# Introduction to Black-Box Optimization

## (and how not to freak out when there are no gradients!)

Presented by Sara Karami

October 2023

Machine Learning for Signal Processing (IFT4030/7030)

# Outline

- Introduction and Background
- Well known methods
- My research

# 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

# Optimization

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

Selecting method according to $f(x)$

Problem difficulty

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

# Optimization

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

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 calculate

# 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
- Laboratory experiment

[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

Real-world examples:

- Computer simulation
- Laboratory experiment

Goal: Find a good enough solution with **minimum evaluations**

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


Schematic of a Black Box [1]

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

# 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 Strategies** (ES) and **Randomized based** methods

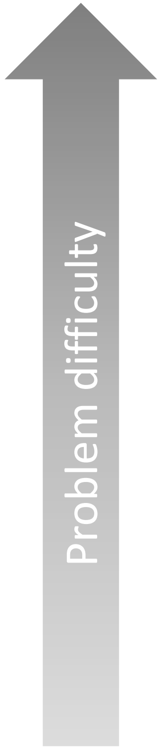
**Derivative free optimization**

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

# BBO Methods

Deciding BBO method

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

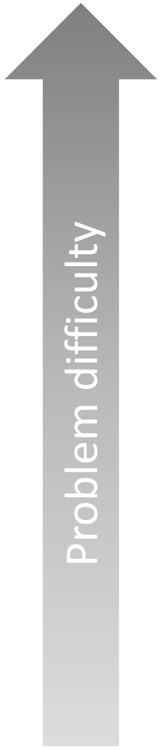



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Bayesian Optimization
- Simulated annealing
- Hill climbing

# BBO Methods

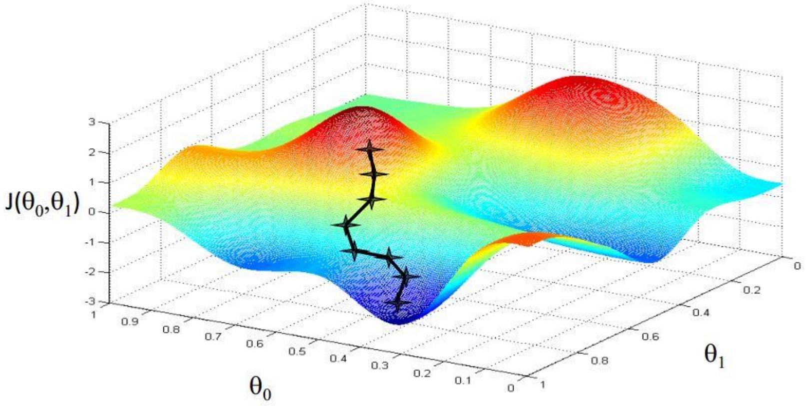
Deciding BBO method

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




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

# Hill Climbing

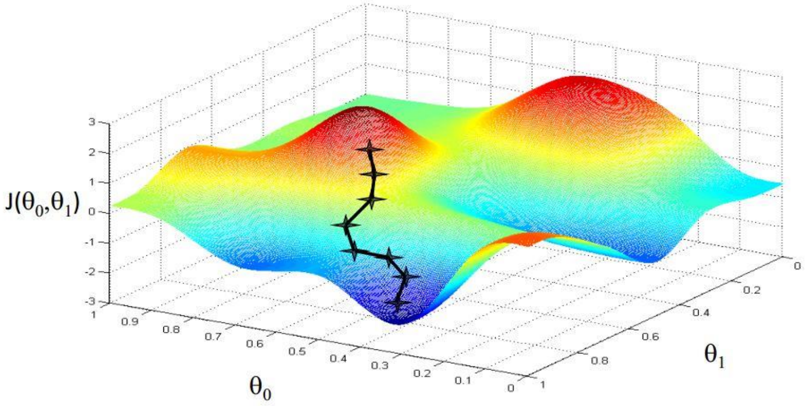



- Starts from a random point
- Randomly selects a neighbor with **better** value

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!

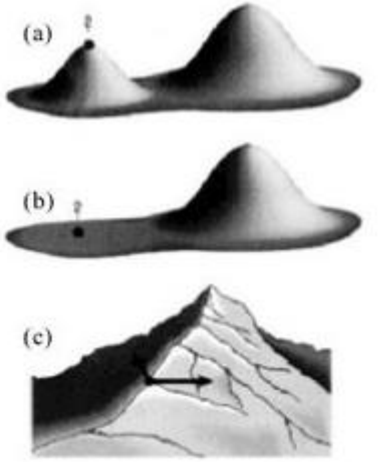
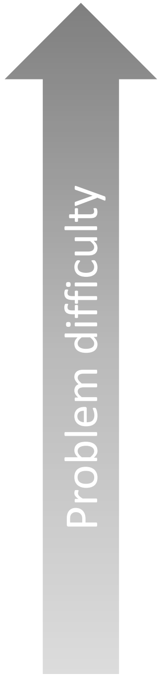


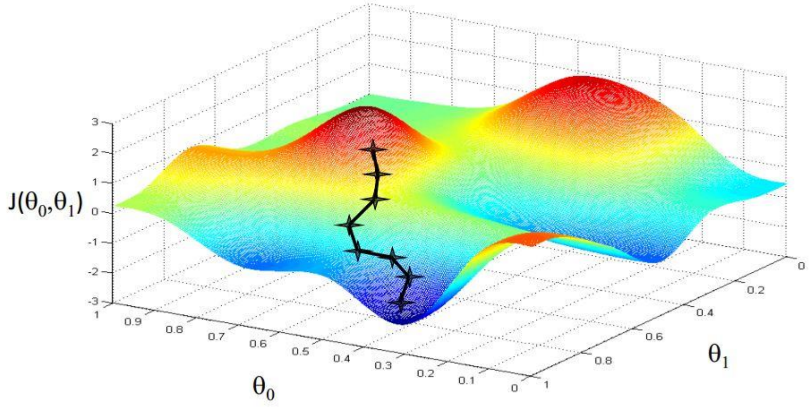

Figure 5.9 Local maxima, Plateaus and ridge situation for Hill Climbing




- 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!

Doesn't it sound like gradient descent?

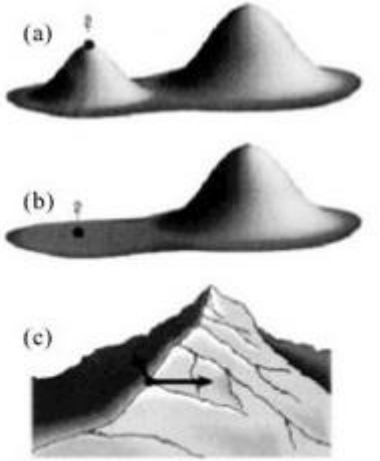
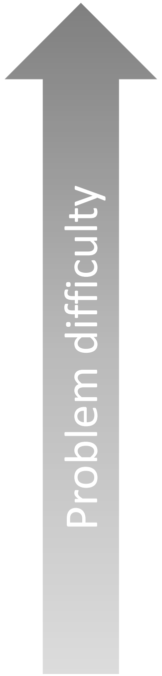


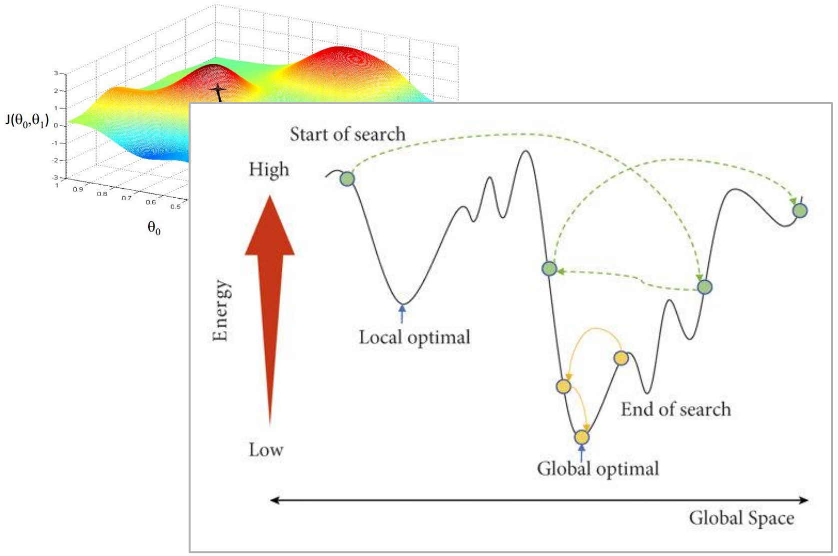

Figure 5.9 Local maxima, Plateaus and ridge situation for Hill Climbing




- 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

# Simulated Annealing



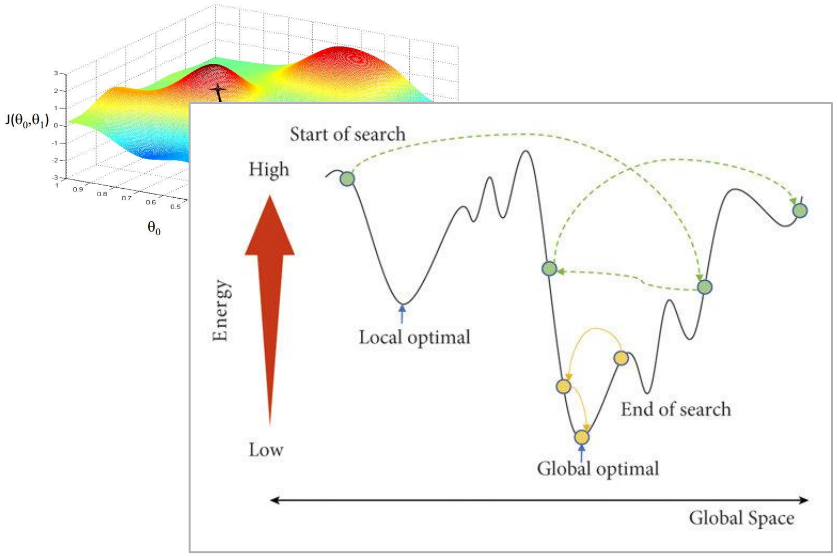

- Accepts **worse** solutions with a probability relative to a **temperature** parameter
- Provides better exploration and can jump out of local optima!

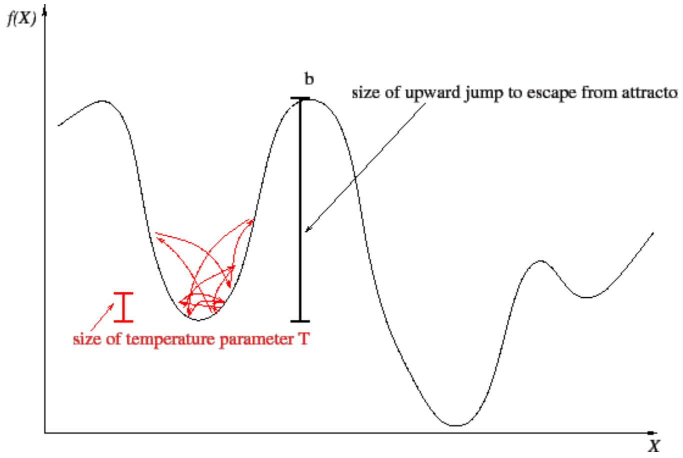
Problem difficulty

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

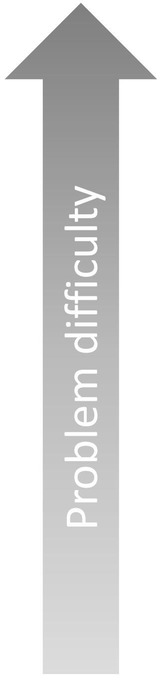
https://www.researchgate.net/figure/Principle-of-the-simulated-annealing-algorithm_fig5_360434054

# Simulated Annealing



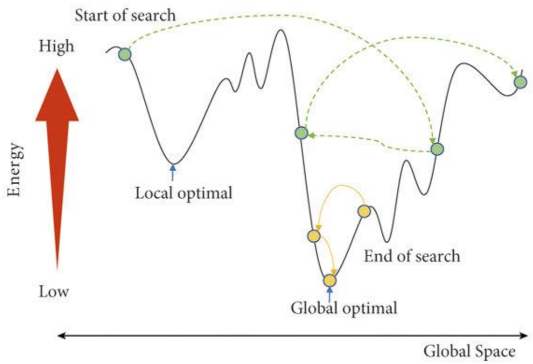

- Accepts **worse** solutions with a probability relative to a **temperature** parameter
- Provides better exploration and can jump out of local optima

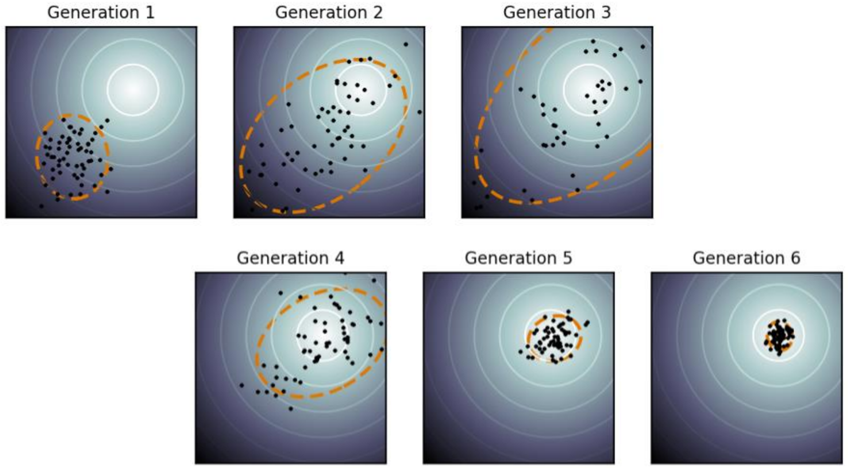



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - Density model-based Evolutionary Strategy (ES)
- **Simulated annealing**
- Hill climbing

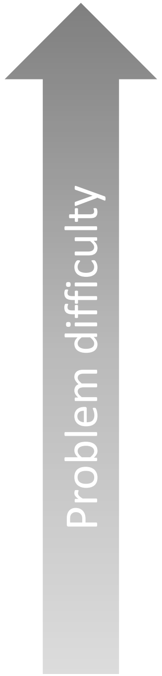
https://www.researchgate.net/figure/Principle-of-the-simulated-annealing-algorithm_fig5_360434054
https://www.researchgate.net/figure/Simulated-Annealing-if-the-temperature-is-very-low-wrt-the-jump-size-SA-risks-a_fig4_30531067

# Density Model-Based ES



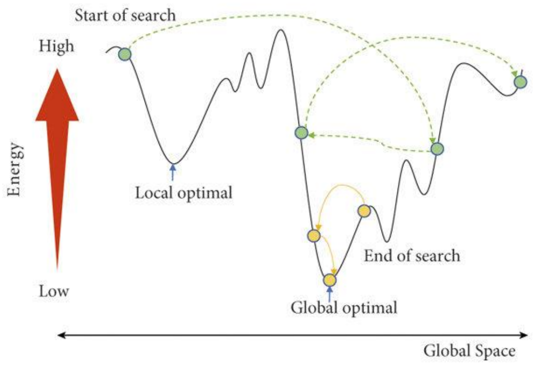

- **Population**: Sampling a bunch of points **(individuals)** from a distribution
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- Updating density model
- Create next **generation**

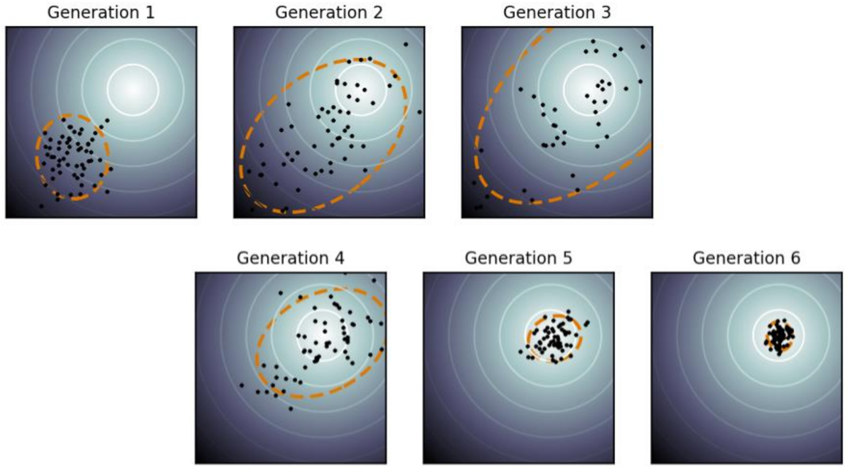



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
    - **Density model-based Evolutionary Strategy (ES)**
- Simulated annealing
- Hill climbing

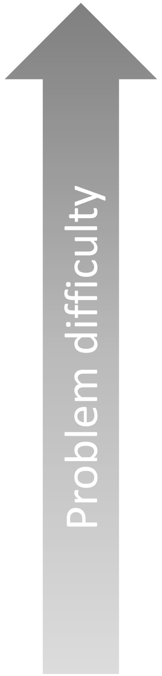
"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

# Density Model-Based ES



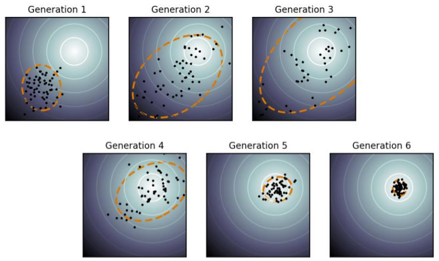

- **Population**: Sampling a bunch of points **(individuals)** from a distribution
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- Updating density model
- Create next **generation**

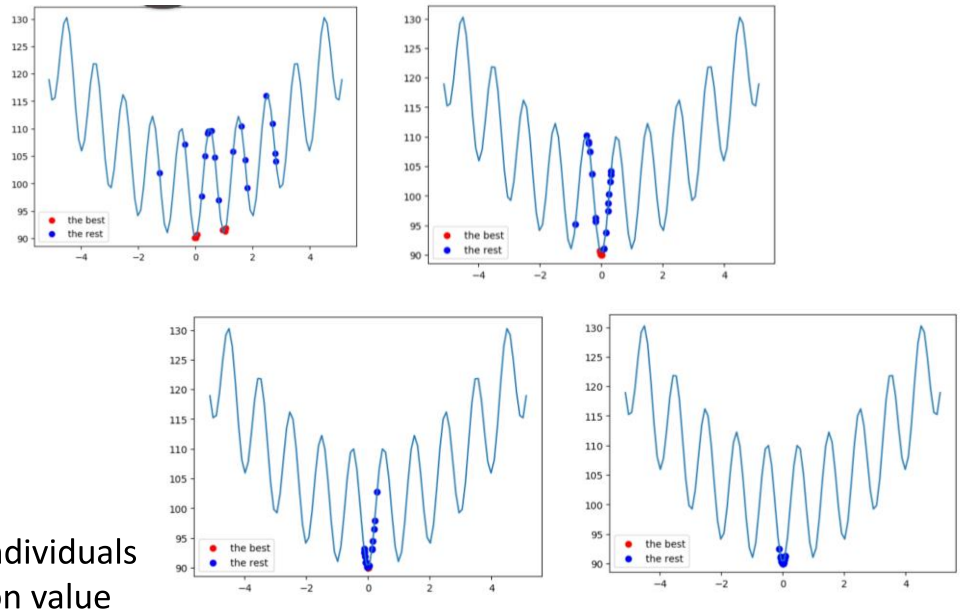



- Genetic Algorithms
- Estimation of Distribution Algorithm (EDA)
  - **Density model-based Evolutionary Strategy (ES)**
- Simulated annealing
- Hill climbing

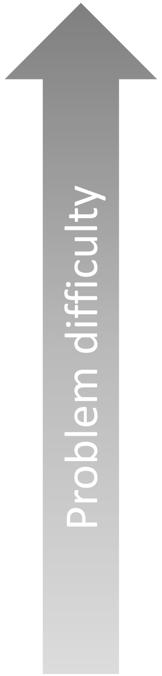

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

# (General ES)




- **Population**: **Randomly** selecting individuals
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- **Perturbing** the fittest to create the next **generation**




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

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

# Genetic Algorithms

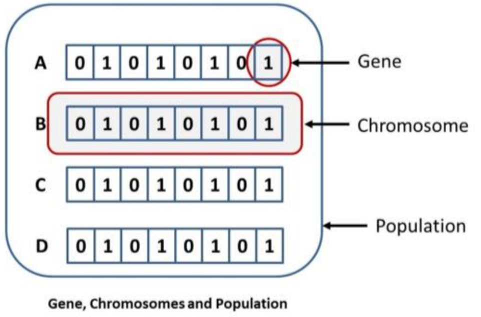



Figure 1.0: Population, Chromosomes and Genes

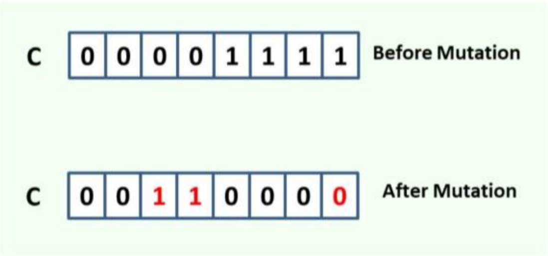



Figure 1.2: Mutation phase illustrated

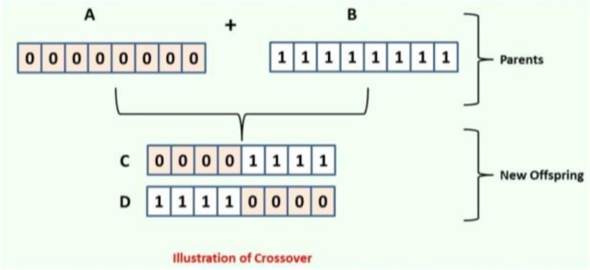
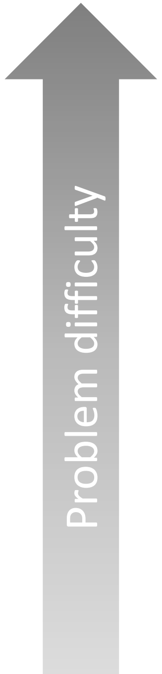



Figure 1.1: Illustration of Cross-over phase

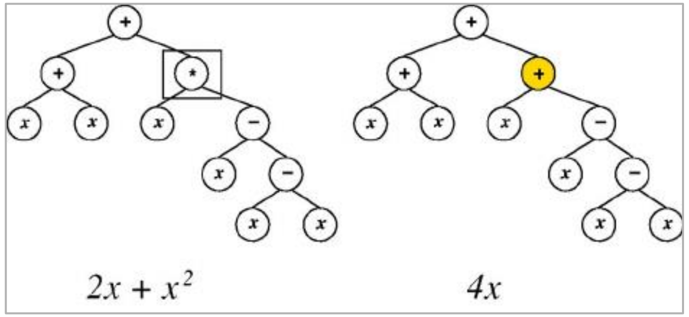
- **Population**: **Randomly** selecting individuals
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the fitter individuals
- **Mutation and Cross-over** the fittest to create the next **generation**



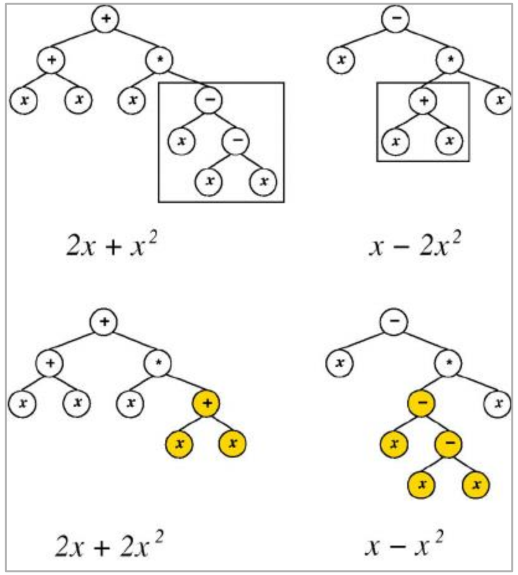

- **Genetic Algorithms**
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- Hill climbing

https://www.kindsonthegenius.com/basics-of-genetic-algorithm-ga-explained-in-simple-terms/

# Genetic Algorithms: Symbolic Regression

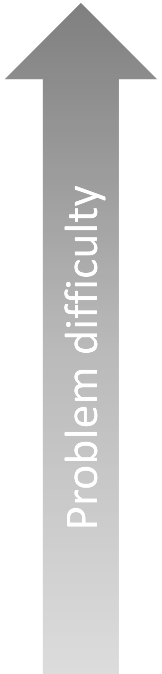



Mutation




Cross-over

- **Population**: **Randomly** selecting individuals
- **Evaluation:** Calculating the function value **(Fitness)** for each individual
- **Selection**: selecting the best individuals
- **Mutation and Cross-over** the fittest to create the next **generation**

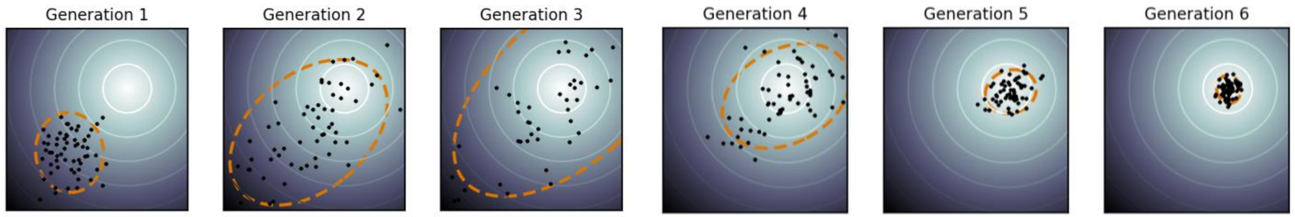



- **Genetic Algorithms**
- Estimation of Distribution Algorithm (EDA)
  - Density model-based Evolutionary Strategy (ES)
- Simulated annealing
- Hill climbing

https://www.researchgate.net/publication/10726448_Dynamics_of_the_evolution_of_learning_algorithms_by_selection?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ

# My Research

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES

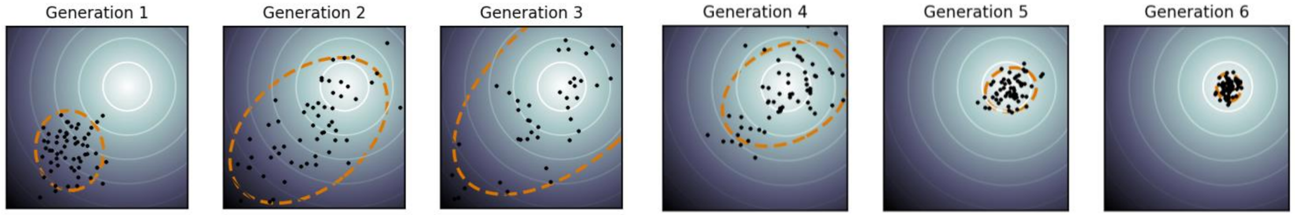


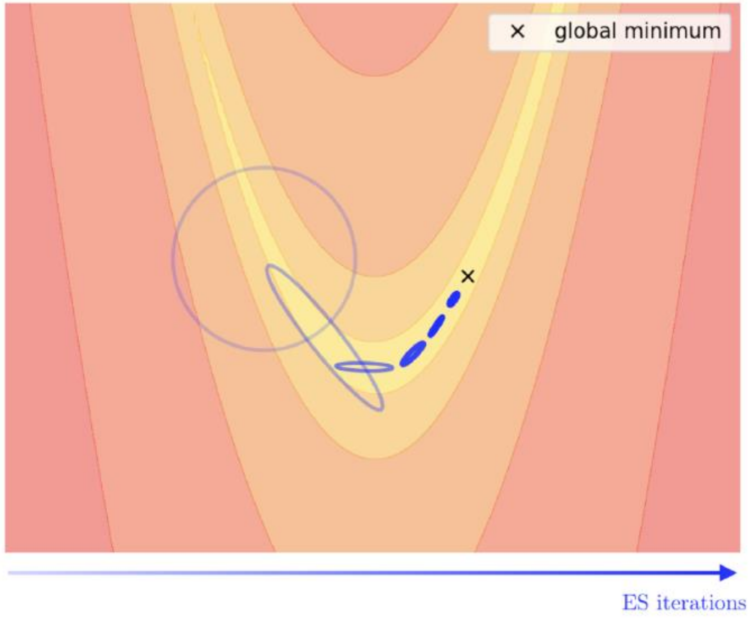

How to make it more sample efficient?

"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.

# My Research

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES



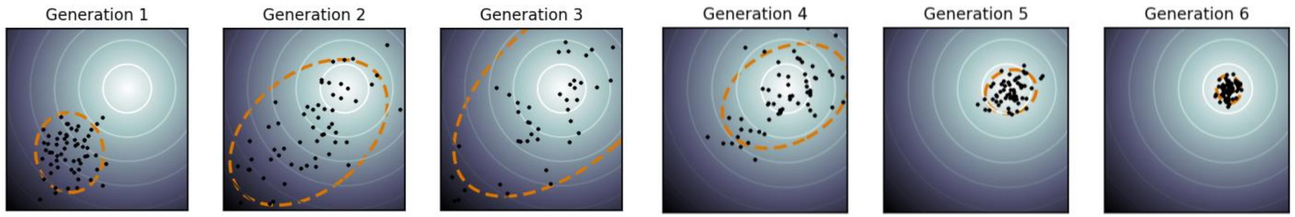

How to make it more sample efficient?

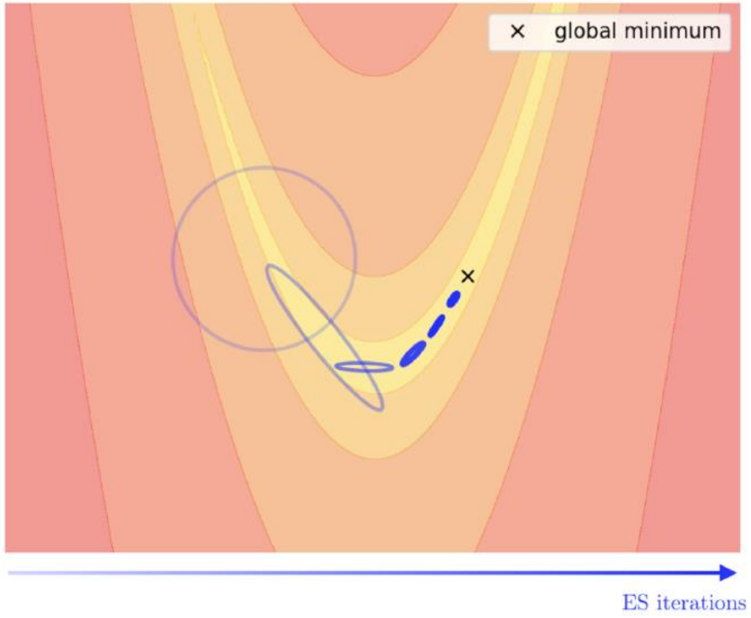
- Arbitrary choice of distribution limits performance

"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.

# My Research

## Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES



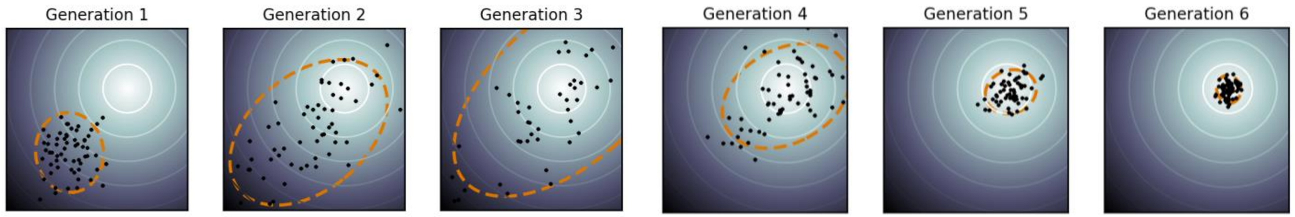

How to make it more sample efficient?

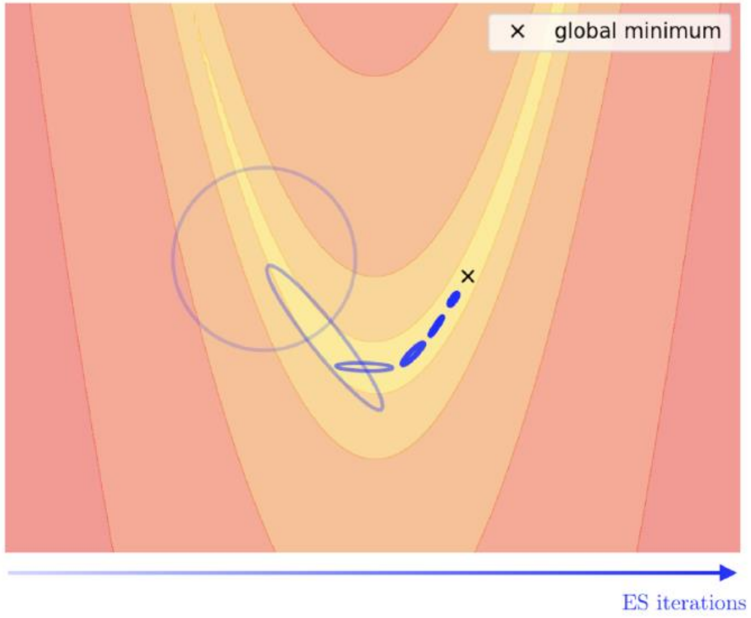
- Arbitrary choice of distribution limits performance
- Starting from scratch and withdrawing gained knowledge


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

# My Research

## Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES




### How to make it more sample efficient?

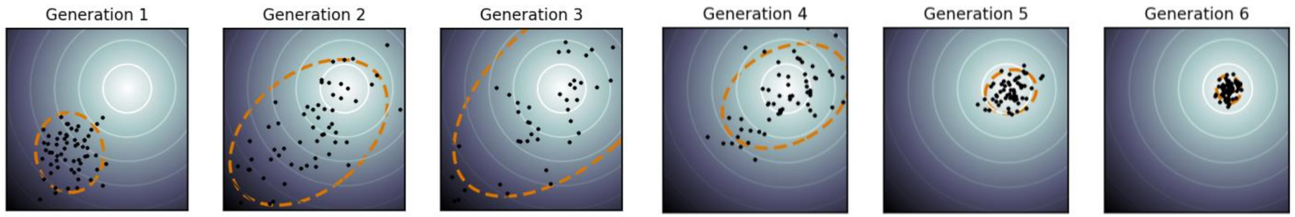
- Arbitrary choice of distribution limits performance ⟶ Flexible density model
- Starting from scratch and withdrawing gained knowledge ⟶ Meta-learning

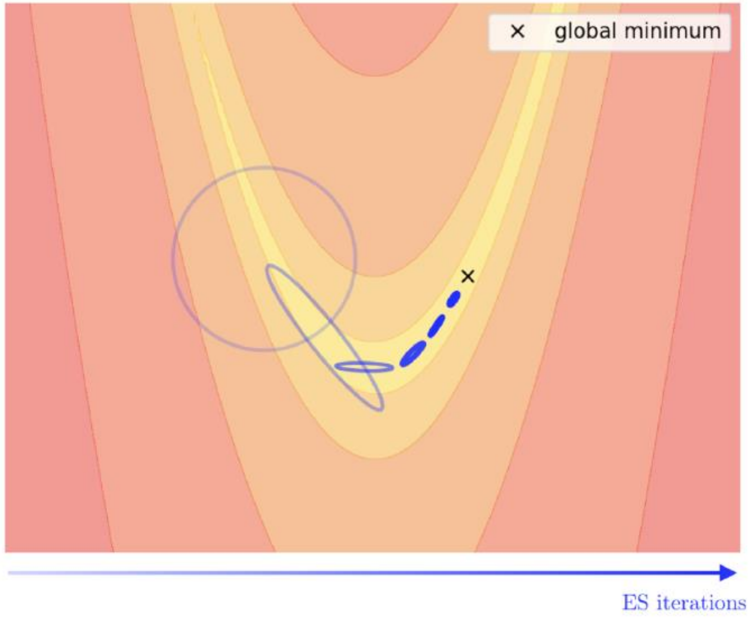

"CMA-ES." wikipedia.org. https ://en.wikipedia.org/wiki/CMAES (accessed Sep. 25,2022)

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

# My Research

## Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES




### How to make it more sample efficient?

- Arbitrary choice of distribution limits performance → Flexible density model
- Starting from scratch and withdrawing gained knowledge → Meta-learning

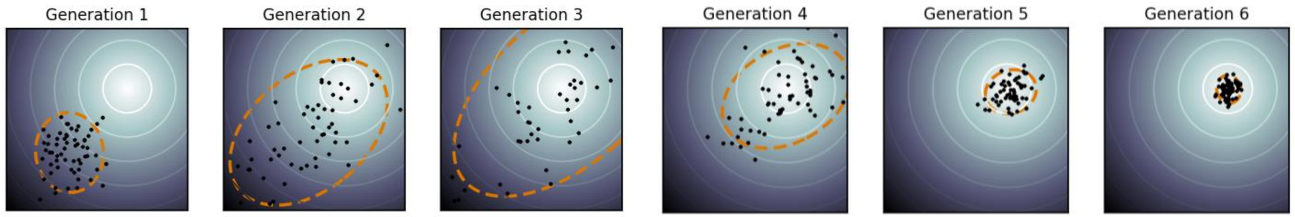
### What we propose

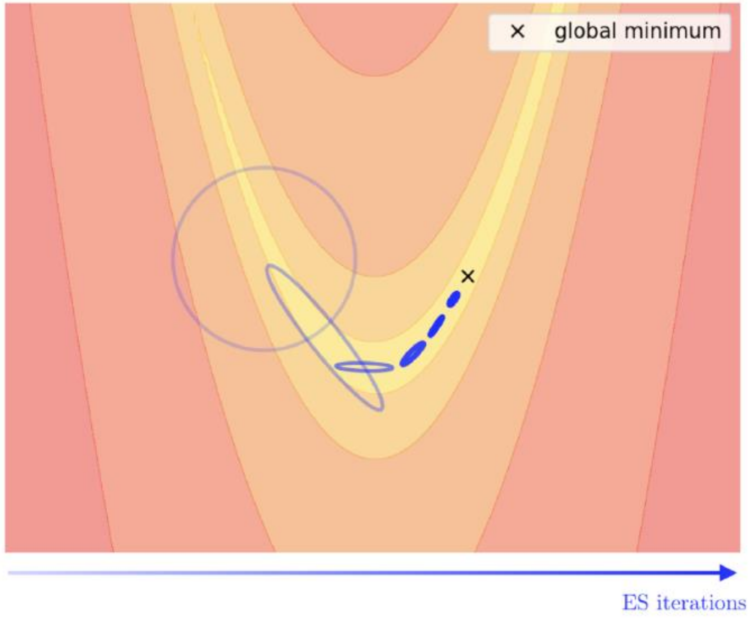
- Normalizing Flows (NF) to model the density
- Meta-optimization to reuse the costly information gained in previous runs


"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.

# My Research

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Example from Density Model-Based ES




How to make it more sample efficient?

- Arbitrary choice of distribution limits performance → Flexible density model
- Starting from scratch and withdrawing gained knowledge → Meta-learning

What we propose

- Normalizing Flows (NF) to model the density
- ~~Meta-optimization to reuse the costly information gained in previous runs~~ 😱


"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

A family of Probabilistic Generative Models: learn $p_x(x)$ over $X$ from observations

# Normalizing Flows

A family of Probabilistic Generative Models: learn $p_x(x)$ over $X$ from observations

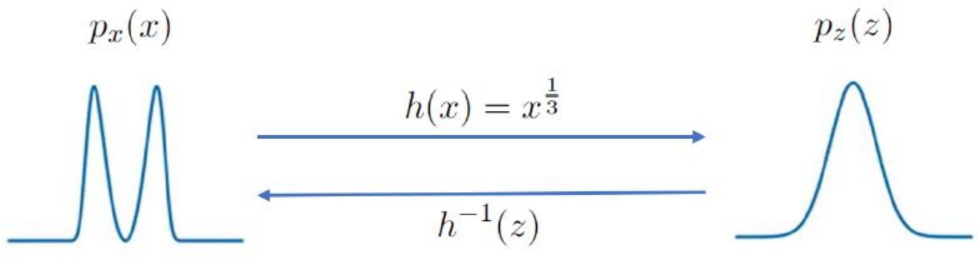
- Structurally restricted Generative Neural Network: bijective Generative NN
- Highly expressive
- Efficient sampling
- Exact and efficient density evaluation
- Based on change of variable formula

# Normalizing Flows

A family of Probabilistic Generative Models: learn $p_x(x)$ over $X$ from observations

- Structurally restricted Generative Neural Network: bijective Generative NN
- Highly expressive
- Efficient sampling
- Exact and efficient density evaluation
- Based on change of variable formula

If you want to know more…




Volume correction

$$p_x(x) = p_z(h(x)).|detDh(x)|$$

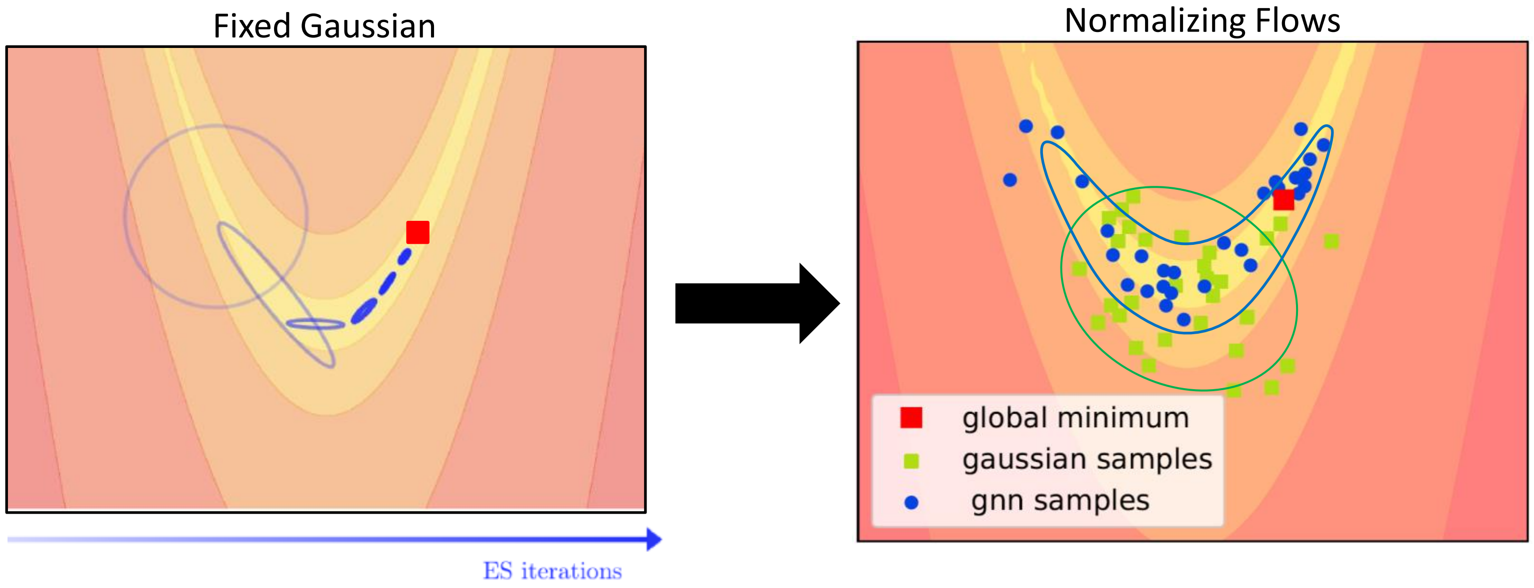
Flow / Bijection

Invertible
Differentiable
Efficiently computable $detDh(x)$

# NF + Density Model-Based ES

Model the density using Normalizing Flows

- More flexible and expressive
- Accelerates the search




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

# Conclusion

- Choice of method depends on the optimization problem
    - We never use BBO and DFO if derivatives exist and are feasible to calculate!

# Conclusion

- Choice of optimization method depends on the problem case
  - We never use BBO and DFO if derivatives exist and are feasible to calculate!

- BBO is the study of algorithms that assume the objectives are given by Black Boxes
  - Function evaluations are **costly** (financially, computationally, time-wise…)
  - General goal is to find a **good enough** solution with **minimum** function evaluations
  - Genetic Algorithms, Evolutionary strategies and their variants are among the most popular and promising methods in this field

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


Schematic of a Black Box [1]

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

# Conclusion

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


Schematic of a Black Box [1]

- Choice of optimization method depends on the problem case
  - We never use BBO and DFO if derivatives exist and are feasible to calculate!
- BBO is the study of algorithms that assume the objectives are given by Black Boxes
  - Function evaluations are **costly** (financially, computationally, time-wise…)
  - General goal is to find a **good enough** solution with **minimum** function evaluations
  - Genetic Algorithms, Evolutionary strategies and their variants are among the most popular and promising methods in this field
- My research is cool! :D

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