Model-based Optimization with Normalizing Flows

Sara Karami

IFT Seminars

April 2024

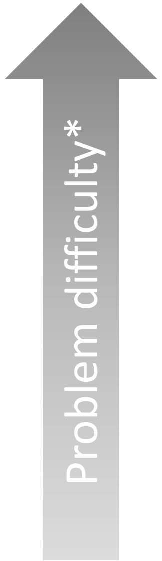
Outline

- Introduction and Background
- Proposed approach
- Experiments and Results
- Conclusion and Future work

Optimization

General form: $\min_{x}\{f(x) : x \in \Omega\}$

Selecting method according to $f(x)$



- Blackbox Optimization (BBO)
- Derivative Free Optimization (DFO)
- Gradient Descent
- Analytical Solvers

*Problem difficulty: Less information / assumptions available

Optimization

General form: $\min_{x} \{ f(x) : x \in \Omega \}$

Selecting method according to $f(x)$

Problem difficulty*

- Blackbox Optimization (BBO) $\rightarrow$ Only gives output values when gets some input queries
- Derivative Free Optimization (DFO) $\rightarrow$ No direct access to derivatives but they exist and there is a level of **smoothness**
- Gradient Descent $\rightarrow$ No regular form but there is a formulation to calculate the derivatives
- Analytical Solvers $\rightarrow$ Formulating the problem and its constraints in specific forms

*Problem difficulty: Less information / assumptions available

Optimization

General form: $\min_{x}\{f(x) : x \in \Omega\}$

Selecting method according to $f(x)$

Problem difficulty

- Blackbox Optimization (BBO)
- Derivative Free Optimization (DFO)
- Gradient Descent
- Analytical Solvers

→ Only if derivatives do not exist or are too costly to compute

Black Box Optimization

Function characteristics:

- Complexity: non-smooth, discontinuous, highly multimodal, noisy
- Dimensionality: large search space
- Separability: dependence between objective variables

$$\min_x \{ f(x) : x \in \Omega \}$$

Schematic of a Black Box [1]

[1] C. Audet, W. Hare, Derivative-Free and Blackbox Optimization, 1st ed. 2017.

Black Box Optimization

Function characteristics:

- Complexity: non-smooth, discontinuous, highly multimodal, noisy
- Dimensionality: large search space
- Separability: dependence between objective variables

$$\min_{x}\{f(x) : x \in \Omega\}$$

Schematic of a Black Box [1]

Real-world examples:

- Computer simulation
- Design of experiments

[1] C. Audet, W. Hare, Derivative-Free and Blackbox Optimization, 1st ed. 2017.

Black Box Optimization

Function characteristics:

- Complexity: non-smooth, discontinuous, highly multimodal, noisy
- Dimensionality: large search space
- Separability: dependence between objective variables

$$\min_{x}\{f(x) : x \in \Omega\}$$

Schematic of a Black Box [1]

Real-world examples:

- Computer simulation
- Design of experiments

Goal: Find a good enough solution with **minimum number of sample evaluations**

[1] C. Audet, W. Hare, Derivative-Free and Blackbox Optimization, 1st ed. 2017.

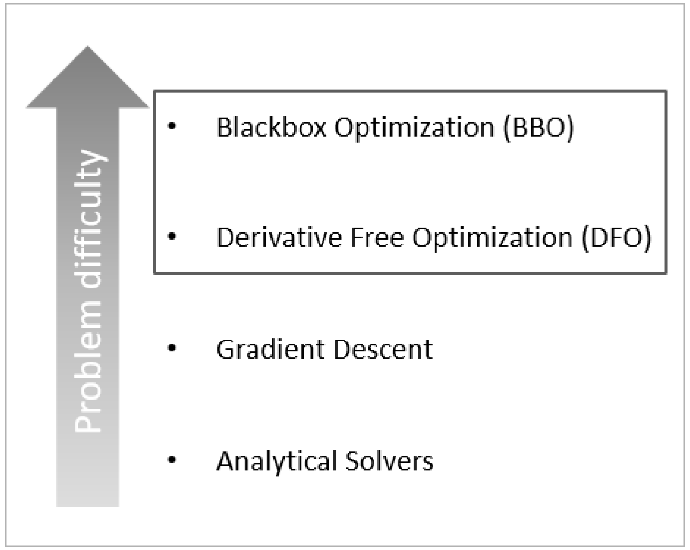
BBO vs DFO

Black-box optimization

- Typically, no assumptions of any form of continuity, differentiability or smoothness on the function
- More on the **heuristic side** without mathematical support
- Main approaches are **Evolutionary Algorithms** (EA) and **Randomized based** methods

Derivative free optimization

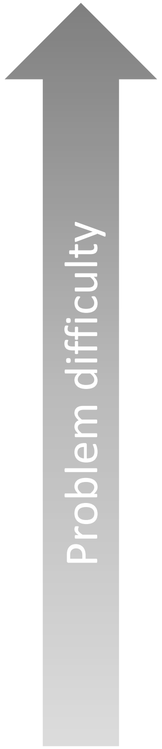
- More mathematically supported: prove of **convergence** and/or **stopping criterion**
- More on the **deterministic** side
- Direct search and surrogate-based



BBO Methods

Deciding BBO method

- Complexity and characteristics of $f(x)$
- Cost of evaluation or evaluation budget



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Bayesian Optimization
- Simulated annealing
- Hill climbing

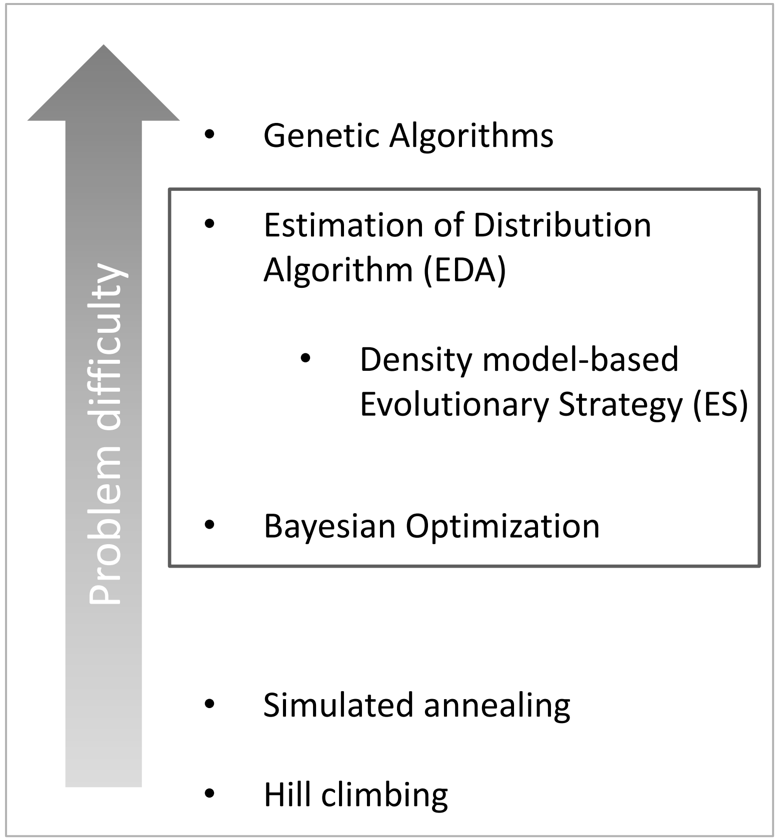
BBO Methods

Deciding BBO method

- Complexity and characteristics of $f(x)$
- Cost of evaluation or evaluation budget

Model-based BBO

- View optimization as a series of incremental updates of a model
    - Gaussian process
    - Probabilistic distribution
- Initialize the model
- Sample candidate solutions
- Evaluate the samples and update the model



BBO Methods

Deciding BBO method

- Complexity and characteristics of $f(x)$
- Cost of evaluation or evaluation budget

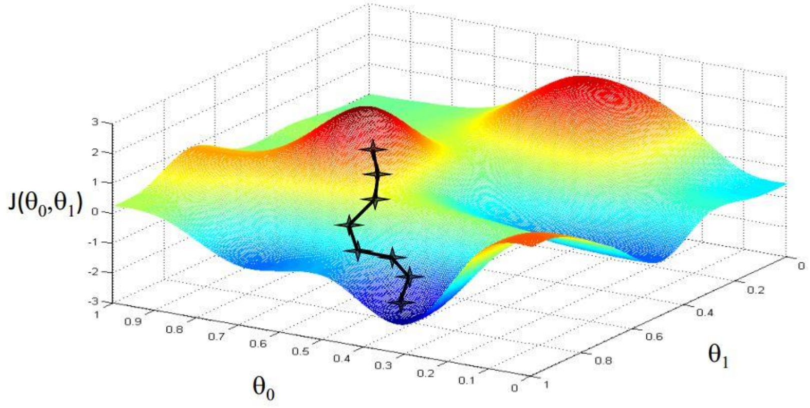
Model-based BBO

- View optimization as a series of incremental updates of a model
    - Gaussian process
    - Probabilistic distribution
- Initialize the model
- Sample candidate solutions
- Evaluate the samples and update the model

Problem difficulty

- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- Bayesian Optimization
- Simulated annealing
- Hill climbing

Hill Climbing

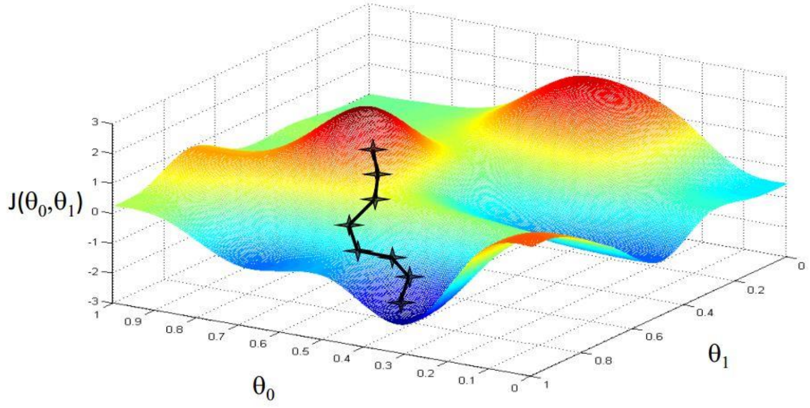


- Starts from a random point
- Randomly selects a neighbor with **better** value

Problem difficulty

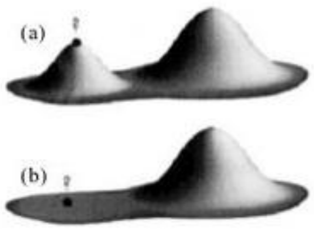
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- **Hill climbing**

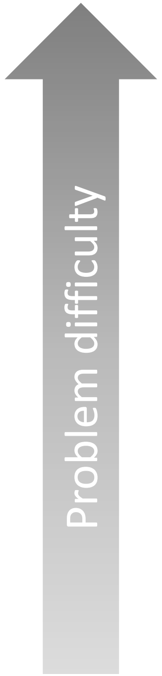
Hill Climbing



- Starts from a random point
- Randomly selects a neighbor with **better** value

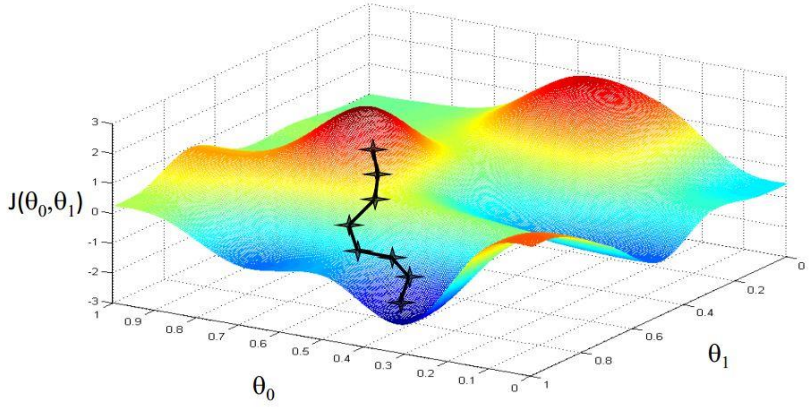
- Highly prone to get stuck in local optima or plateaus!



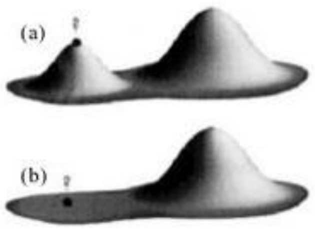


- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- **Hill climbing**

Hill Climbing

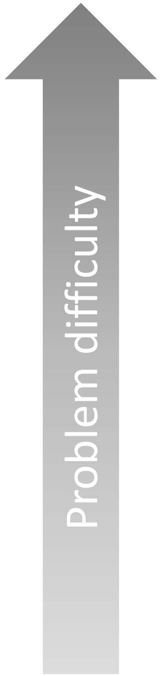


- Starts from a random point
- Randomly selects a neighbor with **better** value



- Highly prone to get stuck in local optima or plateaus!

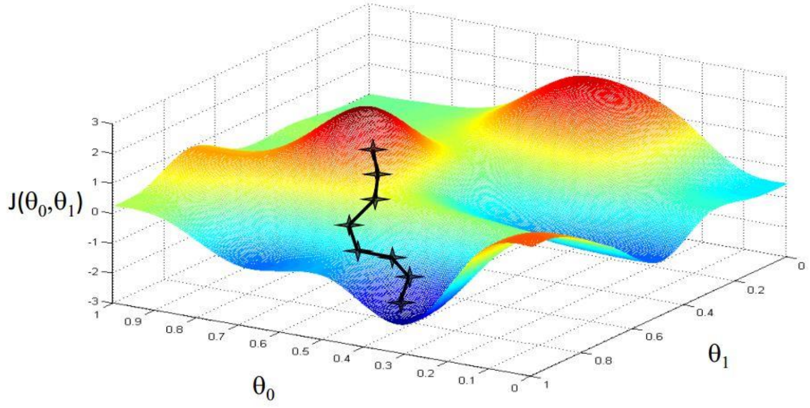
- Hill climbing vs Gradient descent?



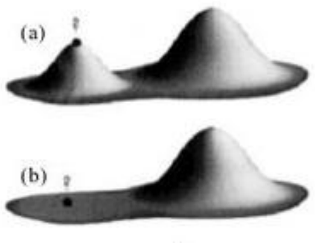
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- **Hill climbing**

https://umu.to/blog/2018/06/29/hill-climbing-irl
https://www.slideshare.net/JismyKJose/heuristc-search-70129411

Hill Climbing

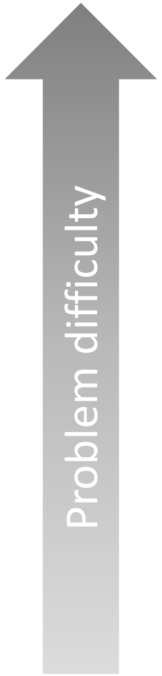


- Starts from a random point
- Randomly selects a neighbor with **better** value



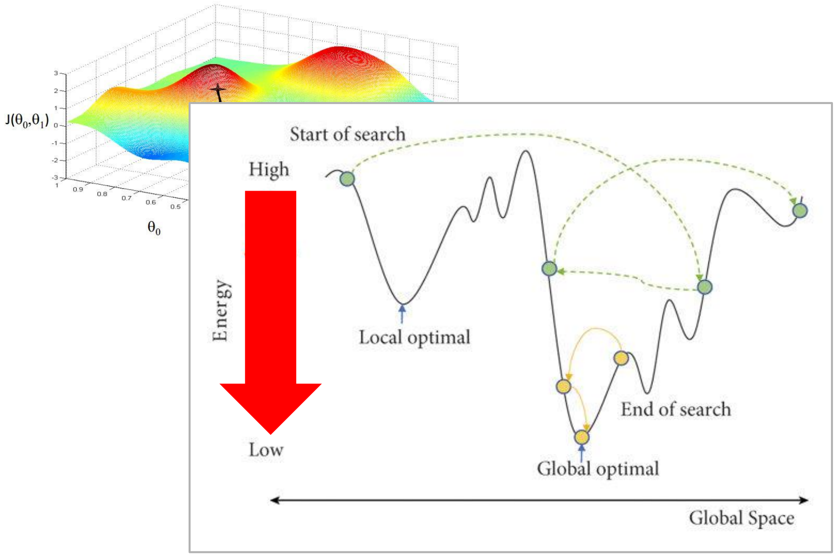
- Highly prone to get stuck in local optima or plateaus!

- Hill climbing vs Gradient descent? **Greedy** choice the next point based on **immediate local information** regardless of gradient amount or direction

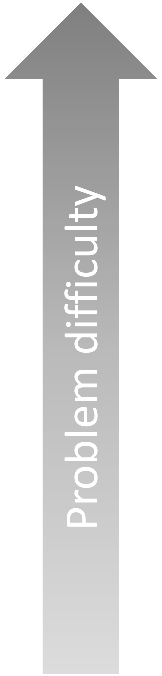


- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- **Hill climbing**

Simulated Annealing

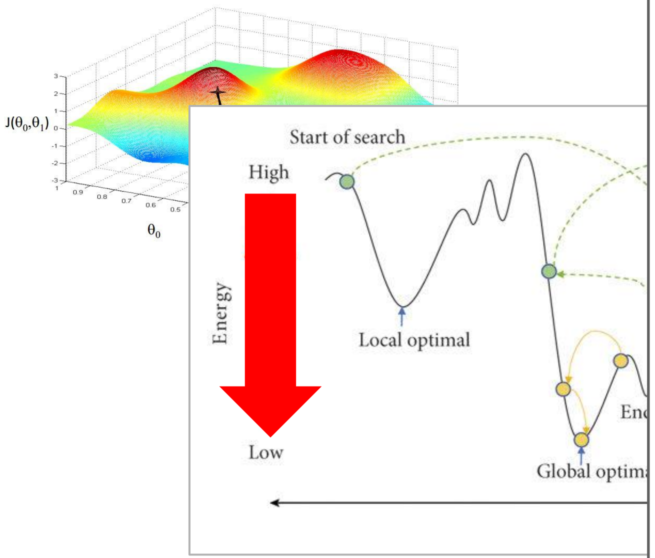


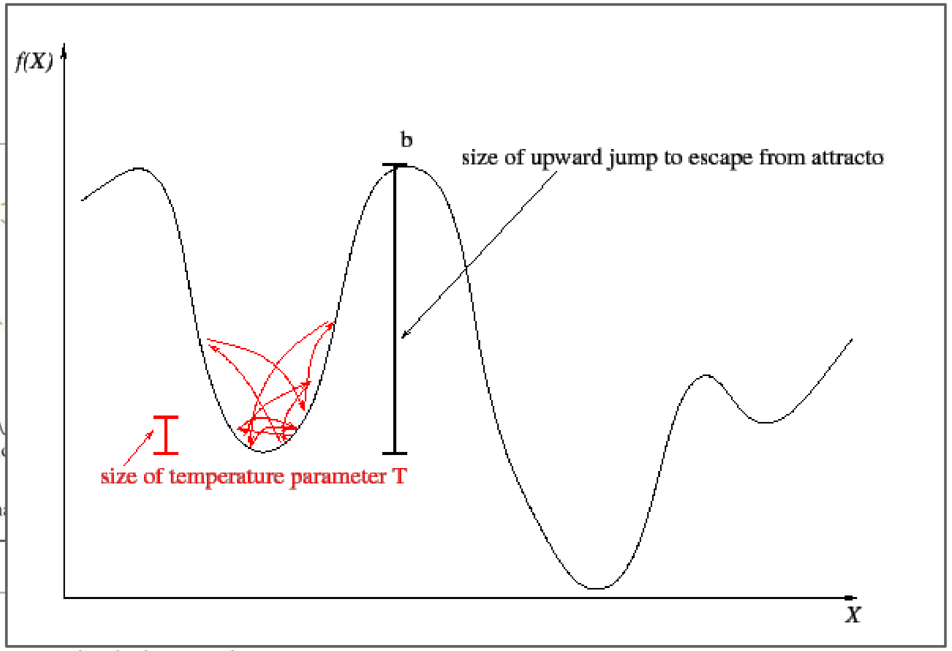
- Accepts **suboptimal** solutions with a probability relative to a **temperature** parameter and has an **adaptive step size**
- Provides better exploration and possibility of jumping out of local optima

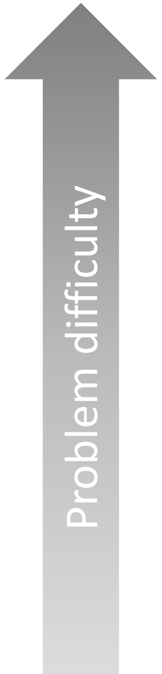


- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- **Simulated annealing**
- Hill climbing

Simulated Annealing







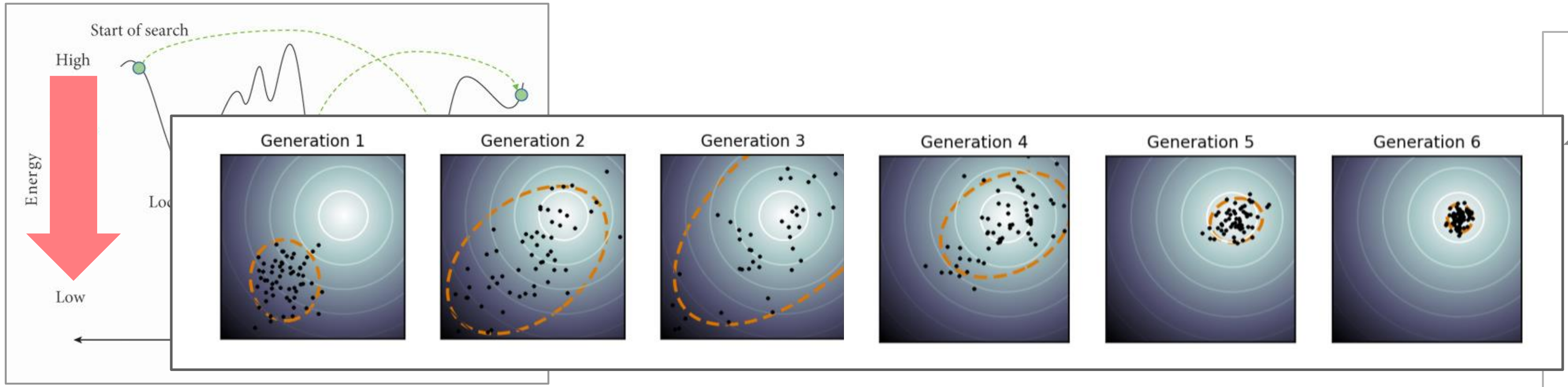
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- **Simulated annealing**
- Hill climbing

- Accepts **suboptimal** solutions with a probability relative to a **temperature** parameter and has an **adaptive step size**
- Provides better exploration and possibility of jumping out of local optima

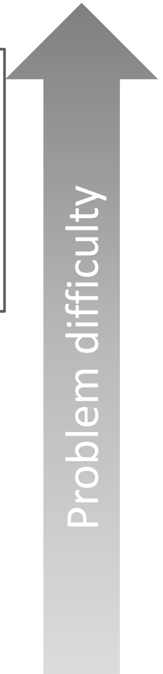
- **One sample at a time** gives a very limited view of the optimization landscape and scheduled temperature leads to premature convergence

https://www.researchgate.net/figure/Principle-of-the-simulated-annealing-algorithm_fig5_360434054
https://www.researchgate.net/figure/Simulated-Annealing-if-the-temperature-is-very-low-wrt-the-jump-size-SA-risks-a_fig4_30531067

Density Model-Based ES

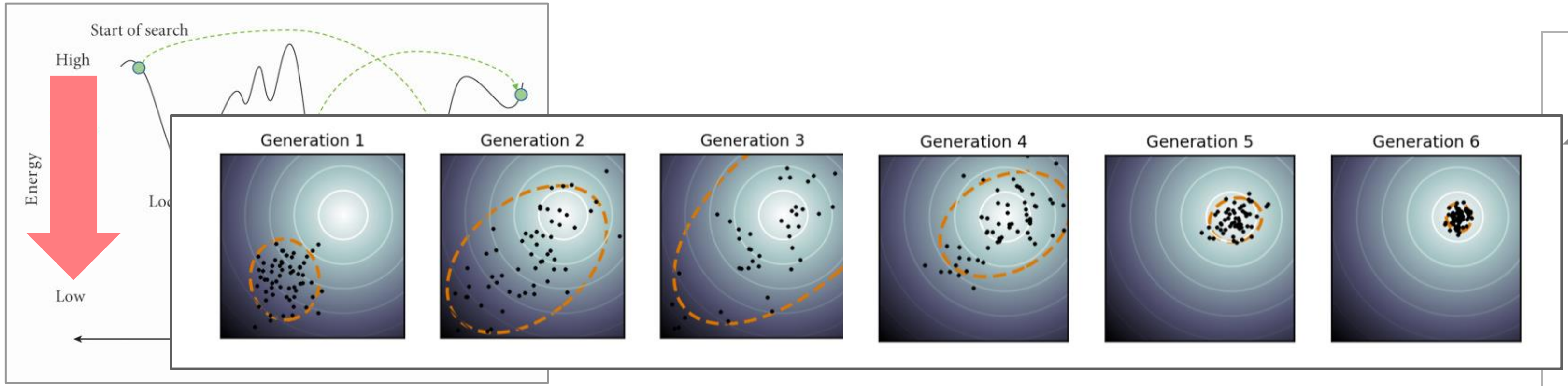


- Randomly Sampling **<u>a number of points</u>** from a distribution

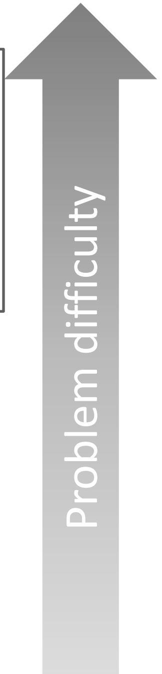


- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **<u>Density model-based Evolutionary Strategy (ES)</u>**
- Simulated annealing
- Hill climbing

Density Model-Based ES

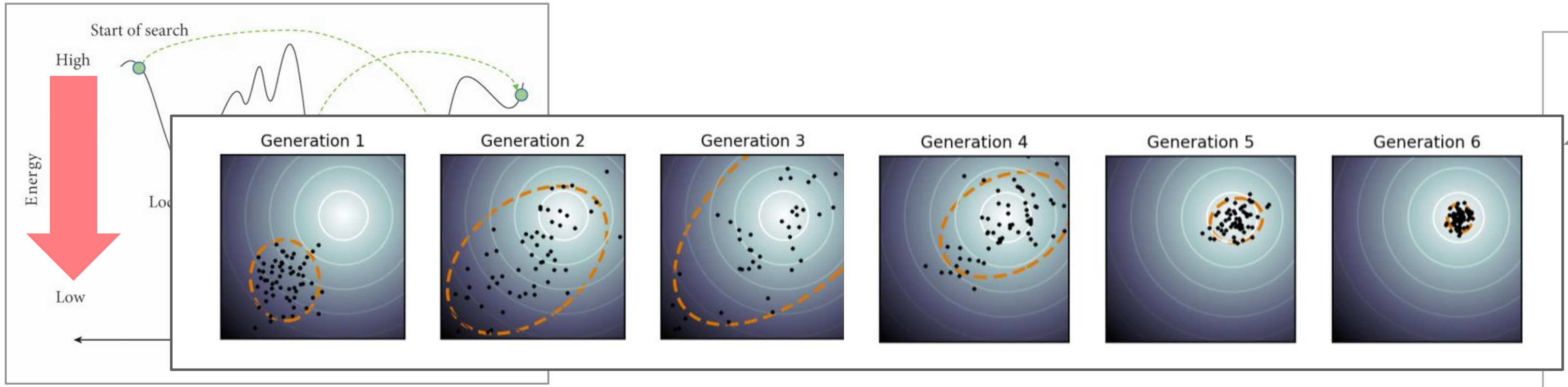


- Randomly Sampling **<u>a number of points</u>** from a distribution: **Generation** and **Individuals**



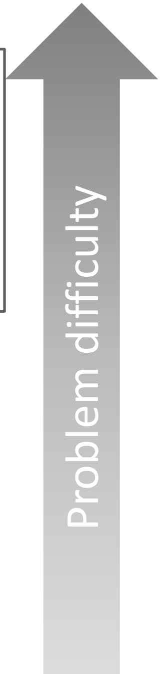
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **<u>Density model-based Evolutionary Strategy (ES)</u>**
- Simulated annealing
- Hill climbing

Density Model-Based ES



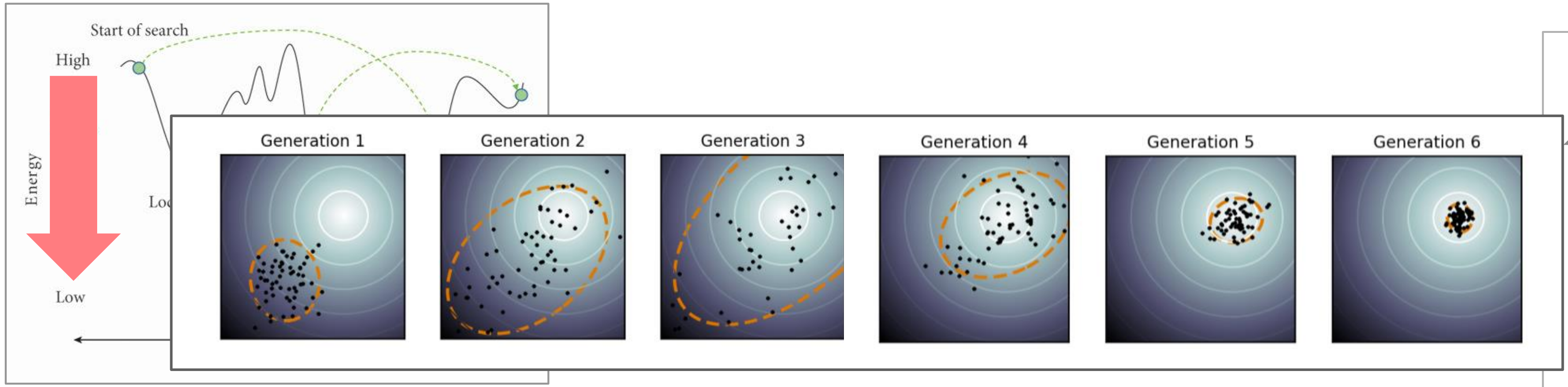
- Randomly Sampling **<u>a number of points</u>** from a distribution: **Generation** and **Individuals**

- **Evaluation:** Calculating the function value **(Fitness)** for each individual

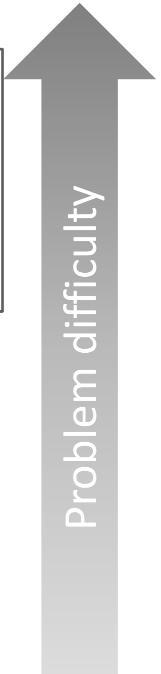


- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **<u>Density model-based Evolutionary Strategy (ES)</u>**
- Simulated annealing
- Hill climbing

Density Model-Based ES

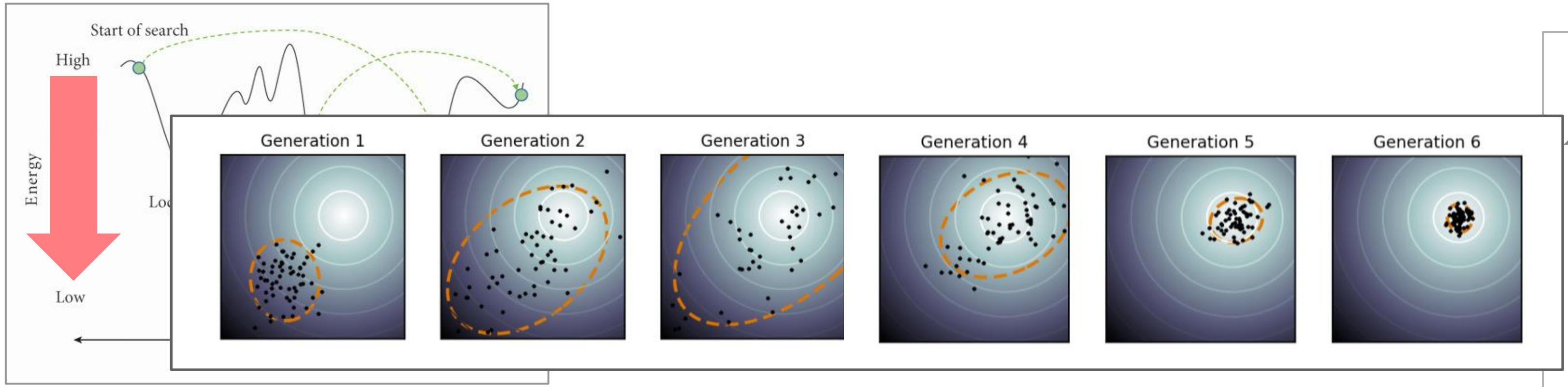


- Randomly Sampling **<u>a number of points</u>** from a distribution: **Generation** and **Individuals**
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **<u>Density model-based Evolutionary Strategy (ES)</u>**
- Simulated annealing
- Hill climbing

Density Model-Based ES



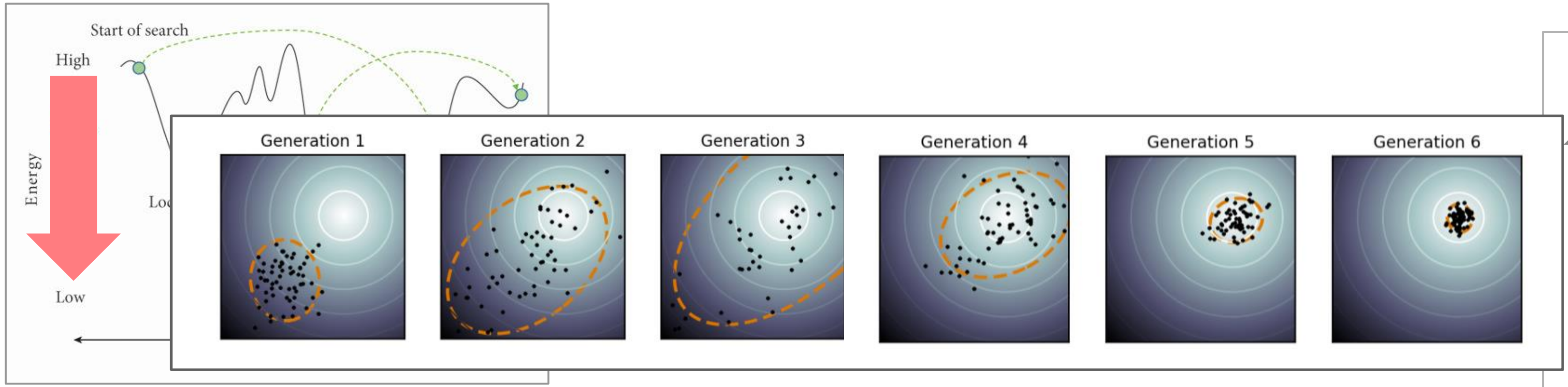
- Randomly Sampling **<u>a number of points</u>** from a distribution: **Generation** and **Individuals**
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- Updating density model

Problem difficulty

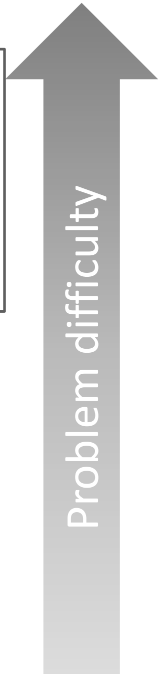
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **<u>Density model-based Evolutionary Strategy (ES)</u>**
- Simulated annealing
- Hill climbing

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

Density Model-Based ES



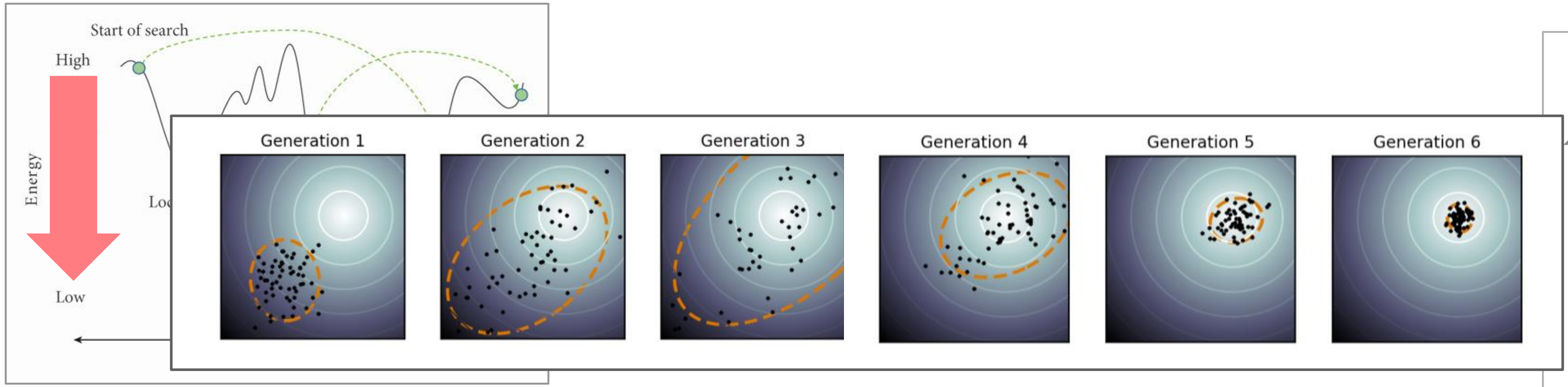
- Randomly Sampling <u>**a number of points**</u> from a distribution: **Generation** and **Individuals**
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- Updating density model
- Creating **next generation**



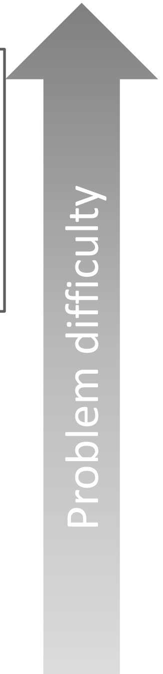
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - <u>**Density model-based Evolutionary Strategy (ES)**</u>
- Simulated annealing
- Hill climbing

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

Density Model-Based ES



- Randomly Sampling **a number of points** from a distribution: **Generation** and **Individuals**
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- Updating density model
- Creating **next generation**



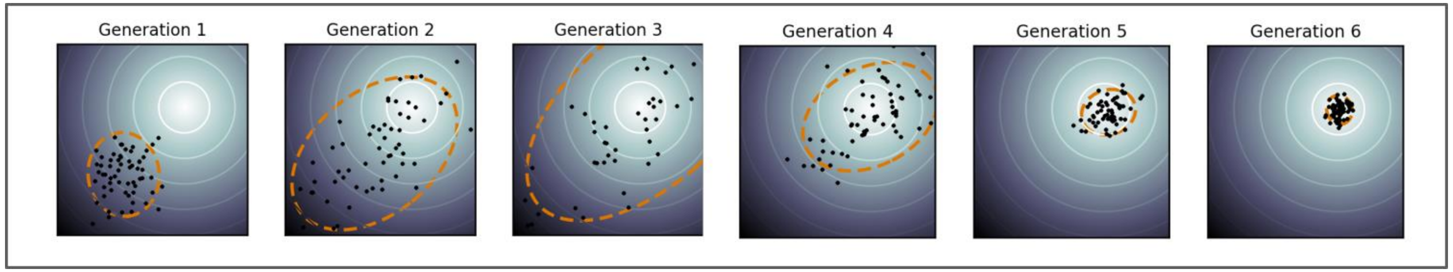
- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **Density model-based Evolutionary Strategy (ES)**
- Simulated annealing
- Hill climbing

Outline

- Introduction and Background
- **Proposed approach**
- Experiments and Results
- Conclusion and Future steps

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



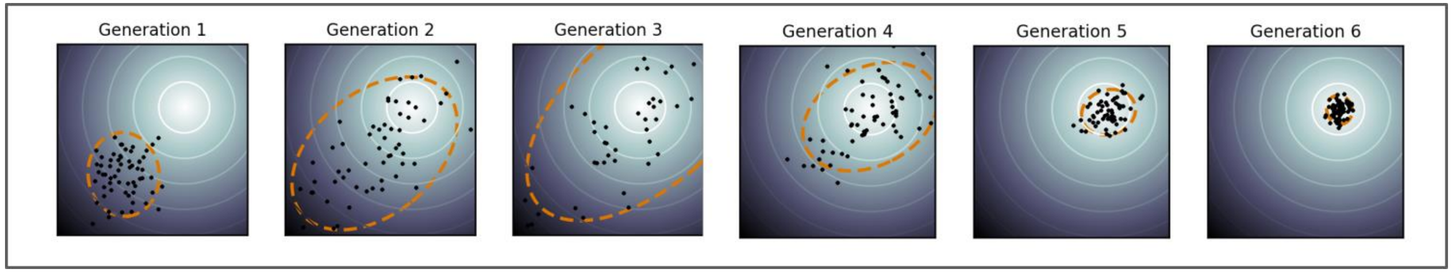
Problem difficulty

- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- Bayesian Optimization
- Simulated annealing
- Hill climbing

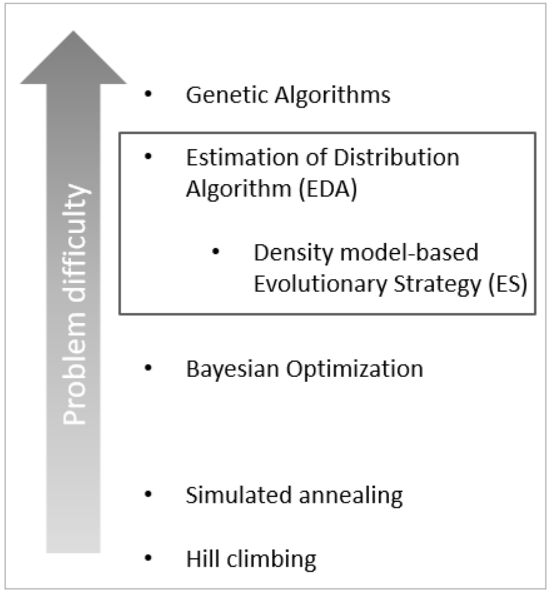
"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



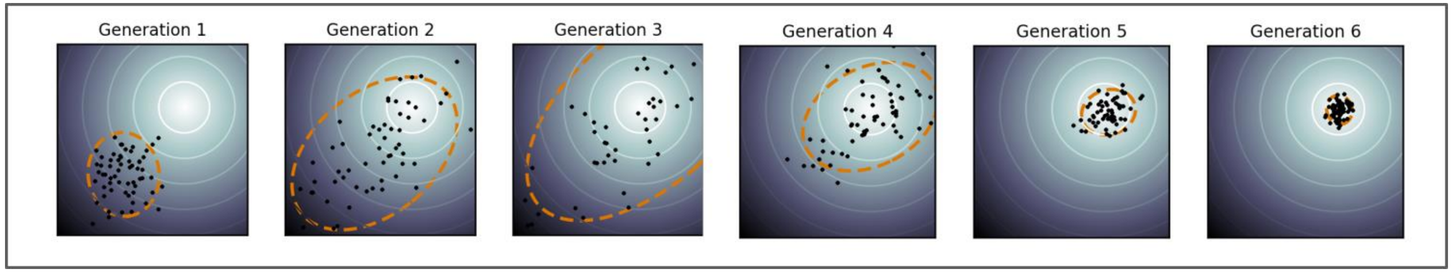
- Randomly sampling the 1st generation from a standard **Gaussian** distribution
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Updating covariance matrix and mean of the Gaussian
- Creating next generation from the updated Gaussian



"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



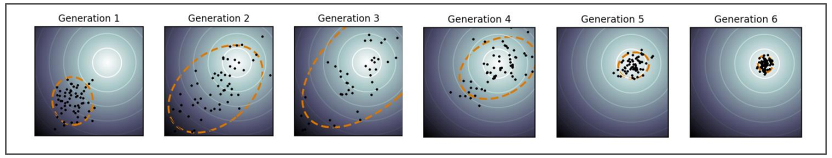
- Randomly sampling the 1st generation from a standard **<u>Gaussian</u>** distribution
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Updating covariance matrix and mean of the Gaussian
- Creating next generation from the updated Gaussian

- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Bayesian Optimization
- Simulated annealing
- Hill climbing

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

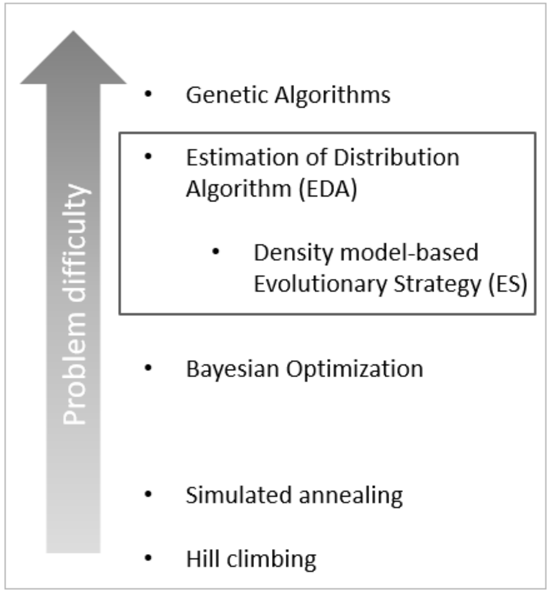
CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



Drawbacks of CMA-ES:

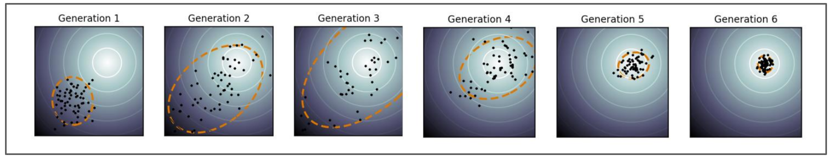
- Arbitrary choice of distribution limits performance

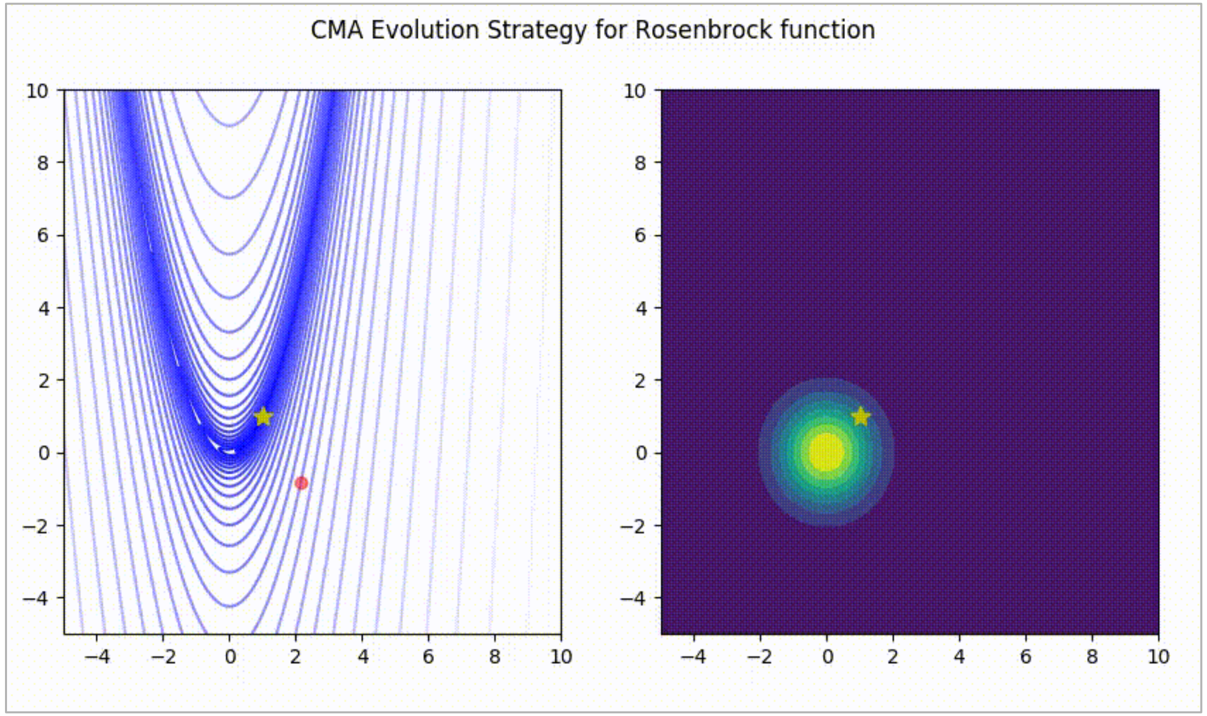


"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES





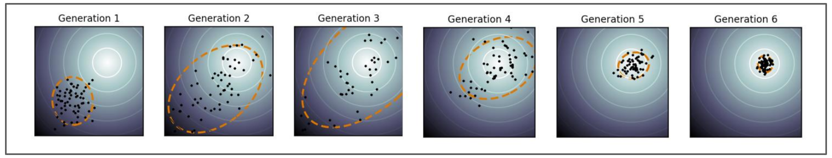
Drawbacks of CMA-ES:

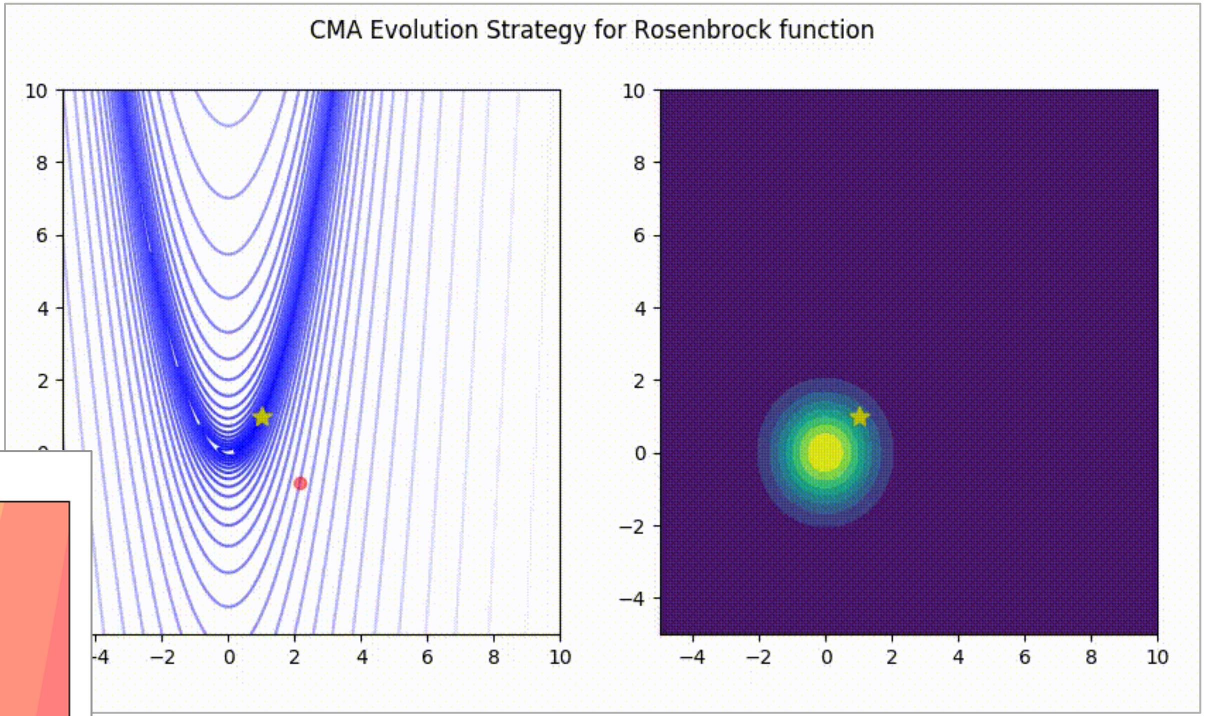
- Arbitrary choice of distribution limits performance

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)., "cmaes 0.3.1." https://pypi.org/project/cmaes/0.3.1/ (accessed Apr. 10, 2024)

CMA-ES

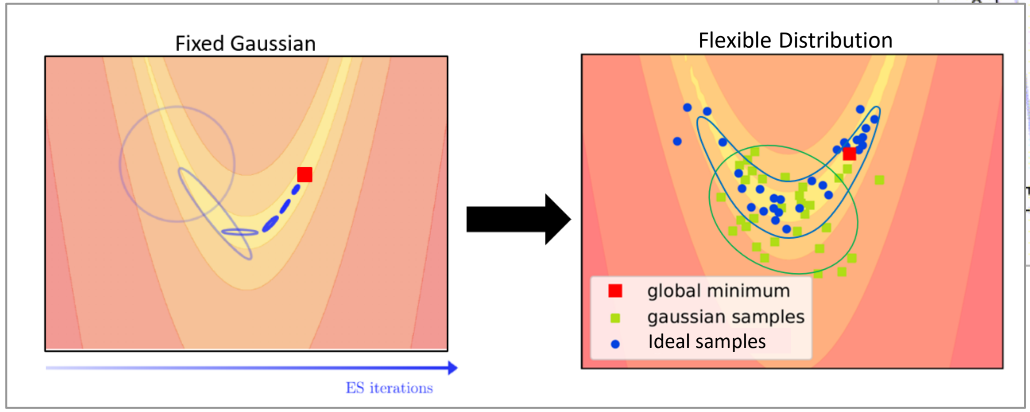
Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES





Drawbacks of CMA-ES:

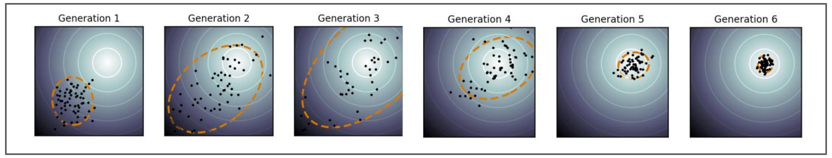
- Arbitrary choice of distribution limits performance



"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)., "cmaes 0.3.1." https://pypi.org/project/cmaes/0.3.1/ (accessed Apr. 10, 2024)
L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



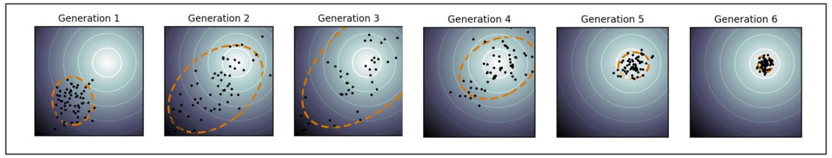
Drawbacks of CMA-ES:

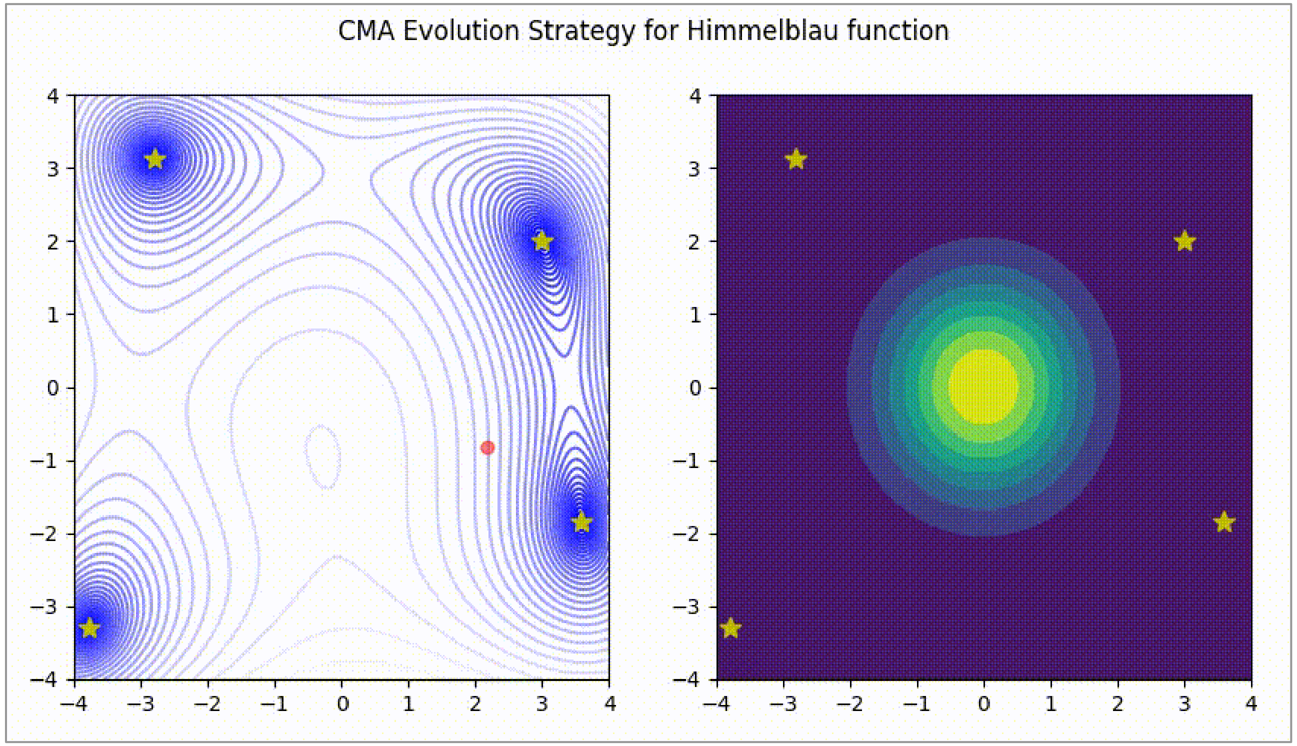
- Arbitrary choice of distribution limits performance
- Starting from scratch and withdrawing gained knowledge

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

CMA-ES

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



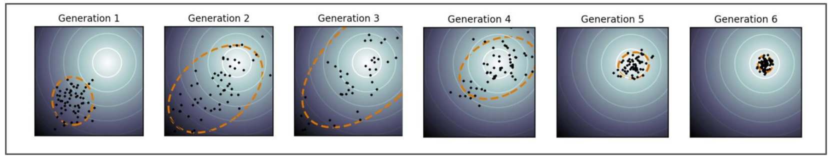


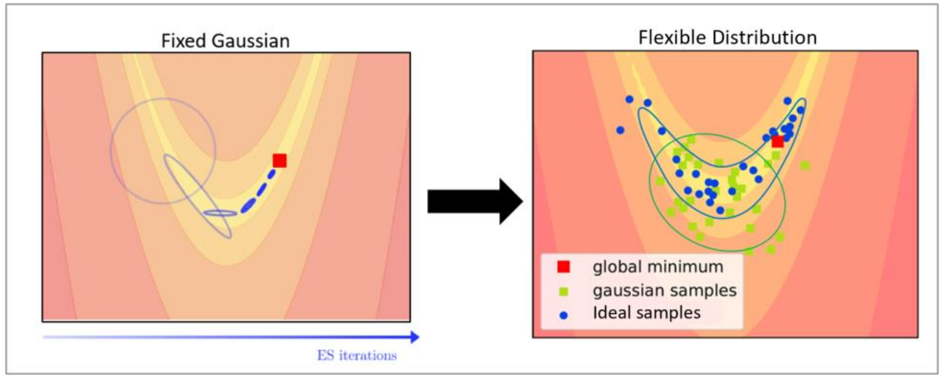
Drawbacks of CMA-ES:

- Arbitrary choice of distribution limits performance
- Starting from scratch and withdrawing gained knowledge

Proposed Approach

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES





Drawbacks of CMA-ES:

- Arbitrary choice of distribution limits performance
- Starting from scratch and withdrawing gained knowledge

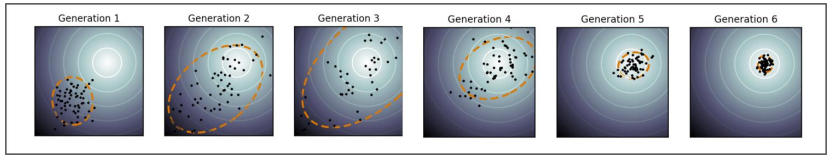
We propose using **Normalizing Flows (NF)** instead of Gaussian to model the density

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

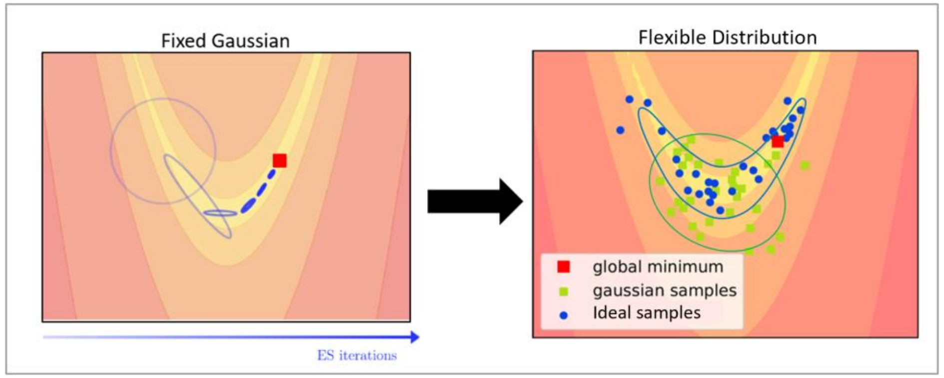
Proposed Approach

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a popular Density Model-Based ES



Drawbacks of CMA-ES:

- Arbitrary choice of distribution limits performance
- Starting from scratch and withdrawing gained knowledge



We propose using **Normalizing Flows (NF)** instead of Gaussian to model the density

- Flexible enough to adapt to the shape of the landscape
- Simultaneous search of multiple modes
- Fine tuning instead of starting from scratch

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)
L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

Normalizing Flows

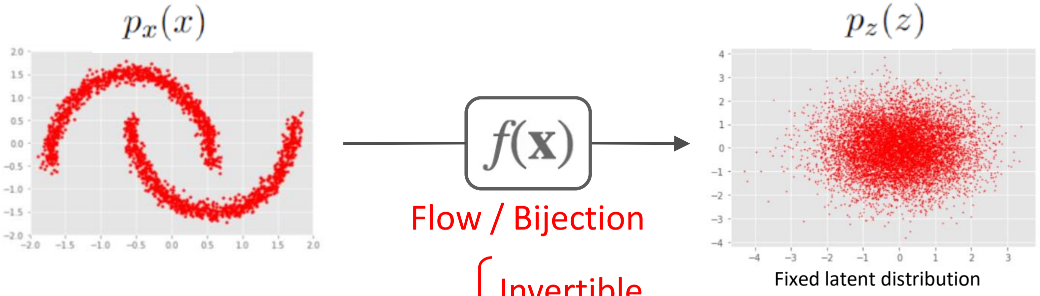
Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



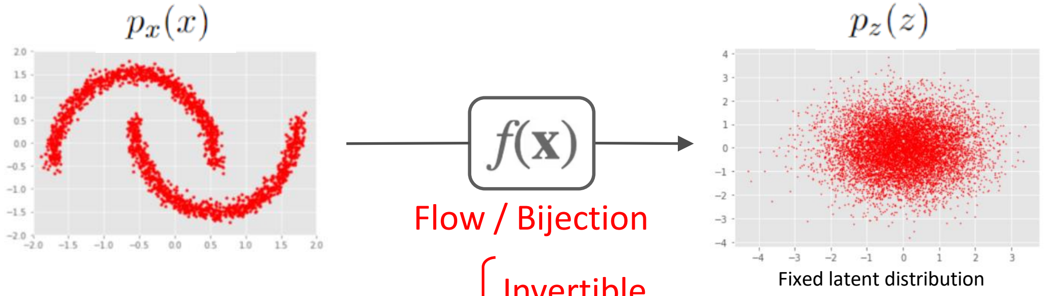
$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{Z}}(f(\mathbf{x})) \left| \det Df(\mathbf{x}) \right|$$

Invertible
Differentiable

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{Z}}(f(\mathbf{x})) \left| \det Df(\mathbf{x}) \right|$$

Invertible
Differentiable

Designing flows is the core research problem in NFs

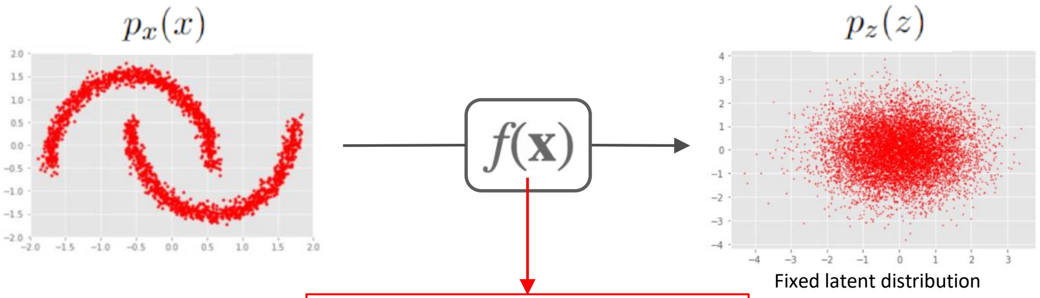
- Each flow layer in NF is a simple invertible, differentiable function
- The overall flow network is a composition of simple flow layers

"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



$$f = f_K \circ f_{K-1} \circ \cdots \circ f_2 \circ f_1$$

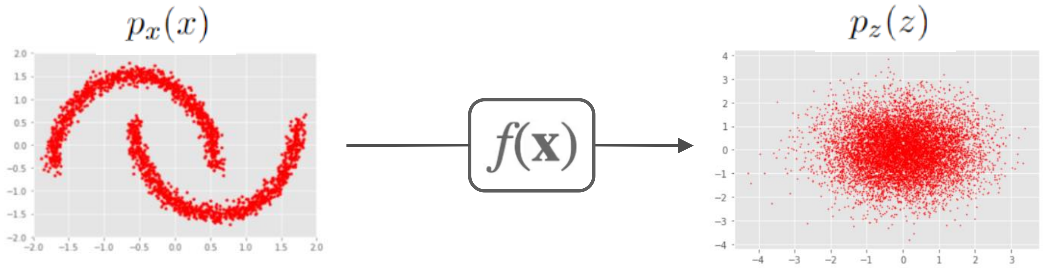
Designing flows is the core research problem in NFs

- Each flow layer in NF is a simple invertible, differentiable function
- The overall flow network is a composition of simple flow layers

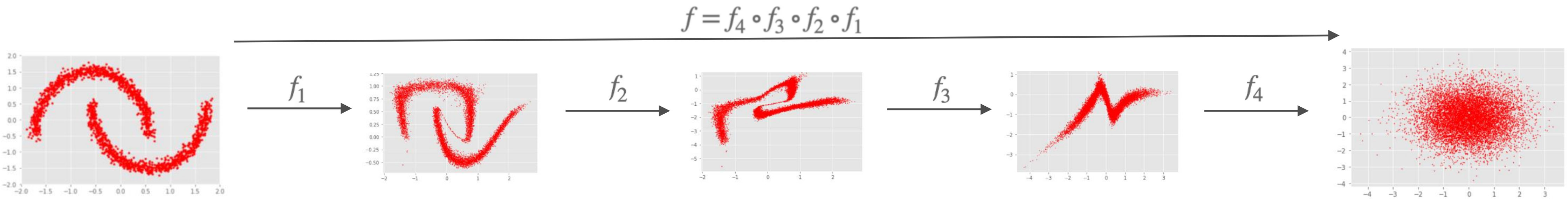
Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



Fixed latent distribution

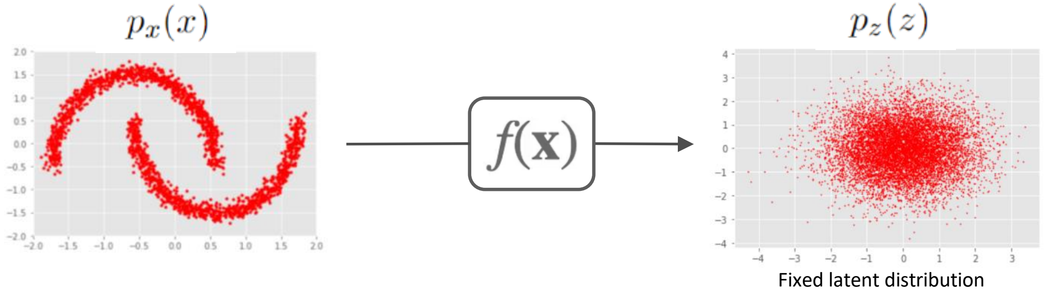


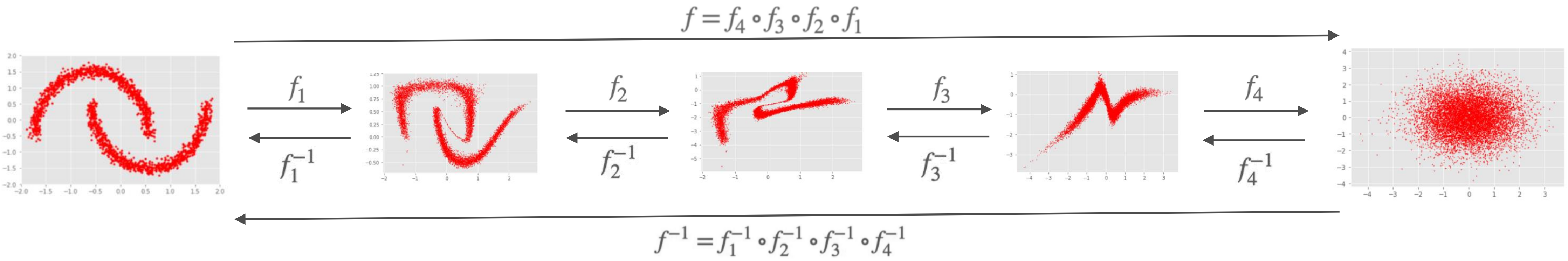
"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



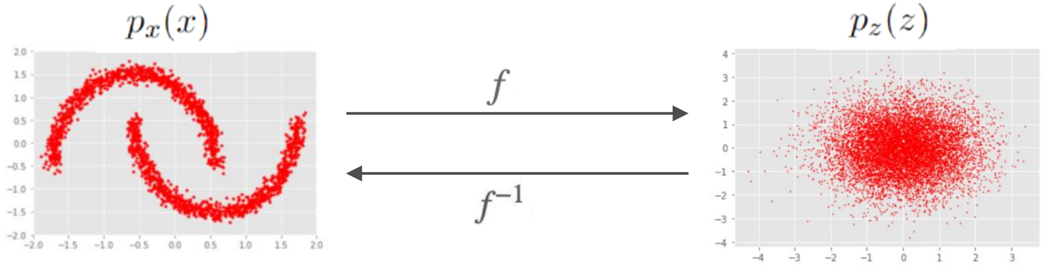


"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



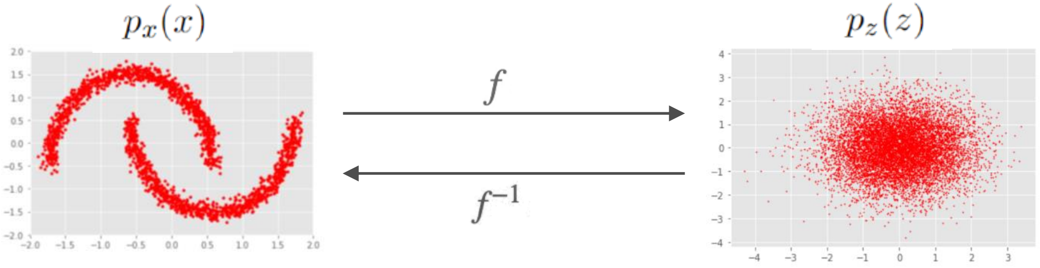
Designing flows is the core research problem in NFs but **NOT** our research goal

"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



Designing flows is the core research problem in NFs but **NOT** our research goal

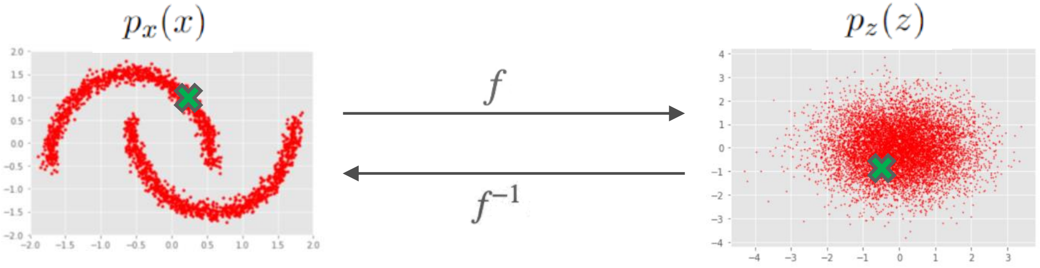
- NF being highly expressive
- Sampling and density evaluation is efficient
- <u>**Exact**</u> forward and backward mapping

"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



Designing flows is the core research problem in NFs but **NOT** our research goal

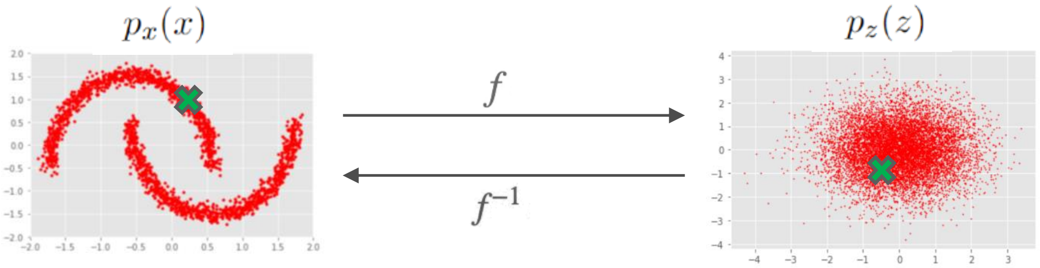
- NF being highly expressive
- Sampling and density evaluation is efficient
- **Exact** forward and backward mapping

"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Normalizing Flows

Structurally restricted Generative Neural Networks (bijective GNNs):

Learn probability distribution $p_x(x)$ over a random variable X from observations



Designing flows is the core research problem in NFs but **NOT** our research goal

- NF being highly expressive
- Sampling and density evaluation is efficient
- **Exact** forward and backward mapping

We adopted **RealNVP** Normalizing Flows for our work

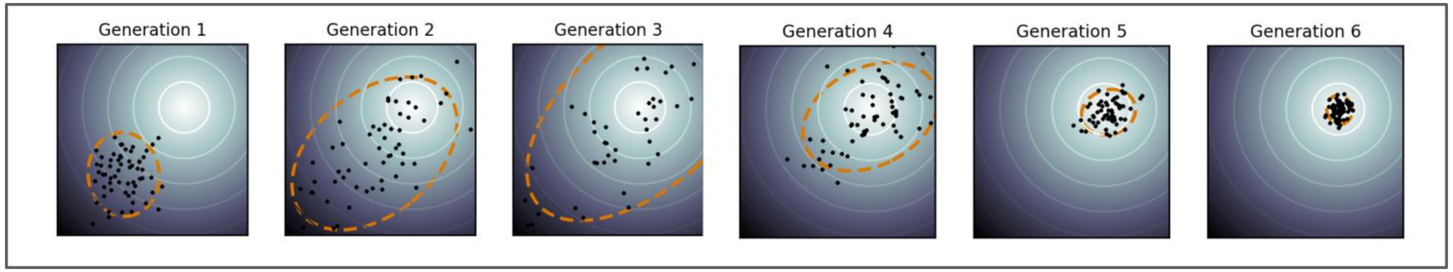
"Tutorial on normalizing flows, part 2." https://engineering.papercup.com/posts/normalizing-flows-part-2/ (accessed Apr. 10, 2024)

Outline

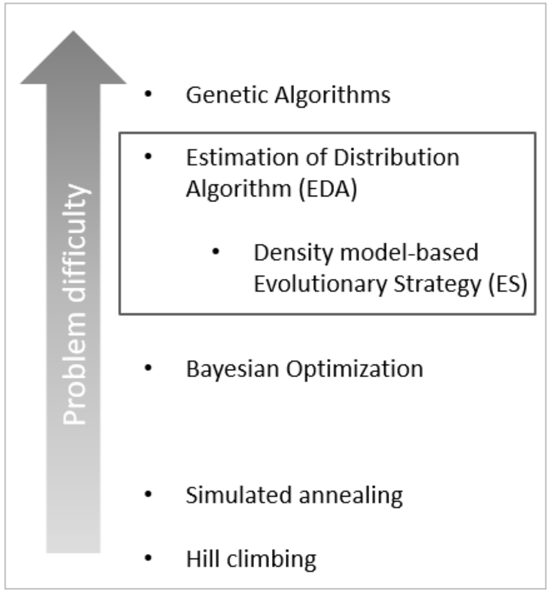
- Introduction and Background
- Proposed approach
- **Experiments and Results**
- Conclusion and Future steps
- Conclusion / PhD Plan

Recap

CMA-ES, a popular Density Model-Based ES



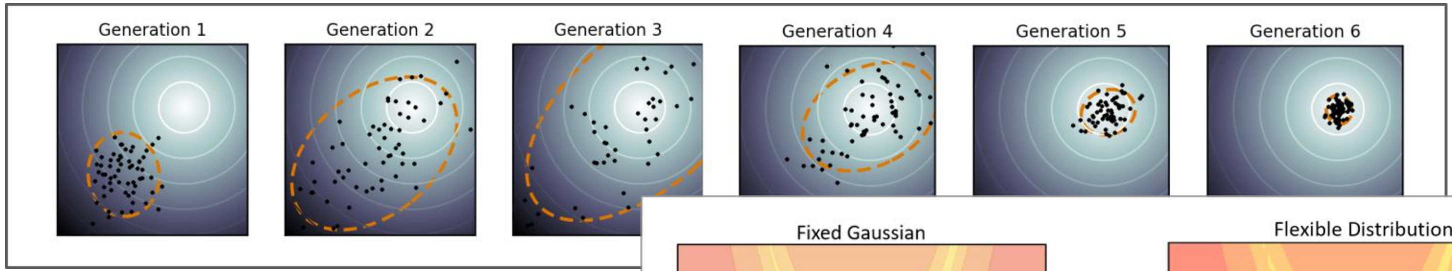
- Randomly sampling the 1st generation from a standard **Gaussian** distribution
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Updating covariance matrix and mean of the Gaussian
- Creating next generation from the updated Gaussian

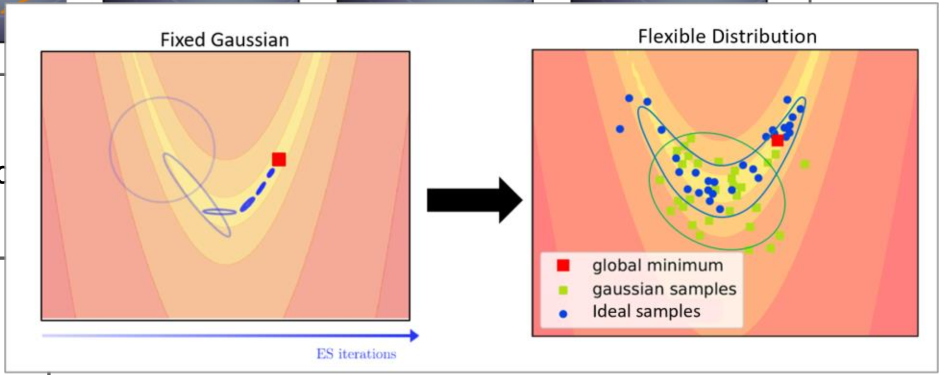


"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

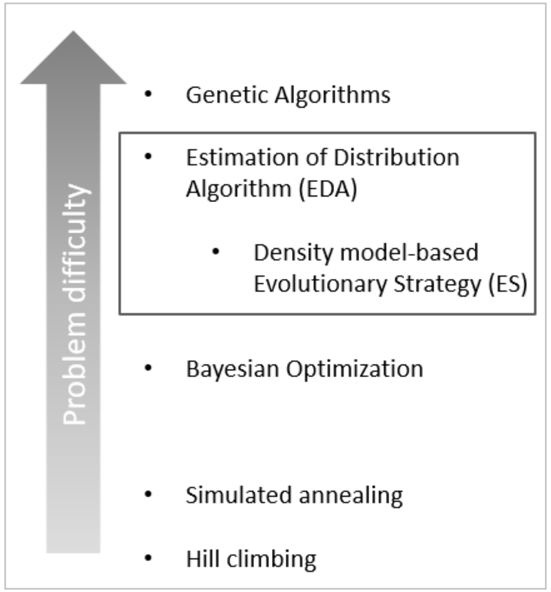
Recap

CMA-ES, a popular Density Model-Based ES





- Randomly sampling the 1st generation from a standard **Gaussian** distribution
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Updating covariance matrix and mean of the Gaussian
- Creating next generation from the updated Gaussian

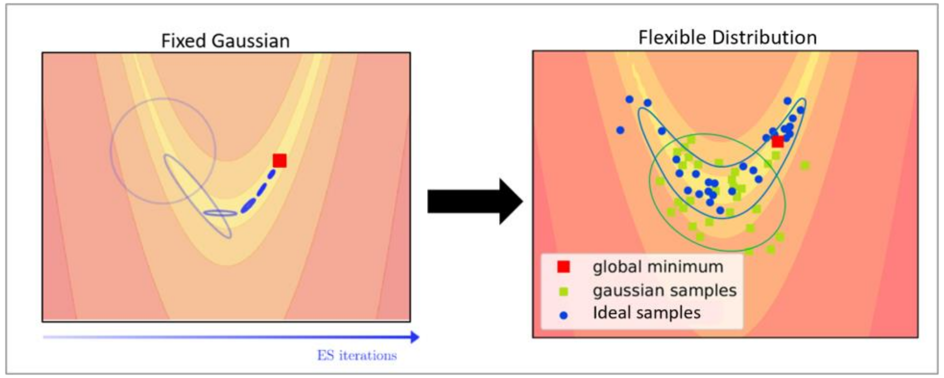


"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)
L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

Proposed Approach

We proposed using **Normalizing Flows (NF)** instead of Gaussian to model the density

- Flexible enough to adapt to the shape of the landscape
- Simultaneous search of multiple modes
- Fine tuning instead of starting from scratch

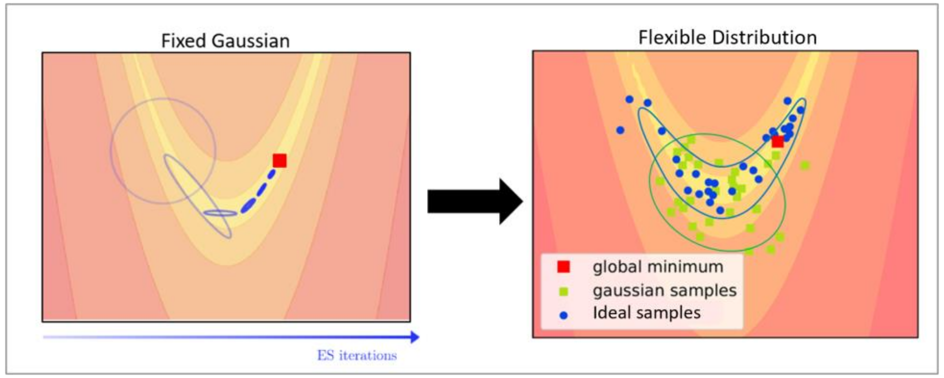


L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

Proposed Approach

We proposed using **Normalizing Flows (NF)** instead of Gaussian to model the density

- Flexible enough to adapt to the shape of the landscape
- Simultaneous search of multiple modes
- Fine tuning instead of starting from scratch



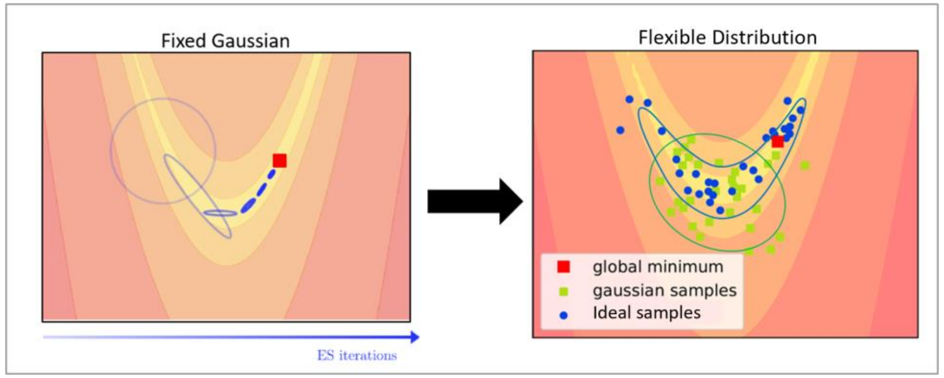
Integrating RealNVP in Evolution Sterategy

L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

Proposed Approach

We proposed using **Normalizing Flows (NF)** instead of Gaussian to model the density

- Flexible enough to adapt to the shape of the landscape
- Simultaneous search of multiple modes
- Fine tuning instead of starting from scratch



Integrating RealNVP in EDA (Estimation of Distribution Alg.)

- Separate the impact of ES
- Investigate the full potential of NF

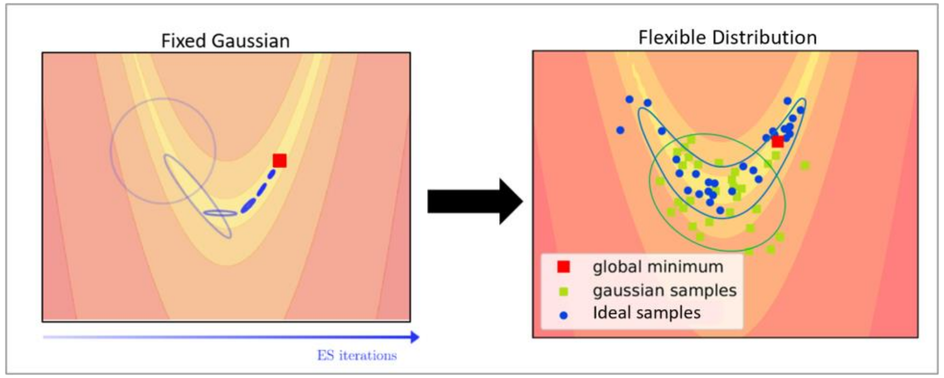
L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

NF + EDA

We proposed using **Normalizing Flows (NF)** instead of Gaussian to model the density

Integrating RealNVP in EDA (Estimation of Distribution Alg.)

- Separate the impact of ES
- Investigate the full potential of NF



- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- Sampling the next generation from the fine-tuned RealNVP

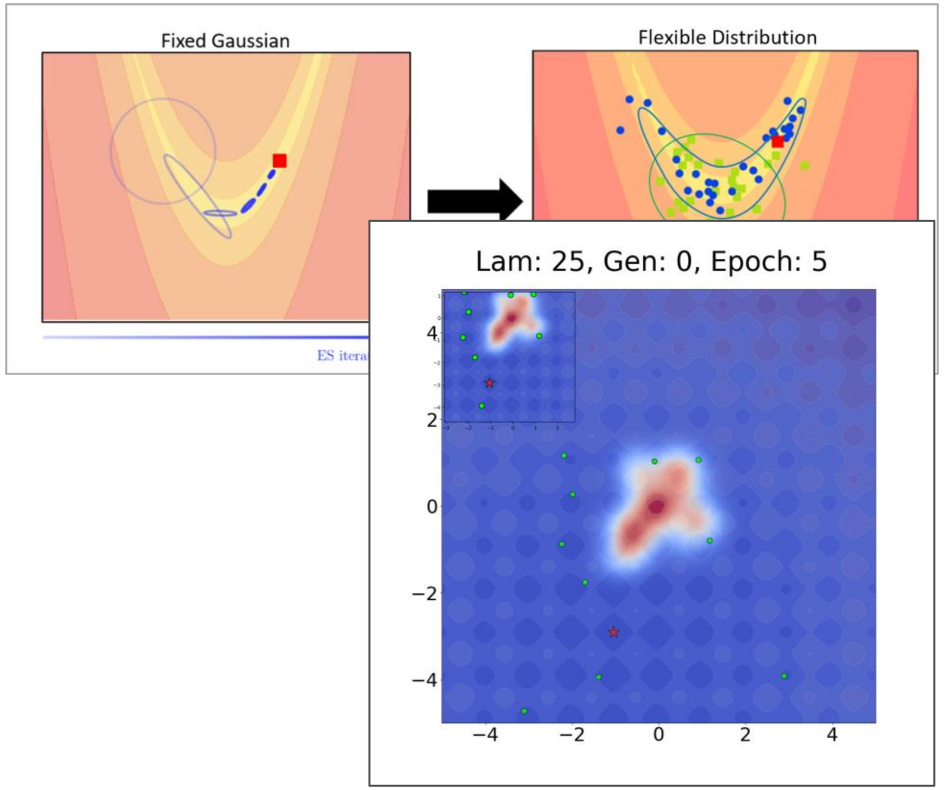
L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

NF + EDA

We proposed using **Normalizing Flows (NF)** instead of Gaussian to model the density

Integrating RealNVP in EDA (Estimation of Distribution Alg.)

- Separate the impact of ES
- Investigate the full potential of NF



- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- Sampling the next generation from the fine-tuned RealNVP

L. Faury, C. Calauzenes, O. Fercoq, and S. Krichen, "Improving Evolutionary Strategies with Generative Neural Networks," arXiv preprint arXiv :1901.11271v1, 2019.

NF + ES

Integrating RealNVP in Evolution Sterategy

- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual in the **original space**
- Selection: selecting the fitter individuals in the **original space**
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- **Remapping** the selected samples to the latent space
- **Taking ES step in the latent space** to generate the next generation
- **Mapping** the new generation to the original space with the fine-tuned RealNVP

Integrating RealNVP in EDA

- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- Sampling the next generation from the fine-tuned RealNVP

NF + ES

Integrating RealNVP in Evolution Sterategy

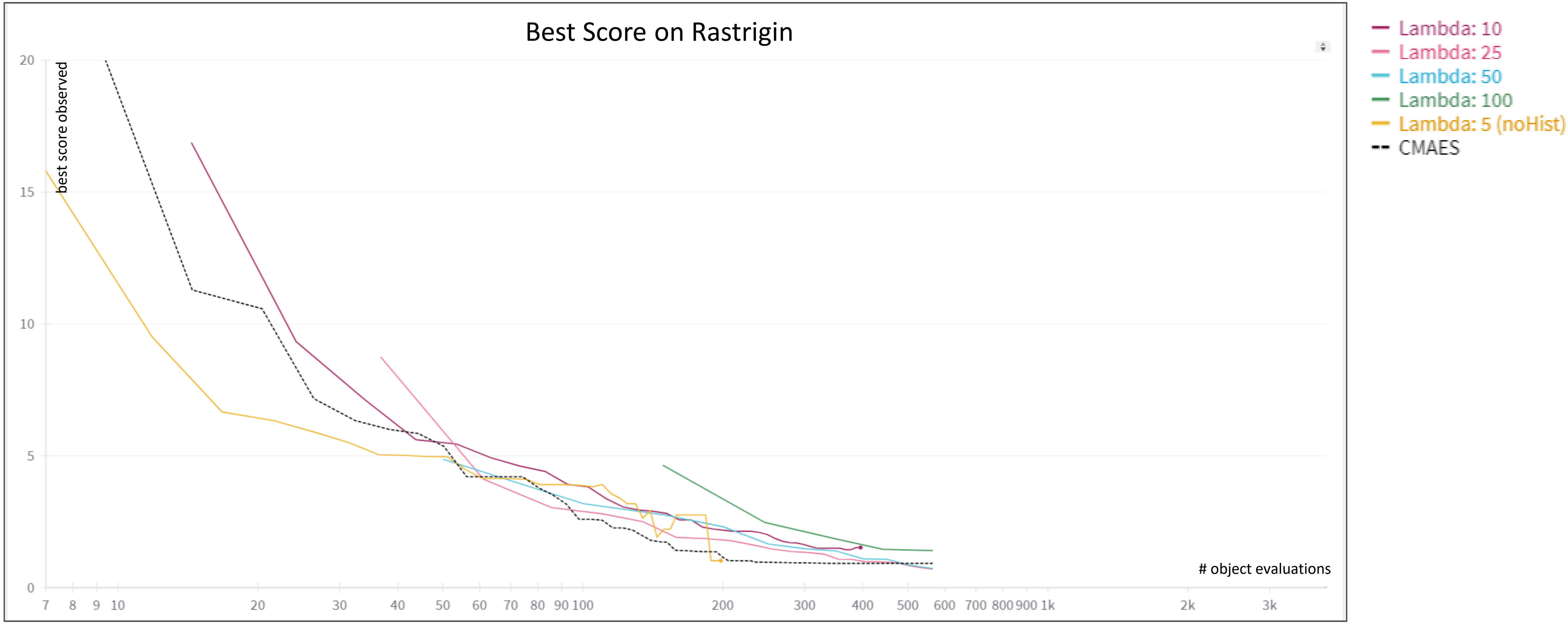
- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual in the **original space**
- Selection: selecting the fitter individuals in the **original space**
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- **Remapping** the selected samples to the latent space
- **Taking ES step in the latent space** to generate the next generation
- **Mapping** the new generation to the original space with the fine-tuned RealNVP

Integrating RealNVP in EDA

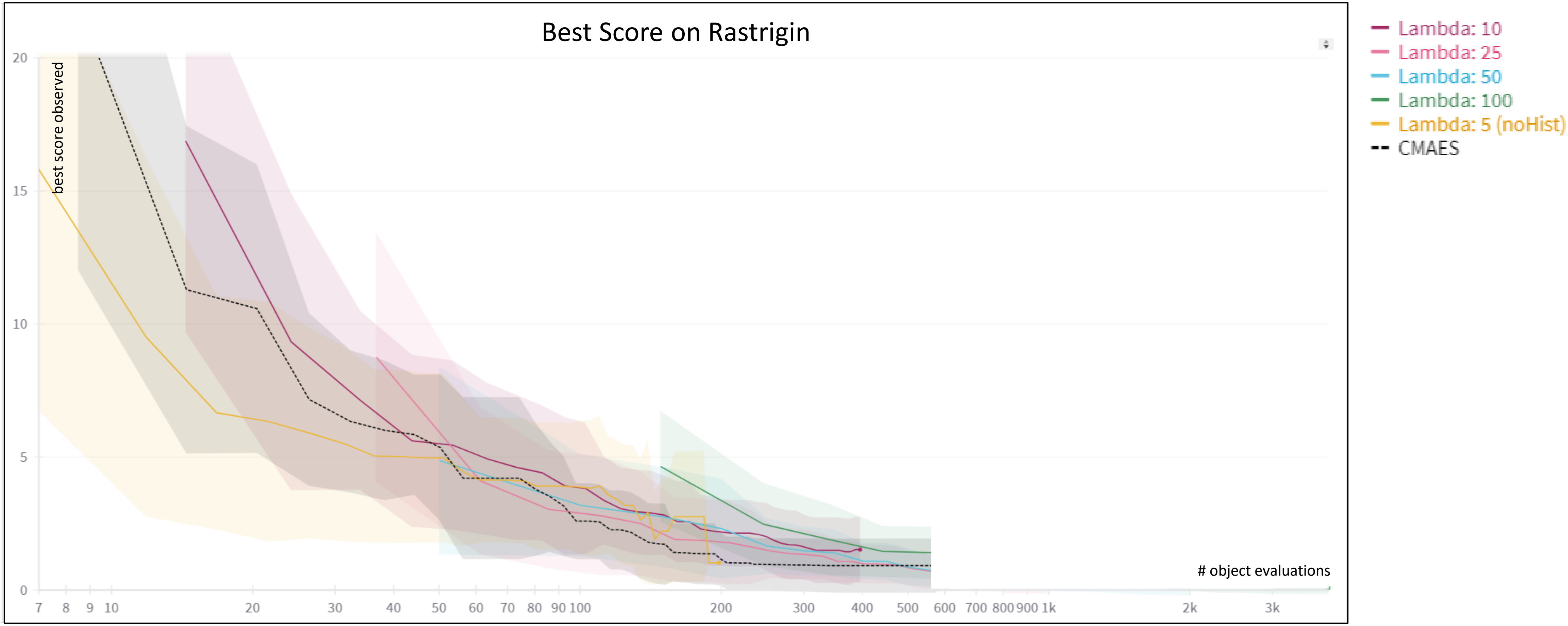
- Sampling the 1st generation from a randomly initialized RealNVP with a GMM base
- Evaluation: calculating fitness for each individual
- Selection: selecting the fitter individuals
- Fine-tuning RealNVP to focus the distribution on the fitter samples
- Sampling the next generation from the fine-tuned RealNVP

NF+ES is expected to outperform NF+EDA

NF + EDA (Estimation of Distribution Alg.)

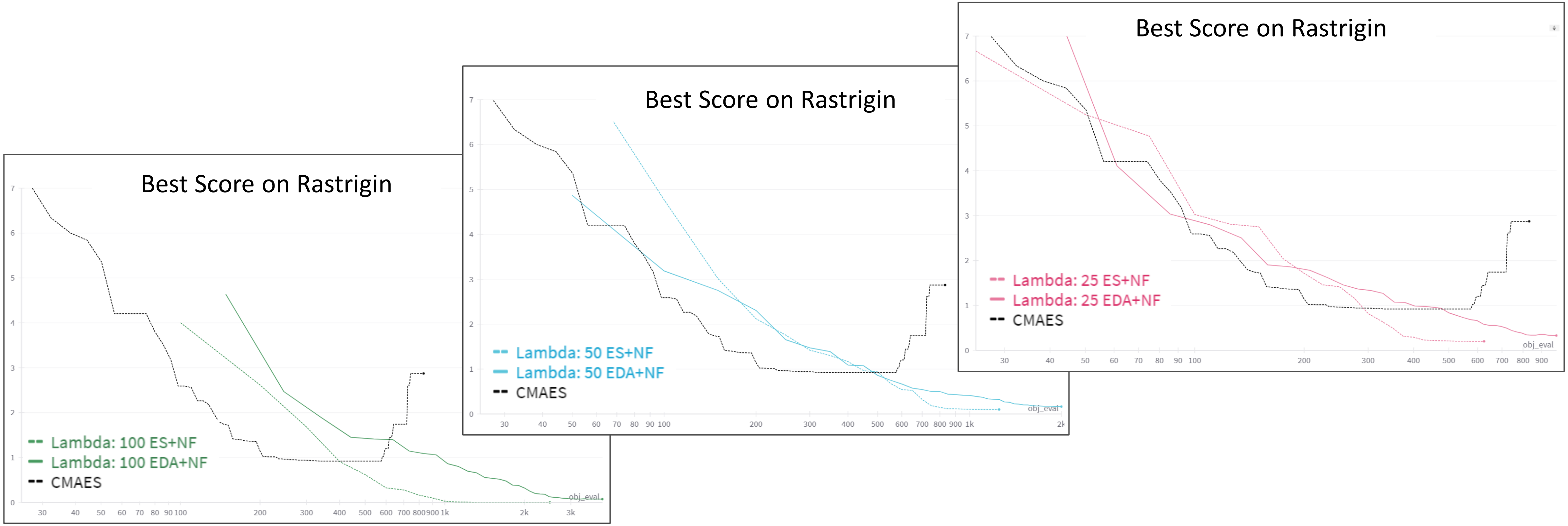


NF + EDA (Estimation of Distribution Alg.)



NF + ES (Evolutionary Strategy)

NF + ES model-based BBO vs NF + EDA Model-based BBO



Outline

- Introduction and Background
- Proposed approach
- Experiments and Results
- **Conclusion and Future steps**

Conclusion and Future Steps

- Preliminary Results:
  - NF has the required capability to enhance model-based BBO
  - NF+EDA results are promising despite the simplicity of the method and with minimum hyperparameter tuning
  - NF+ES results outperform NF+EDA and are competent with CMA-ES even in the preliminary experiments without hyperparameter tuning
- Challenges ahead:
  - Model convergence and Exploration/Exploitation balance
  - Hyperparameter tuning
  - Conducting experiments on real-world data
- Next phase:
  - Applying Meta-optimization to further improve performance and sample efficiency