Multi-armed Bandit Based Covariance Matrix Adaptation Evolution Strategy

Chuan-Che Yen
Advisor: Dr. Tian-Li Yu

TEILab

Nov 23, 2014

Outline

Outline

Real-valued Function Optimization

- Real-valued function
  - $f\colon \mathcal{S} \subset \mathbb{R}^n \to \mathbb{R}, x \mapsto f(x)$
  - $\mathcal{S}$: search space
  - Elements of $\mathcal{S}$: candidates or solutions
- Optimization
  - $\arg\min_x f(x)$, where x are within given bounds.
  - Maximizing f is equivalent to minimizing $-f$.
- Example
  - $\arg\min_x 2x^3 - 3x^2 - 36x - 14$.
  - Design of aircraft wings.

Black-box Optimization

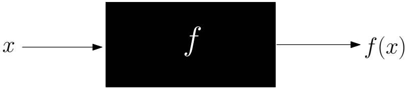


Figure: Black-box function

- The only information is the interaction between input and output.
- The key point is investigating the trade-off between exploration and exploitation.
  - Exploration is the capability of having an overview for the search space.
  - Exploitation is the capability of generating high resolution candidates.

Difficulties

- Non-convex

Difficulties

- Non-convex
  - Local optima are less important for global optimum.

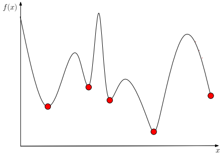


Figure: Non-convex function

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness

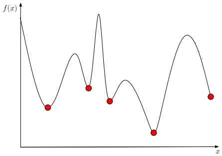


Figure: Non-convex function

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.

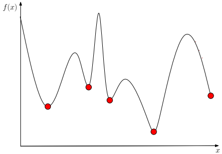


Figure: Non-convex function

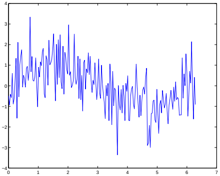


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.

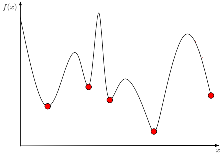


Figure: Non-convex function

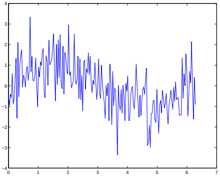


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality

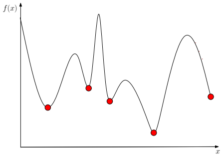


Figure: Non-convex function

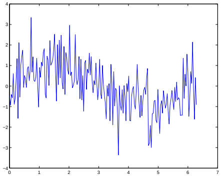


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality
  - Search space grows exponentially

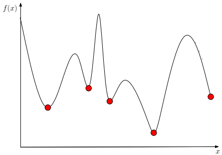


Figure: Non-convex function

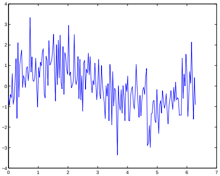


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality
  - Search space grows exponentially
- Non-separable

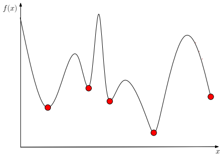


Figure: Non-convex function

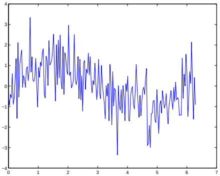


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality
  - Search space grows exponentially
- Non-separable
  - Dependencies between decision variables

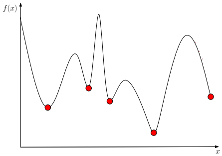


Figure: Non-convex function

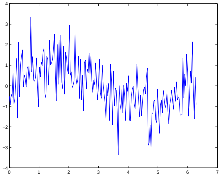


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality
  - Search space grows exponentially
- Non-separable
  - Dependencies between decision variables
- Ill-conditioned

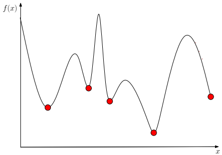


Figure: Non-convex function

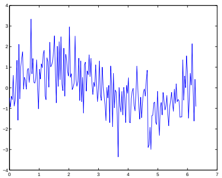


Figure: $sin(x)$ with noise

Difficulties

- Non-convex
  - Local optima are less important for global optimum.
- Ruggedness
  - Perturbated by noise.
  - Non-smooth.
- Dimensionality
  - Search space grows exponentially
- Non-separable
  - Dependencies between decision variables
- Ill-conditioned
  - Unable to extract gradient information

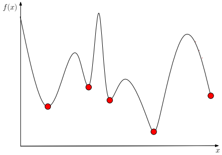


Figure: Non-convex function

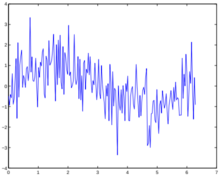


Figure: $sin(x)$ with noise

Outline

1. Real-valued Function Optimization

2. Related Approaches
   - Real-coded Extended Compact Genetic Algorithm
   - Covariance Matrix Adaptation Evolution Strategy

3. Summary

4. Conclusion

Related Approaches

- Optimizing black-box problems
  - No deterministic way to evolve global optimum.
  - Applying stochastic algorithms for approximation.
- Stochastic algorithms
  - Iteratively generating better solutions.
- Two major approaches
  - Estimation of distribution algorithm (EDA).
  - Evolution strategy (ES).

Discretization

- Continuous domain → Discrete domain
- Finding good solutions → Finding promising regions
- 2 traditional discretization methods
    - Fixed Height Histogram (FHH)
    - Fixed Width Histogram (FWH)

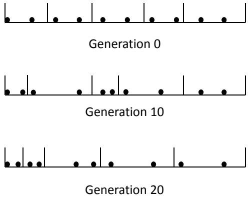


Figure: Illustration of FHH

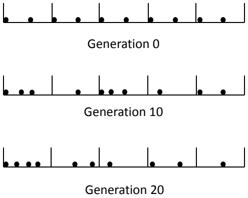


Figure: Illustration of FWH

Split on Demand

- Solutions in each bin should not exceed γN.
  - N is the population size.
  - γ defines the rate of one region.
- γ decays with a factor ϵ.

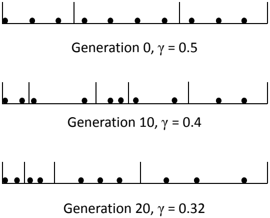


Figure: Illustration of SoD

EDA

- Also known as Probabilistic Model Building GA (PMBGA).

EDA

- Also known as Probabilistic Model Building GA (PMBGA).
  - Building model explicitly.
  - Linkage between decision variables are provided.

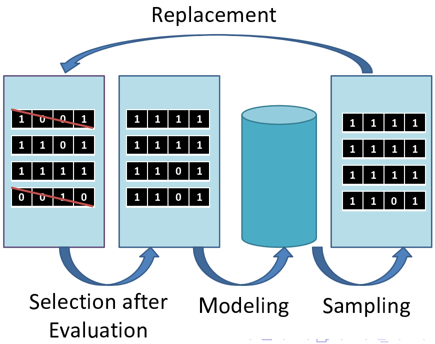
EDA

- Also known as Probabilistic Model Building GA (PMBGA).
  - Building model explicitly.
  - Linkage between decision variables are provided.
- Mechanism difference with traditional GA.
  - Operators 'crossover' and 'mutation' are replaced with 'modeling' and 'sampling'.

EDA

- Also known as Probabilistic Model Building GA (PMBGA).
  - Building model explicitly.
  - Linkage between decision variables are provided.

- Mechanism difference with traditional GA.
  - Operators ‘crossover’ and ‘mutation’ are replaced with ‘modeling’ and ‘sampling’.



Extended Compact Genetic Algorithm (ECGA)

- Each EDA is different from the others in model building.
- ECGA was proposed by Harik (1999).
- Good probabilistic model inspires good linkage learing
  - Model is built according to population distribution.
  - Applying greedy search to refine model iteratively.
- ECGA focuses on bitstring, discrete problems.
  - χ-ECGA.
  - An interface for real-valued function is demanded.

Real-coded ECGA with SoD

Real-coded ECGA with SoD

1. Preparing discretization

Real-coded ECGA with SoD

1. Preparing discretization
2. Integrating discretized results into ECGA.

Real-coded ECGA with SoD

1. Preparing discretization
2. Integrating discretized results into ECGA.
3. ECGA builds model accordingly, output the promising regions.

Real-coded ECGA with SoD

1. Preparing discretization
2. Integrating discretized results into ECGA.
3. ECGA builds model accordingly, output the promising regions.
4. Sampling accordingly.

Real-coded ECGA with SoD

1. Preparing discretization
2. Integrating discretized results into ECGA.
3. ECGA builds model accordingly, output the promising regions.
4. Sampling accordingly.
5. For every L generations, a local optimizer is adopted to obtain high resolution solutions

Real-coded ECGA with SoD

1. Preparing discretization
2. Integrating discretized results into ECGA.
3. ECGA builds model accordingly, output the promising regions.
4. Sampling accordingly.
5. For every L generations, a local optimizer is adopted to obtain high resolution solutions
6. If model does not converge, goto 1.

Evolution Strategy (ES)

- A search template for black-box optimization.
  - Encoded in continuous domain.
- New search points are generated based on current population.
- (μ, λ)-ES and $(\mu + \lambda)$-ES.
- $x_i^{t+1} = m^t + \sigma N_i(0, C)$.
  - x_i^{t+1}: i-th generated solution at generation $t + 1$.
  - m^t: weighted mean of population at generation t.
  - σ: step size.
  - C: Estimated distribution.

Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

- A famous derivation of ES.
- Importance of σ and C.
  - Larger step size reinforces exploration while smaller reinforces exploitation.
    - Choosing an fixed, appropriate number?
  - Covariance matrix determines the shape of estimated distribution.
    - Determining the length of each axis.
    - Representing the dependency among decision variables.
- CMA-ES features in the adoption of historical information.
  - σ and C are adjusted accordingly.

Illustration of σ and C

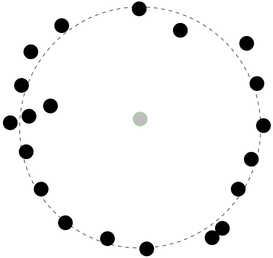


Figure: $t = 0$

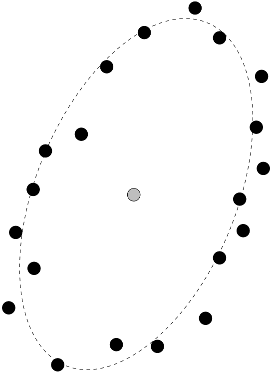


Figure: $t = 10$

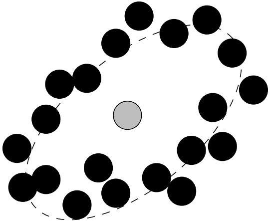


Figure: $t = 20$

Outline

Outline

1. Real-valued Function Optimization
2. Related Approaches
3. Summary
4. Conclusion