Bias–variance tradeoff

A multi-objective approach using MAMaLGaM

Matteo Medioli - Group 10

July 2019

Evolutionary Algorithm
TU Delft

Introduction

Basis of Comparisons

Different analysis:

1. Shared ”resource” Hyperparameters:
    - Max. number of evaluations
    - Elitist archive size
    - Approximation set size
    - Problem dimension (K*3)
2. Optimal Population size
    - Bisection method
3. Algorithm-based hyperparameters
    - Number of mixture components

Hypervolume: metric of comparison

MAMaLGaM: Overview

1. Domination based MO
2. Local Optimizer:
   - Incremental ML approach to estimate **Covariance Matrix** $\hat{\Sigma}$

$$\hat{\boldsymbol{\mu}}(t) = \frac{1}{|S|} \sum_{i=0}^{|S|-1} S_i$$

$$\hat{\Sigma}(t) = \left(1 - \eta^{\boldsymbol{\Sigma}}\right) \hat{\Sigma}(t-1) + \quad \eta^{\boldsymbol{\Sigma}} \frac{1}{|S|} \sum_{i=0}^{|S|-1} \left(\boldsymbol{\mathcal{S}}_i - \hat{\boldsymbol{\mu}}(t)\right) \left(\boldsymbol{\mathcal{S}}_i - \hat{\boldsymbol{\mu}}(t)\right)^T \tag{1}$$

   - Random variable **dependencies** stored in $\hat{\Sigma}$
   - **Offspring**: sampling based on Cholesky Decomposition of $\hat{\Sigma}$

$$\begin{aligned} \hat{\Sigma} &= L_{\hat{\Sigma}} \left(L_{\hat{\Sigma}}\right)^T \\ Z_i &\sim N(0,1) \end{aligned} \tag{2}$$

MAMaLGaM: Overview

1. Domination based MO
2. Local Optimizer:
   - Incremental ML approach to estimate **Covariance Matrix** $\hat{\Sigma}$

$$\hat{\boldsymbol{\mu}}(t) = \frac{1}{|S|}\sum_{i=0}^{|S|-1} S_i$$

$$\hat{\Sigma}(t) = \left(1 - \eta^{\boldsymbol{\Sigma}}\right)\hat{\Sigma}(t-1) + \quad \eta^{\boldsymbol{\Sigma}}\frac{1}{|S|}\sum_{i=0}^{|S|-1}\left(\boldsymbol{\mathcal{S}}_i - \hat{\boldsymbol{\mu}}(t)\right)\left(\boldsymbol{\mathcal{S}}_i - \hat{\boldsymbol{\mu}}(t)\right)^T \tag{1}$$

   - Random variable **dependencies** stored in $\hat{\Sigma}$
   - **Offspring**: sampling based on Cholesky Decomposition of $\hat{\Sigma}$

$$\begin{aligned}\hat{\Sigma} &= L_{\hat{\Sigma}}\left(L_{\hat{\Sigma}}\right)^T \\ Z_i &\sim N(0, 1)\end{aligned} \tag{2}$$

$$X_{new} = \hat{\mu} + L_{\hat{\Sigma}} Z \tag{3}$$

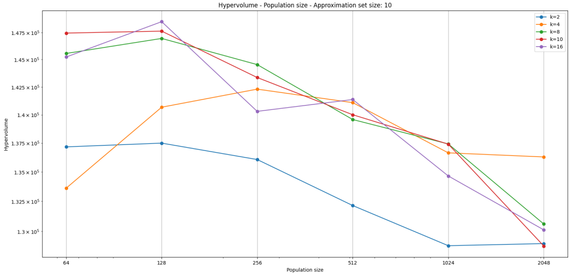
Black-box Optimization Analysis

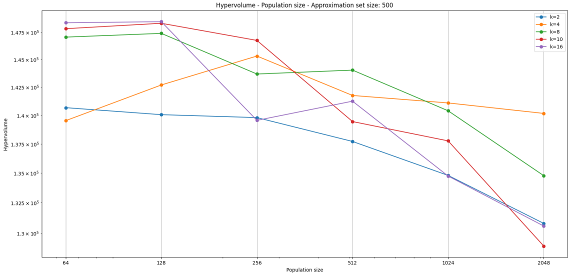
Analyzed hyperparameters

- Approximation set size
- Population size
- Number of mixing components

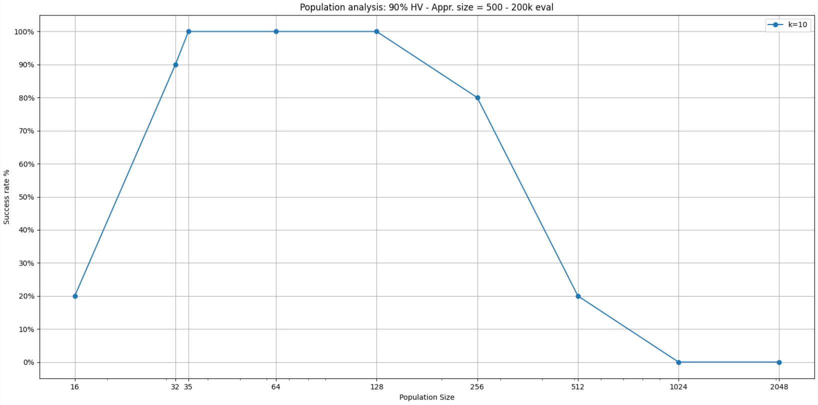
Test runs: 10

Approximation set size

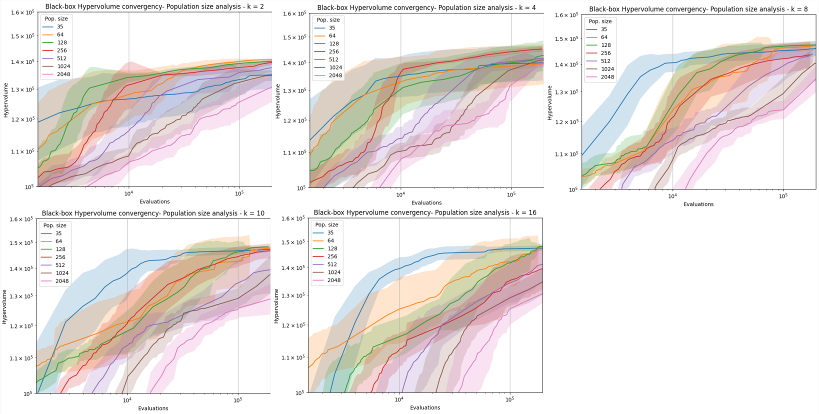




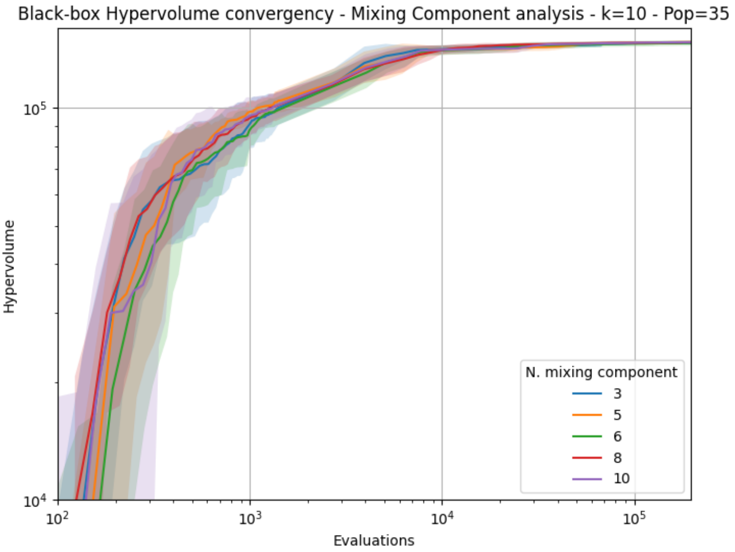
Population size



Optimal Population size: influence of K



Number of mixing components



White-box Optimization: improvements, analysis and results

Worst Basis Identification and Score Function

Main idea: iterative elimination of 3 parameters w_i, μ_i and σ_i related to $basis_i \quad \forall i \in 0, K-1$

$$x = \left(\underbrace{w_0, \mu_0, \sigma_0}_{basis_0} \cdots \underbrace{w_i, \mu_i, \sigma_i}_{basis_i} \cdots \underbrace{w_{K-1}, \mu_{K-1}, \sigma_{K-1}}_{basis_{K-1}} \right) \tag{4}$$

Worst Basis Identification and Score Function

Main idea: iterative elimination of 3 parameters w_i, μ_i and σ_i related to $basis_i \quad \forall i \in 0, K-1$

$$x = \left(\underbrace{w_0, \mu_0, \sigma_0}_{basis_0} \cdots \underbrace{w_i, \mu_i, \sigma_i}_{basis_i} \cdots \underbrace{w_{K-1}, \mu_{K-1}, \sigma_{K-1}}_{basis_{K-1}} \right) \tag{4}$$

$$\forall i \in \{0, K-1\}$$

$$\hat{x}_i = \left(w_0, \mu_0, \sigma_0 \cdots w_{i-1}, \mu_{i-1}, \sigma_{i-1}, w_{i+1}, \mu_{i+1}, \sigma_{i+1} \cdots w_{K-1}, \mu_{K-1}, \sigma_{K-1} \right) \tag{5}$$

Worst Basis Identification and Score Function

Main idea: iterative elimination of 3 parameters w_i, μ_i and σ_i related to $basis_i \quad \forall i \in 0, K-1$

$$x = \Big(\underbrace{w_0, \mu_0, \sigma_0}_{basis_0} \cdots \underbrace{w_i, \mu_i, \sigma_i}_{basis_i} \cdots \underbrace{w_{K-1}, \mu_{K-1}, \sigma_{K-1}}_{basis_{K-1}} \Big) \tag{4}$$

$$\forall i \in \{0, K-1\}$$

$$\hat{x}_i = \Big(w_0, \mu_0, \sigma_0 \cdots w_{i-1}, \mu_{i-1}, \sigma_{i-1}, w_{i+1}, \mu_{i+1}, \sigma_{i+1} \cdots w_{K-1}, \mu_{K-1}, \sigma_{K-1} \Big) \tag{5}$$

$$score \;\; D_i = \frac{f_0(x) - f_0(\hat{x}_i)}{f_0(x)} \tag{6}$$

Worst Basis Identification and Score Function

Main idea: iterative elimination of 3 parameters w_i, μ_i and σ_i related to $basis_i \quad \forall i \in 0, K-1$

$$x = \left(\underbrace{w_0, \mu_0, \sigma_0}_{basis_0} \cdots \underbrace{w_i, \mu_i, \sigma_i}_{basis_i} \cdots \underbrace{w_{K-1}, \mu_{K-1}, \sigma_{K-1}}_{basis_{K-1}} \right) \tag{4}$$

$$\forall i \in \{0, K-1\}$$

$$\hat{x}_i = \left(w_0, \mu_0, \sigma_0 \cdots w_{i-1}, \mu_{i-1}, \sigma_{i-1}, w_{i+1}, \mu_{i+1}, \sigma_{i+1} \cdots w_{K-1}, \mu_{K-1}, \sigma_{K-1} \right) \tag{5}$$

$$score \;\; D_i = \frac{f_0(x) - f_0(\hat{x}_i)}{f_0(x)} \tag{6}$$

The higher the D_i value, the lower the Mean squared error $f0$ evaluated on $\hat{x}_i$. The basis that most negatively affects fitness $f0$ is the missing one in $\hat{x}_i$ with the highest D_i score.

$$i_{worst} = \underset{i \in \{0, K-1\}}{\operatorname{argmax}} \; D_i \tag{7}$$

Worst Basis Mutation

- Two techincs:
  1. Noising parameters with weighted Normal distribution

$$
\begin{aligned}
w_{iworst} += \mathcal{N}(0,1) * s(w_{iworst}) \\
\mu_{i_{worst}} += \mathcal{N}(0,1) * s(\mu_{i_{worst}}) \\
\sigma_{iworst} += \mathcal{N}(0,1) * s(\sigma_{iworst})
\end{aligned}
\tag{8}
$$

  where:

$$
s(\rho_i) = \frac{\sqrt{|\rho_i|} * D_i}{\textbf{varm}} \tag{9}
$$

Worst Basis Mutation

- Two techincs:
  1. Noising parameters with weighted Normal distribution

$$\begin{aligned} w_{iworst} + &= \mathcal{N}(0,1) * s(w_{iworst}) \\ \mu_{i_{worst}} + &= \mathcal{N}(0,1) * s(\mu_{i_{worst}}) \\ \sigma_{iworst} + &= \mathcal{N}(0,1) * s(\sigma_{iworst}) \end{aligned} \tag{8}$$

  where:

$$s(\rho_i) = \frac{\sqrt{|\rho_i|} * D_i}{\text{varm}} \tag{9}$$

  2. Re-sampling of independent standard Gaussian used for offspring

$$\hat{\Sigma} = L_{\hat{\Sigma}} \left(L_{\hat{\Sigma}}\right)^T \tag{10}$$

$$Z_i \sim N(0,1) \tag{11}$$

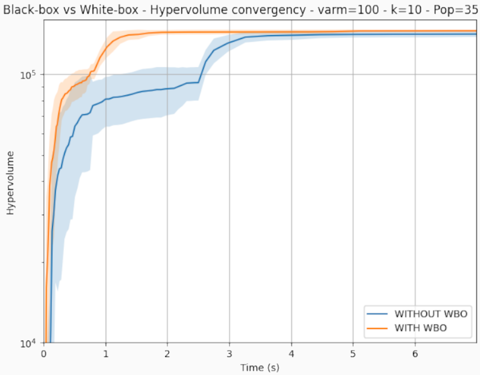
$$X_{new} = \hat{\mu} + L_{\hat{\Sigma}} Z \tag{12}$$

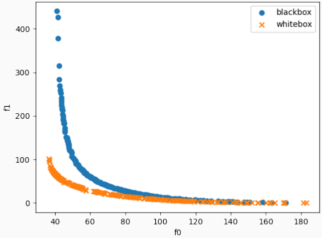
$$\hat{Z} = Z[i_{worst} : i_{worst} + 2] \sim N(0,1) \tag{13}$$

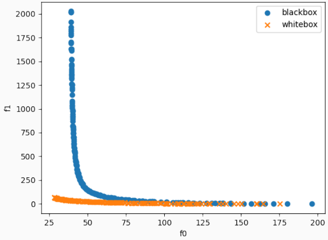
$$\hat{X}_{new} = \hat{\mu} + L_{\hat{\Sigma}} \hat{Z} \tag{14}$$

$$X_{new}[i_{worst} : i_{worst} + 2] = \hat{X}_{new}[i_{worst} : i_{worst} + 2] \tag{15}$$

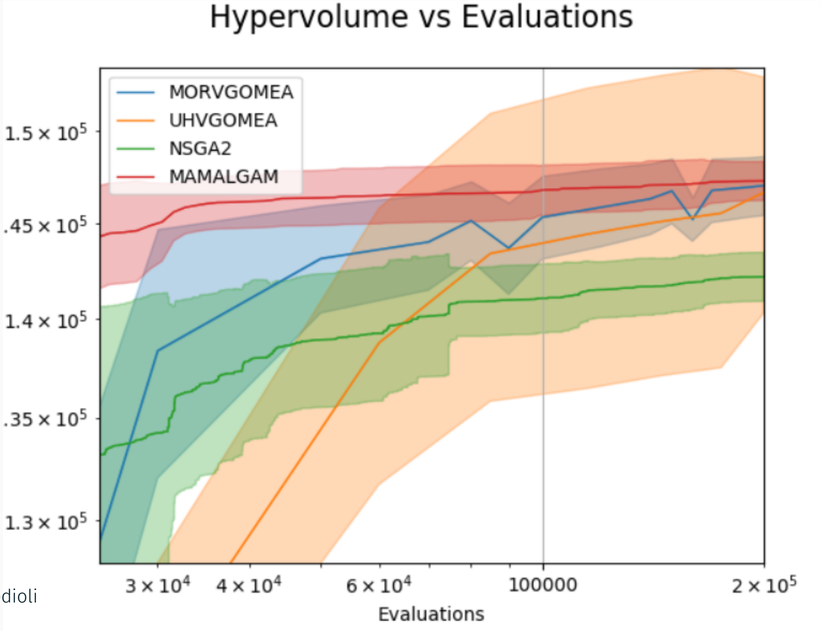
WBO - Results


Black-box vs White-box - Hypervolume convergency - varm=100 - k=10 - Pop=35
Hypervolume
Time (s)
WITHOUT WBO
WITH WBO


blackbox
whitebox
f1
f0


blackbox
whitebox
f1
f0

Comparison - Results



Thank you for your attention