



DEPARTMENT DE SCIENCES ET INGÉNIERIE

---

UE PROJET

# Comparaison entre Augmented Random Search (ARS) et Cross-Entropy Method (CEM)

---

*Auteurs*

Joe CHAMOUN 21312860

Jawher JRIBY 3702809

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Méthodologie</b>	<b>3</b>
2.1	Algorithme ARS . . . . .	3
2.2	Algorithme CEM . . . . .	3
2.3	Configuration Expérimentale . . . . .	4
<b>3</b>	<b>Résultats</b>	<b>4</b>
3.1	Résultats de ARS . . . . .	4
3.2	Résultats du CEM . . . . .	5
<b>4</b>	<b>Comparaison des Performances</b>	<b>5</b>
4.1	Analyse des Courbes d'Apprentissage . . . . .	5
4.2	Comparaison des Récompenses Cumulatives . . . . .	5
4.3	Évaluation de la Stabilité . . . . .	6
4.4	Analyse des Stratégies d'Exploration . . . . .	6
4.5	Adaptation des Politiques . . . . .	6
<b>5</b>	<b>Discussion</b>	<b>6</b>
5.1	Analyse des Performances . . . . .	6
5.2	Stabilité et Convergence . . . . .	6
5.3	Complexité Computationnelle . . . . .	7
<b>6</b>	<b>Conclusion</b>	<b>7</b>
<b>7</b>	<b>Travaux Futurs</b>	<b>7</b>

# 1 Introduction

L'apprentissage par renforcement (RL) est une branche de l'intelligence artificielle qui permet aux agents d'apprendre à prendre des décisions optimales par essais et erreurs dans des environnements complexes. Deux outils importants dans ce domaine sont Augmented Random Search (ARS) et Benchmarking and Baselines for Reinforcement Learning (BBRL). ARS est un algorithme d'apprentissage par renforcement connu pour sa simplicité et son efficacité, utilisant des perturbations aléatoires pour optimiser les politiques sans nécessiter de gradients. D'autre part, BBRL est une bibliothèque conçue pour faciliter la recherche en RL en fournissant des environnements standardisés et des implémentations de base d'algorithmes, permettant aux chercheurs de comparer leurs travaux de manière cohérente et reproductible. Dans le cadre de notre projet, nous présentons une analyse comparative entre deux algorithmes d'optimisation et d'apprentissage par renforcement : l'Augmented Random Search (ARS) en BBRL dans ses versions v1 et v2, ainsi que la méthode Cross-Entropy (CEM). Nous avons développé les implémentations des versions v1 et v2 de l'ARS et avons utilisé une implémentation existante pour le CEM. Ces algorithmes ont été testés sur des environnements de simulation continue et discrète. L'objectif principal est de comprendre les forces et les faiblesses de chaque algorithme en termes de performances, de stabilité et de complexité computationnelle. L'ARS est connu pour sa simplicité et son efficacité, tandis que CEM est apprécié pour sa capacité à explorer l'espace des solutions de manière approfondie. Nous explorerons les différences de performance et de stabilité entre ces algorithmes.

## 2 Méthodologie

### 2.1 Algorithme ARS

ARS est une méthode d'optimisation simple et efficace pour l'apprentissage par renforcement. Elle consiste à perturber les paramètres du modèle avec des vecteurs gaussiens, à évaluer les performances de ces perturbations et à ajuster les paramètres en fonction des résultats obtenus. Nous avons implémenté les deux versions principales de ARS :

- **ARS v1** : Cette version de ARS met à jour les paramètres de la politique en se basant sur les récompenses moyennes obtenues à partir des perturbations positives et négatives. Sa simplicité réside dans les mises à jour directes de la politique à l'aide de calculs statistiques de base.
- **ARS v2** : En s'appuyant sur l'ARS v1, l'ARS v2 introduit une normalisation des récompenses par l'écart type, améliorant ainsi la stabilité des mises à jour des paramètres. En tenant compte de la variabilité des récompenses, l'ARS v2 facilite des ajustements de politique plus robustes.

### 2.2 Algorithme CEM

La méthode Cross-Entropy (CEM) est un algorithme d'optimisation stochastique qui se concentre sur la mise à jour des paramètres d'un modèle en se basant sur les meilleurs échantillons obtenus. CEM fonctionne en générant une population de solutions, en évaluant leurs performances, puis en mettant à jour les paramètres en se basant sur les meilleurs échantillons. Les étapes principales de l'algorithme CEM sont :

- Génération d'une population initiale de solutions.
- Évaluation des performances de chaque solution.
- Sélection des meilleures solutions (élites) en fonction de leurs performances.
- Mise à jour des paramètres en utilisant les élites.
- Répétition du processus jusqu'à convergence.

## 2.3 Configuration Expérimentale

- **Environnements d'Évaluation** : Nous avons évalué ARS et CEM en utilisant deux types d'environnements distincts :
  - **Environnement Continu (CartPoleContinuous-v1)** : Dans ce cadre, l'agent doit équilibrer un poteau sur un chariot en utilisant des actions continues, posant ainsi un défi de contrôle continu.
  - **Environnement Discret (CartPole-v1)** : Ici, l'agent équilibre un poteau avec des actions discrètes, mettant en évidence les stratégies de contrôle discret.

## 3 Résultats

### 3.1 Résultats de ARS

Les performances de ARS ont été évaluées en termes de récompenses cumulées obtenues par l'agent au fil des épisodes. Les figures suivantes montrent les courbes d'apprentissage pour les deux versions de ARS pour les environnements continus et discrets.

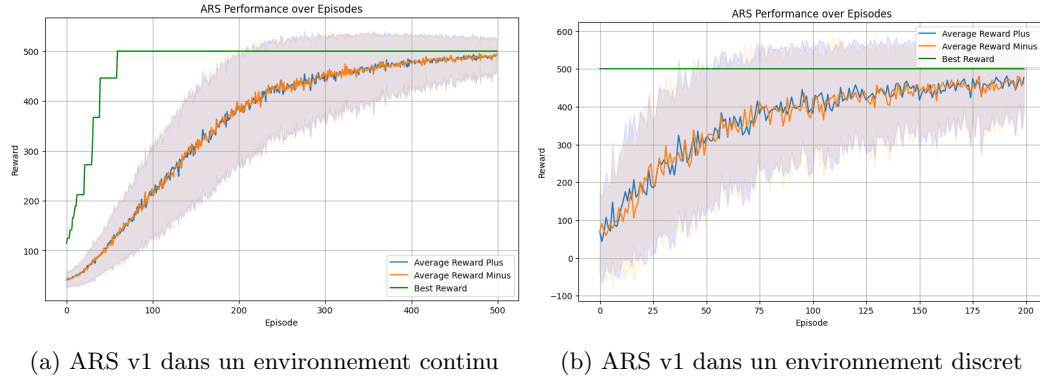


Figure 1: Performances de l'ARS v1

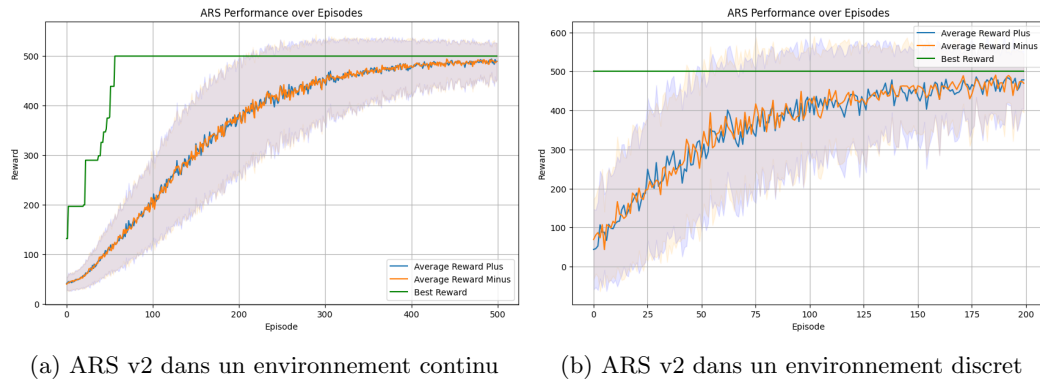


Figure 2: Performances de l'ARS v2

Les résultats montrent que la version 2 de ARS présente une trajectoire plus cohérente et se stabilise à une récompense plus élevée plus rapidement. En effet ARS v2 normalise les récompenses, ce qui aide à gérer différentes échelles de récompenses et rend le processus d'apprentissage plus stable. Cela conduit à une amélioration constante et à une convergence plus rapide vers des politiques à haute récompense. En outre, en normalisant les observations pour avoir une moyenne nulle et une

variance unitaire, ARS v2 garantit que le processus d'apprentissage n'est pas affecté par les échelles variées des caractéristiques d'entrée, conduisant à un apprentissage plus robuste.

## 3.2 Résultats du CEM

Les performances du CEM ont également été évaluées de manière similaire. Les figures suivantes montrent les courbes d'apprentissage du CEM dans les environnements continus et discrets.

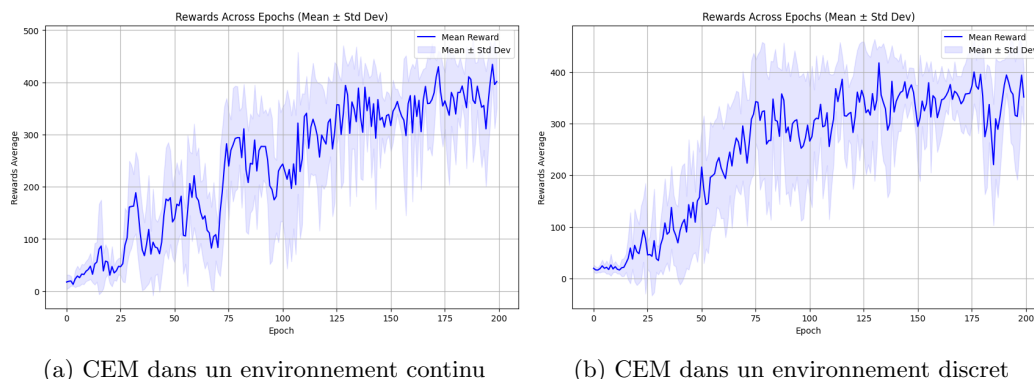


Figure 3: Performances du CEM

Les résultats montrent que le CEM peut atteindre des récompenses élevées, mais il peut également souffrir d'une variance élevée dans les performances, surtout dans les environnements discrets.

## 4 Comparaison des Performances

### 4.1 Analyse des Courbes d'Apprentissage

Les courbes d'apprentissage fournissent un aperçu essentiel de la progression des algorithmes ARS et CEM au fil des épisodes. Elles illustrent comment les performances évoluent pendant l'apprentissage, ce qui permet de comparer la convergence, la stabilité et l'efficacité de ces deux approches.

Dans les environnements discrets, nous observons que ARS v2 converge plus rapidement que ARS v1. Cette amélioration de la convergence est attribuable à la normalisation par l'écart type introduite dans ARS v2, qui permet des ajustements plus précis des paramètres de la politique. Par conséquent, ARS v2 atteint généralement des récompenses plus élevées à un rythme plus rapide que ARS v1.

D'autre part, CEM présente une convergence plus lente mais des performances globales comparables dans les environnements discrets. Cette approche met l'accent sur la sélection des élites pour mettre à jour les paramètres, ce qui peut nécessiter plus d'itérations pour converger vers des politiques efficaces. Cependant, CEM maintient souvent une stabilité accrue tout au long du processus d'apprentissage.

### 4.2 Comparaison des Récompenses Cumulatives

L'évaluation des récompenses cumulatives moyennes et maximales obtenues par ARS et CEM offre des perspectives détaillées sur leur efficacité globale. Sur les environnements continus, ARS v2 affiche souvent des performances supérieures en termes de récompenses cumulatives par rapport à CEM.

Cette différence peut s'expliquer par la nature de l'approche d'exploration de ARS v2, qui utilise des perturbations normalisées pour explorer efficacement l'espace des politiques. En revanche, bien que le CEM puisse atteindre des récompenses compétitives, il tend à présenter une variance plus faible des performances, ce qui peut être bénéfique dans des applications où la stabilité est primordiale.

### 4.3 Évaluation de la Stabilité

ARS v2 se distingue par sa stabilité accrue par rapport à ARS v1, grâce à la normalisation par l'écart type qui régule les variations des récompenses entre les épisodes. Cette stabilité améliorée favorise une convergence plus fiable et régulière vers des politiques optimales. En revanche, bien que CEM soit plus stable que ARS v1, elle peut encore présenter des fluctuations dans les performances en raison de sa stratégie de sélection d'élites.

### 4.4 Analyse des Stratégies d'Exploration

Les stratégies d'exploration utilisées par ARS et CEM influencent directement leur capacité à découvrir des politiques efficaces dans des espaces d'actions complexes.

ARS utilise des perturbations gaussiennes pour explorer l'espace des paramètres de manière aléatoire mais contrôlée. Cette approche favorise une exploration diversifiée, ce qui peut être bénéfique pour éviter les minimas locaux et découvrir des politiques prometteuses dans des environnements dynamiques.

En revanche, CEM se concentre sur une élite de solutions pour ajuster les paramètres de la politique. Cette approche privilégie le choix sélectif des échantillons les plus performants, ce qui peut conduire à des stratégies plus ciblées mais potentiellement moins diversifiées.

### 4.5 Adaptation des Politiques

ARS ajuste les paramètres de la politique en fonction des performances relatives des perturbations, ce qui encourage une exploration adaptative et itérative de l'espace des politiques. Cette approche favorise la diversité des solutions et permet de découvrir des stratégies efficaces dans des environnements variés.

En contraste, CEM se concentre sur des ajustements plus ciblés basés sur des élites sélectionnées. Bien que cette approche puisse être efficace pour affiner les paramètres vers des solutions optimales, elle peut également être limitée par la diversité des politiques explorées, ce qui peut être critique dans des environnements dynamiques ou non déterministes.

## 5 Discussion

### 5.1 Analyse des Performances

L'analyse des courbes d'apprentissage révèle plusieurs points clés concernant les performances des deux algorithmes :

- **ARS v1 vs ARS v2** : La version 2 de ARS montre une meilleure stabilité et des récompenses plus élevées grâce à la normalisation par l'écart type, qui permet une mise à jour plus robuste des paramètres.
- **ARS vs CEM** : Bien que les deux algorithmes puissent atteindre des performances élevées, ARS tend à être plus stable, tandis que CEM peut parfois montrer des fluctuations importantes dans les récompenses obtenues.

### 5.2 Stabilité et Convergence

ARS v2, grâce à la normalisation par l'écart type, offre une meilleure stabilité par rapport à ARS v1 et CEM. En revanche, CEM, bien que potentiellement capable de trouver des solutions optimales, peut nécessiter plus d'itérations pour converger de manière stable, surtout dans des environnements complexes.

### 5.3 Complexité Computationnelle

En termes de complexité computationnelle, ARS est généralement plus léger et plus rapide à exécuter que CEM, car il repose sur des mises à jour simples basées sur des perturbations aléatoires. CEM, en revanche, nécessite des calculs supplémentaires pour maintenir et mettre à jour la matrice de covariance, ce qui peut augmenter le temps de calcul.

## 6 Conclusion

En résumé, l'analyse comparative entre l'Augmented Random Search (ARS) et la méthode Cross-Entropy (CEM) dans des environnements discrets et continus met en évidence à la fois leurs différences et leurs similitudes en termes de performances, de stabilité et de complexité computationnelle.

ARS se distingue par sa simplicité et son efficacité d'exécution, tandis que CEM offre une approche plus robuste pour explorer efficacement les espaces de politiques. Le choix entre ces deux méthodes dépend des exigences spécifiques de l'application, en prenant en compte des facteurs tels que la stabilité requise, la complexité computationnelle et la nature du problème à résoudre.

Ce rapport a présenté une comparaison entre ARS et CEM dans des environnements continus et discrets. Les résultats montrent que ARS, notamment sa version 2, présente une stabilité supérieure et une convergence plus rapide dans la plupart des cas. Cependant, bien qu'avec une variance plus importante, CEM peut atteindre des niveaux de performance élevés. En fin de compte, le choix de la méthode appropriée dépend des besoins et des contraintes spécifiques de l'application concernée.

## 7 Travaux Futurs

Pour améliorer cette étude, plusieurs pistes peuvent être explorées :

- **Optimisation des Hyperparamètres** : Tester différents ensembles d'hyperparamètres pour voir comment ils affectent les performances des algorithmes.
- **Environnements Plus Complexes** : Évaluer les algorithmes dans des environnements plus complexes et plus réalistes pour voir comment ils se comportent dans des situations plus difficiles.
- **Combinaison d'Algorithmes** : Explorer des approches hybrides qui combinent les forces de ARS et CEM pour améliorer les performances globales.

## References

1. H. Mania, A. Guy, and B. Recht, "Simple random search provides a competitive approach to reinforcement learning," 2018. *arXiv:1803.07055*. Available online: <https://arxiv.org/html/1803.07055>
2. A. Guy, H. Mania, and B. Recht, "Augmented Random Search (ARS)," 2018. Available online: <https://github.com/modestyachts/ARS>
3. A. Tiwari and S. V. Nadimpalli, "Augmented Random Search for Quadcopter Control: An alternative to Reinforcement Learning," 2019. Available online: [https://www.researchgate.net/publication/337671730\\_Augmented\\_Random\\_Search\\_for\\_Quadcopter\\_Control\\_An\\_alternative\\_to\\_Reinforcement\\_Learning](https://www.researchgate.net/publication/337671730_Augmented_Random_Search_for_Quadcopter_Control_An_alternative_to_Reinforcement_Learning)