problems, by using a third party Credit Assignment scheme that aggregates both fitness and diversity. Finally, on the Black-Box Optimization Benchmarking (BBOB) scenario, not only the performance of the techniques was assessed, but also their robustness, given the very heterogeneous benchmark function set provided by it. Besides, in this case, yet another problem domain was tackled, the continuous one, by coupling the AOS schemes with a Differential Evolution (DE) algorithm.

After the description of the contributions for AOS in Chapter 5, these experiments provided enough empirical evidence to draw, under different benchmarking conditions, the following conclusions, which match most of our expectations:

1. **DMAB Operator Selection technique performs better than MAB, AP, Naive uniform, and possibly equivalently to Oracle whenever available:**
   - **True.** Indeed, DMAB is the overall winner technique in most cases. The price to pay for this gain in performance with respect to the standard MAB is the need to tune two very sensitive and problem-dependent hyper-parameters, the scaling factor $C$ and the change-detection threshold $\gamma$. Given this mentioned problem-dependency, DMAB consequently does not perform well on scenarios considering many different problems, such as BBOB.

2. **SLMAB Operator Selection technique performs equivalently or better than DMAB, while having one hyper-parameter less:**
   - **False.** This is the only notable deception with respect to the original expectation. In fact, SLMAB performed equivalently or better than DMAB only in some artificial scenarios (Section 6.3), and on the experiments within BBOB, in which both DMAB and SLMAB performed rather poorly. In the other cases using real EAs, however, SLMAB was not able to efficiently follow the dynamics of the operator qualities (see, e.g., its behavior plot on the OneMax problem, in Figure 6.12d), being outperformed even by the standard MAB in some cases. Besides, as for the other bandit-based approaches, there is still the need to tune the (still very sensitive) scaling factor $C$.

3. **Extreme Credit Assignment scheme performs better than the Instantaneous and Average ones:**
   - **True.** Except for the artificial scenarios with small $\Delta T$, in which more up-to-date
6.8 General Discussion

information is needed in order to follow the very quick abrupt changes (thus preferring the Instantaneous scheme), Extreme performs better in most cases, including the boolean benchmark and the SAT problems. On the BBOB scenario, however, only Extreme was tried to date; the evaluation of the Average and Instantaneous counterparts is left for further work.

4. Normalized versions perform equivalently or better than the Absolute versions of the Credit Assignment schemes, while being more robust with respect to their hyper-parameters:
   **True.** In terms of performance, on the analyzed scenarios, they performed equivalently in most cases, while being better in a few cases. The gain in robustness by the use of this simple normalization scheme is clearly shown in the analysis on the transformed OneMax functions, presented in Section 6.7.2. However, it can be said to be significantly outperformed, in terms of robustness, by the rank-based schemes, which should thus be preferred, as discussed in the following.

5. RMAB Operator Selection with the rank-based Credit Assignment schemes perform equivalently or better than the other AOS combinations and naive strategies, while being very robust with respect to their hyper-parameters:
   **True.** Even in the artificial and boolean benchmark problems, in which the methods were off-line tuned for each problem (thus hindering the effects of the higher robustness), the different combinations of RMAB showed to be able of following closely the performance of the best AOS combinations in most cases. And, intuitively, the more heterogeneous the scenario, *i.e.*, the more different problems need to be tackled by the same hyper-parameter configuration, the clearer the benefits brought by the higher robustness, as shown in the BBOB scenario, in which RMAB is the clear winner. It is important to note that, besides the robustness, most of the efficiency in adaptation is also due to the Credit Assignment schemes in this case, while the RMAB technique is responsible for controlling the EvE balance with respect to Operator Selection. Furthermore, between the AUC and SR, the former showed to outperform the latter in the vast majority of the cases, being equivalent otherwise. Finally, between the AUC and the FAUC, the latter should be preferred just in case the robustness with respect to monotonous transformations is needed, otherwise the AUC should be employed.

Summarizing all these empirical evidences, to date, the AUC-RMAB remains as the recommended technique in case one wants to employ the AOS paradigm on his own algorithm. The last choice that needs to be made, between the Decay and the NDCG alternatives, is not critical and depends on whether there is some available budget for the tuning of the decay factor $D$ or not, as the NDCG has shown to present similar but slightly inferior performance in most cases.
Part IV

General Conclusion
Chapter 7

Final Considerations

Evolutionary Algorithms (EAs) are stochastic algorithms that tackle search and optimization problems based on the Darwinian evolution paradigm. EAs have already shown to perform well in many different domains of application that are not tractable by standard methods, mainly because they do not make any strong assumption about the problem to be solved, and also due to the many parameters that enable the user to adapt the algorithm to the problem at hand, as discussed in Chapter 2. These many parameters, although providing the mentioned flexibility, are the main responsible factor for the fact that EAs are rarely used by researchers from other domains, as there are no general guidelines for their setting.

After the survey on parameter setting in EAs presented in Chapter 3, instead of using an off-line tuning technique (that would provide a static strategy), the behavior of the algorithm should be rather adapted while solving the problem, according to the current needs of the search with respect to the Exploration versus Exploitation (EvE) balance. Taking as an example the choice of which operator should be applied, more disruptive operators should be used to explore the search space in the initial stages of the search (or when there is the need of escaping from stagnation), while fine-tuning operators should be preferred whenever there are promising regions that need to be further verified.

The use of feedback from the search to adapt on-line the selection of the operator to be applied is commonly referred to as Adaptive Operator Selection (AOS). In this work, we proposed different contributions to AOS, which will be summarized in Section 7.1. Section 7.2 will conclude this manuscript by sketching possible directions for further work.

7.1 Summary of Contributions

In order to perform AOS, one needs to define two elements, as described in Chapter 4 and depicted in Figure 4.1. After an operator application, the Credit Assignment scheme transforms its impact on the search process into a numerical reward, which is used to maintain an up-to-date estimation of the performance of this operator. Based on these empirical estimates, the Operator Selection mechanism selects the next operator to be applied.
Chapter 7. Final Considerations

Different approaches for both Credit Assignment and Operator Selection components have been proposed in this thesis, resulting in novel AOS methods. These contributions are detailed in Chapter 5, and can be briefly summarized as follows.

The proposed Credit Assignment schemes use as a measure of impact the fitness improvement of the offspring with respect to its parent (or to the best of its parents in the case of crossover operators). The first proposal, referred to as Extreme Credit Assignment, rewards the operator with the maximum fitness improvement recently achieved by it, based on the assumption that outlier high improvements might be equally or more important than frequent but moderate ones. Although showing to be efficient, this scheme provides a very problem-dependent behavior to the AOS methods implementing it. This happens mainly due to the fact that (i) different problems have different fitness ranges, consequently reward ranges; and (ii) the magnitude of the fitness improvements tends to vary as the search advances (the closer to the optimum, usually the rarer and smaller the improvements). In the quest for a higher robustness, a simple normalization scheme was proposed, and this alleviated, but not eliminated the problem-dependency. This led us to further propose two Credit Assignment schemes based on ranks over the fitness improvements, which showed to be very robust: the first method is inspired by a Machine Learning paradigm, referred to as Area-Under-Curve (AUC), while the second simply uses the Sum-of-Ranks (SR) to evaluate the operator qualities. An extra level of robustness (with a small price to pay in terms of efficiency) was further achieved by the use of ranks over fitness values instead of ranks over fitness improvements: the resulting algorithms are totally comparison-based, i.e., invariant with respect to monotonous transformations over the original fitness function. The respective methods are referred to as Fitness-based Area-Under-Curve (FAUC) and Fitness-based Sum-of-Ranks (FSR).

The Operator Selection issue was tackled as another instance of the Exploration versus Exploitation (EvE) dilemma: the operator that is found to perform better than the others should be used as much as possible (exploitation), while other operators should also be tried from time to time (exploration), as one of them might become the new best one at a further (unknown) instant of the search. This dilemma has been intensively studied in the context of yet another Machine Learning paradigm (more specifically in Game Theory), the Multi-Armed Bandit (MAB). A first tentative application of this paradigm to the Operator Selection problem used the Upper Confidence Bound (UCB) algorithm [Auer et al., 2002], which provides asymptotic optimality guarantees with respect to total cumulative reward. But these guarantees hold only in the original and stationary context of MAB problems, while the AOS context is very dynamic: the performance of the operators continuously vary as the search goes on. As a consequence, different proposals were made in order to be able to efficiently use the MAB paradigm in the AOS context. The first proposal on this direction, referred to as Dynamic Multi-Armed Bandit (DMAB), uses a statistical test to trigger a restart of the MAB process whenever a change on the operator quality distribution is detected. This method has shown to be very efficient, but different problems have fitness landscapes with different dynamics, what makes its restarting mechanism to be also highly problem-dependent. This led us to propose the Sliding Multi-Armed Bandit (SLMAB), which accounts for the AOS dynamics by continuously adapting the weight of the received rewards, according to how frequent each operator has been applied: the less frequent,
the more outdated is its corresponding empirical estimate, consequently the higher should be the weight of the instant reward received, and vice-versa. In practice, however, this mechanism did not show to perform as good as expected. The last proposal, referred to as RMAB, is indeed the simplest one. In this method, the AOS dynamics are in fact handled on the Credit Assignment side by the rank-based schemes in a transparent way: as the ranking considers the rewards received by all operators, the application of one operator affects the perceived quality of all the others; in this way, their quality estimates are always sufficiently up-to-date by construction.

A last contribution concerns the empirical assessment of the resulting AOS methods. Firstly, some artificial scenarios were proposed to analyze their behavior on different controlled environments. And secondly, a very extensive empirical analysis of all the proposed AOS methods, also compared with some baseline methods, was performed. They were assessed within a Genetic Algorithm applied to the proposed artificial scenarios, to other boolean benchmark problems, as well as to the hard Boolean Satisfiability problems; and within a Differential Evolution algorithm applied to a comprehensive benchmark set of continuous functions.

Based on the evidences gathered from the extensive empirical analysis done, in case one wants to apply the AOS paradigm to his own algorithm/problem, the recommended AOS method as of today is the combination of the AUC Credit Assignment scheme with the RMAB Operator Selection mechanism. Compared to other adaptive and naive methods, it achieves state-of-the-art or comparable performance, while also being very robust with respect to its hyper-parameters when applied to different problems.

It is worth noting that the use of off-line tuning prior to every experiment was motivated by the intention of comparing the different AOS methods at their best. But, ideally, whenever a new problem needs to be tackled, no off-line tuning should be required – this is why so much effort was put towards achieving higher robustness. In the case of the recommended AOS method, the same hyper-parameter configuration \((C,5W50)\) was found to be the best over very different problem classes in the context of the BBOB framework, and neighboring values (considering the range of values tried) were used by the winner configurations on the OneMax and Royal Road problems \((C.1W100)\) and on most of the artificially generated problems \((C \in \{.1,.5,1\},W \in \{50,100\})\). The employment of prior off-line tuning remains thus as an optional step, as reasonable performance can be achieved by the recommended AOS method when using hyper-parameter values around the mentioned configurations. And even if one opts for off-line tuning, the given AOS method has only two hyper-parameters that will enable it to efficiently follow the best operator during the search process; while the Static off-line tuned case\(^1\) needs the setting of the application rate for each operator (\(i.e.,\) the number of parameters to be set are a multiple of the number of considered operators), besides providing, as its name says, a strategy that remains static during all the optimization run.

\(^1\)We consider here only the parameters related to the variation operators, although many other parameters could be set by the same tuning procedure, or controlled at the same time by other adaptive schemes.
7.2 Further Work

A major drawback of the final recommended AOS method is that its remarkable robustness and its state-of-the-art performance remain limited, mostly, to unimodal problems. In order to efficiently tackle multi-modal problems, the impact of the operator application with relation to the population diversity should also be considered somehow, while only the fitness is being currently regarded. This explains why none of the methods were able to perform well on the multi-modal functions of the BBOB test-bed. Some preliminary work has been done on this direction, using the Compass method from Université d’Angers as Credit Assignment, assessed within a GA in the context of Boolean Satisfiability (SAT) problems (see Section 6.5). But in this work, only the very sensitive and problem-dependent DMAB technique was used for Operator Selection. Further work should concern the preservation of the achieved level of robustness and efficiency in the framework of the Pareto Dominance-based Credit Assignment scheme proposed in [Maturana et al., 2010b], which is a follow-up of the Compass method.

An alternative approach for tackling multi-modal problems can be found in the state-of-the-art optimizer IPOP-CMA-ES [Auger and Hansen, 2005], which was used as a baseline for comparison on the experiments within BBOB. Instead of modifying its adaptive mechanism, it uses a deterministic parameter control of the population size: after some number of generations, if no improvement has been achieved, the size of the population is doubled and the search is restarted from scratch. Bigger the population, higher are the chances of finding the global optimum, as it enables a better parallel exploration of the multiple peaks of the fitness landscape. This approach should be tried in the near future, by incorporating the restarting/population size control mechanism into the underlying algorithms that were tried with the AOS schemes, namely, GAs and DE.

Coming back to the work on SAT problems, the AOS methods were used to select between rather naive operators, thus not achieving competitive performance. In order to possibly take part in SAT races, further work should concern, thus, the autonomous selection between state-of-the-art operators for SAT. Indeed, we are currently working in collaboration with University of British Columbia on this topic; but, as of today, there are no conclusive results yet.

Along the same line, in the work within the DE algorithm applied to continuous problems, the AOS methods were combined with the very standard version of DE. Several more efficient DE variants exist nowadays, such as the JADE algorithm [Zhang and Sanderson, 2009], which, besides using improved mutation strategies, also employs the on-line adaptation of some of the DE parameters, namely, the mutation scaling factor $F$ and the crossover rate $CR$. As a continuation of the collaboration work with the China University of Geosciences, the PM-AdapSS-DE method (Section 6.6.3) was tried within JADE, achieving significantly better results than when combined with the standard algorithm [Gong et al., 2011]. A natural next step in this case would be to try our rank-based AOS methods with JADE, in order to possibly achieve more competitive results when compared to state-of-the-art optimizers.

Another path for further work, that is also being currently explored in the scope of a collaboration, this time with the City University of Hong Kong, involves the DE algorithm
7.2 Further Work

again, but applied to multi-objective problems. This work is still in a very preliminary stage, as acknowledged in [Li et al., 2011]. The main difficulty for the time being is (as in the standard non-adaptive framework) to define how to efficiently evaluate the quality of the solutions (and consequently the impact of the operator application) with respect to the different objectives.

As can be seen from all these on-going collaborations, the AOS paradigm is indeed very useful and general enough to be applied to many different contexts. But up to now, we have considered their application only to the selection of operators within EAs. The developed AOS methods should also be further assessed within different kinds of meta-heuristics and stochastic algorithms; and at a higher level of abstraction, selecting between heuristics instead of operators, what is commonly referred to as Hyper-Heuristics. For instance, the upcoming “International Cross-Domain Heuristic Search Challenge” seems to be an interesting experimental framework to evaluate the AOS mechanisms at the level of Hyper-Heuristics: there is a set of low-level heuristics defined for each problem domain (including MAX-SAT, bin-packing, flow-shop and scheduling), and the task consists into automatically managing their use, in an efficient way, over all the problem domains.

A different issue concerns the scalability of the AOS methods with respect to the number of operators being controlled. In the experimental settings used in this work, there were at most six operators. Ideally, the same level of performance should be attained no matter the number of operators. In practice, the more operators are considered, the more exploration is needed, consequently degrading the performance of the algorithm.

A recently proposed approach, referred to as Adaptive Operator Management (AOM) [Maturana et al., 2010b], tackles this problem by adapting on-line the list of variation operators that are available to the AOS method: operators found to be inefficient are momentarily disabled, while others are included in the active operator set. This seems to be a prominent and complementary research direction, which could be further investigated.

On the application side, it is true that the gain brought by the use of AOS, in terms of number of fitness function evaluations, might become more “valuable” in real-world problems, in which the fitness evaluation is usually very expensive. During the next year, we intend to apply the developed adaptive mechanisms to problems linked to sustainable development. More specifically, we are currently starting to work on the optimization of the design and materials used in the construction of buildings, aiming at a better energy efficiency. This is a multi-objective problem by definition, with many degrees of freedom resulting in a huge search space, and the evaluation of a solution requires the execution of a complete energy consumption simulation, which might take up to several minutes.

Finally, many research colleagues have been recently demonstrating interest on the code of the proposed AOS methods. These requests are being made mainly (i) to employ the AOS paradigm on their own optimization algorithm/problem, and/or (ii) to use our AOS methods as baseline to compare with their own methods (e.g., [Verel et al., 2010]). In order to “complete” the scientific contributions proposed in this thesis, we intend to prepare and make freely available a well-documented package containing the source code of the proposed methods.

http://www.asap.cs.nott.ac.uk/chesc2011
Bibliography

[Abdullah et al., 2007] cited page(s) 25

[Anderson et al., 2000] cited page(s) 15

[Angeline, 1995] cited page(s) 40

[Angelov et al., 2003] cited page(s) 26

[Arabas et al., 1994] cited page(s) 32

[Arbib, 2002] cited page(s) 12

[Auer et al., 2002] cited page(s) xi, xi, 5, 52, 68, 68, 81, 81, 83, 188

[Auger and Hansen, 2005] cited page(s) 95, 161, 168, 190

193
[Auger and Teytaud, 2010] cited page(s) 30

[Bäck et al., 2000] cited page(s) 32

[Bäck, 1992] cited page(s) 39

[Balaprakash et al., 2007] cited page(s) 36, 36, 37, 103

[Barbosa and Sá, 2000] cited page(s) 48, 49, 50, 52, 57

[Bartz-Beielstein et al., 2005] cited page(s) 37, 103

[Battiti et al., 2008] cited page(s) 39

[Bentley and Corne, 2002] cited page(s) 26

[Beyer, 1995] cited page(s) 41

[Bibai et al., 2010] cited page(s) 25


[Clune et al., 2005] cited page(s) 37

[Collet and Schoenauer, 2003] cited page(s) vii, 3

[Conover, 1999] cited page(s) 103

[Cook, 1971] cited page(s) 15, 151

[Da Costa and Schoenauer, 2009] cited page(s) vii, 3, 20

[Da Costa et al., 2008] cited page(s) xi, xii, 5, 50, 68, 69, 82, 85, 85, 91, 105, 171

[Das et al., 2009] cited page(s) 25

[Davis, 1989] cited page(s) 48, 48, 49, 50, 52, 56, 56

[De Jong and Spears, 1990] cited page(s) 30, 30, 30, 30

[De Jong, 2006] cited page(s) 3, 12, 18, 21


[Fogel et al., 1966] cited page(s) 12, 22

[Fogel, 1995] cited page(s) 12, 22

[Fogel, 1998] cited page(s) 12

[Gagliolo and Schmidhuber, 2008] cited page(s) xi, 68

[Garnier and Kallel, 2000] cited page(s) 140, 140, 141, 142

[Gelly et al., 2007] cited page(s) 6

[Giger et al., 2007] cited page(s) 48, 49, 49, 59

[Goldberg et al., 1991] cited page(s) 34

[Goldberg et al., 1992] cited page(s) 32

[Goldberg et al., 1995] cited page(s) 20

[Goldberg, 1989] cited page(s) 12, 22

[Goldberg, 1990] cited page(s) 52, 52
[Gong et al., 2010] cited page(s) 62, 158, 159, 160, 168

[Gong et al., 2011] cited page(s) 62, 190

[Grefenstette, 1986] cited page(s) 30, 30, 30, 30

[Hansen and Ostermeier, 2001] cited page(s) 6, 21, 25, 33, 39, 40

[Hansen et al., 2009a] cited page(s) 159, 160, 163

[Hansen et al., 2009b] cited page(s) 159

[Hansen et al., 2009c] cited page(s) 25

[Hansen et al., 2010a] cited page(s) 158

[Hansen et al., 2010b] cited page(s) 158

[Hansen, 2008] cited page(s) 34
BIBLIOGRAPHY

[Hansen, 2009a] cited page(s) 161

[Hansen, 2009b] cited page(s) 25

[Hartland et al., 2007] cited page(s) 85

[Hatta et al., 1997] cited page(s) 57

[Hatta et al., 2001] cited page(s) 57, 57

[Herrera and Lozano, 1998] cited page(s) 32

[Herrera and Lozano, 2001] cited page(s) 60

[Hesser and Männer, 1990] cited page(s) 39

[Hinkley, 1970] cited page(s) xi, 5, 68

[Ho et al., 1999] cited page(s) 48, 49, 58, 60
[Holland, 1975] cited page(s) 12, 22

[Holland, 1993] cited page(s) 145, 146

[Hoos and Stützle, 2000] cited page(s) 152

[Horn et al., 1994] cited page(s) 137

[Horn, 1997] cited page(s) 27

[Hutter and Hamadi, 2005] cited page(s) 42

[Hutter et al., 2006] cited page(s) vii, 4

[Hutter et al., 2009] cited page(s) 37, 37

[Ibanez et al., 2009] cited page(s) 25

[Igel et al., 2007] cited page(s) 41
[Jansen et al., 2005] cited page(s) 32

[Järvelin and Kekäläinen, 2000] cited page(s) 74

[Jones, 1994] cited page(s) 145

[Julstrom, 1995] cited page(s) 48, 49, 49, 50, 52, 56, 56

[Julstrom, 1997] cited page(s) 48, 49, 49, 50, 52, 56, 56, 57

[Kamalian et al., 2005] cited page(s) 26

[Keijzer et al., 2002] cited page(s) vii, 3, 20, 94

[Klir and Yuan, 1995] cited page(s) 12

[Koza et al., 2000] cited page(s) 26

[Koza et al., 2003] cited page(s) 16
[Koza, 1992] cited page(s) 12, 22

[Koza, 1994] cited page(s) 12, 22

[Koza, 2010] cited page(s) 26

[Kramer, 2010] cited page(s) 39

[Krasnogor, 2002] cited page(s) 18

[Laarhoven and Aarts, 1987] cited page(s) 14

[Lai and Robbins, 1985] cited page(s) xi, 5, 68, 81, 81

[Lardeux et al., 2006] cited page(s) 15, 18

[Lee and Takagi, 1993] cited page(s) 60

[Li et al., 2011] cited page(s) 190

[Lobo and Goldberg, 1997] cited page(s) 48, 49, 52, 57

204
BIBLIOGRAPHY

[Lobo et al., 2007] cited page(s) vii, 4, 31, 196, 197

[Luchian and Gheorghies, 2003] cited page(s) 49, 49, 52, 58

[Luke and Panait, 2006] cited page(s) 22

[Martikainen and Ovaska, 2006] cited page(s) 15

[Maturana and Saubion, 2007a] cited page(s) 60

[Maturana and Saubion, 2007b] cited page(s) 60

[Maturana and Saubion, 2008a] cited page(s) xiii, 48, 49, 50, 51, 52, 58, 58, 59, 63, 150, 151, 152

[Maturana and Saubion, 2008b] cited page(s) 60

[Maturana et al., 2009a] cited page(s) xi, 4, 7, 51, 63, 69, 85, 150

[Maturana et al., 2009b] cited page(s) 59
[Maturana et al., 2010a] cited page(s) 7, 51, 150

[Maturana et al., 2010b] cited page(s) 49, 49, 59, 95, 157, 189, 191

[Maturana, 2009] cited page(s) 41

[Merz and Freisleben, 1997] cited page(s) 15, 25

[Michalewicz, 1996] cited page(s) 3

[Mitchell et al., 1992] cited page(s) 145

[Mitchell, 1998] cited page(s) 23

[Moore, 1991] cited page(s) 31

[Mueller-Gritschneder et al., 2009] cited page(s) 15

[Nannen and Eiben, 2007] cited page(s) 38, 171

[Niehaus and Banzhaf, 2001] cited page(s) 49, 49, 52, 61
BIBLIOGRAPHY

[Obradovic and Srikumar, 2000] cited page(s) 16

[Ong and Keane, 2004] cited page(s) 162

[Page, 1954] cited page(s) 84, 85, 85

[Pitman and King, 2009] cited page(s) 26

[Porumbel et al., 2010] cited page(s) 25

[Price et al., 2005] cited page(s) 12, 23, 24, 24, 25

[Qin et al., 2009] cited page(s) 61, 170

[Quick et al., 1996] cited page(s) 145

[Quiroz et al., 2007] cited page(s) 15, 26

[Rechenberg, 1972] cited page(s) 12, 21

[Romero and Machado, 2007] cited page(s) 26


[Ursem, 2002] cited page(s) 59

[Vajda et al., 2008] cited page(s) 32, 32

[Verel et al., 2010] cited page(s) 94, 191

[Wei et al., 2008] cited page(s) 63, 158

[Whitacre et al., 2006] cited page(s) 6, 49, 50, 52, 61, 71

[Whitley and Watson, 2005] cited page(s) 30

[Whitley, 1994] cited page(s) 23

[Wolpert and Macready, 1997] cited page(s) 30, 41

[Wong et al., 2003] cited page(s) 48, 49, 52, 58

[Yoemans et al., 2003] cited page(s) 26

210
[Yu et al., 2008] cited page(s) 3, 25

[Yuan and Gallagher, 2004] cited page(s) 36

[Yuan and Gallagher, 2007] cited page(s) 37

[Zaharie, 2009] cited page(s) 25

[Zhang and Sanderson, 2009] cited page(s) 190