GECCO '21, July 10-14, 2021, Lille, France

A BENCHMARK PROBLEM SET

The problems considered in Section 5 are described in paper [2] and implemented in pure Python except for problems PB2.40 and PB2.41 where we used routines from the numpy library. Also note that in references [2, 15, 18] there is a replication error defining the g_1 constraint function for the Himmelblau (HB) problem: the coefficient attached to the term x_1x_4 is given as 0.00026 whereas it should be 0.006262 [16] and we remind the correct description below:

Problem HB (Himmelblau's nonlinear optimization problem)

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) = 5.3578547x_3^2 + 0.8356891x_1x_5 \\ & + 37.293239x_1 - 40, 792.141 \\ \text{subject to} & 0 \leq g_1(\mathbf{x}) \leq 92 \\ & 90 \leq g_2(\mathbf{x}) \leq 110 \\ & 20 \leq g_3(\mathbf{x}) \leq 25 \\ & 78 \leq x_1 \leq 102 \\ & 33 \leq x_2 \leq 45 \\ & 27 \leq x_i \leq 45 \ (i = 3, 4, 5) \end{array}$$

 $g_1(\mathbf{x}) = 85.334407 + 0.0056858x_2x_5 + 0.006262x_1x_4 - 0.0022053x_3x_5$

$$g_2(\mathbf{x}) = 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2$$

$$g_3(\mathbf{x}) = 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_4$$

B FITTING A LOG-NORMAL DISTRIBUTION TO RUNTIMES DATA

For the active (1+1)-CMA-ES the paper [2] reports the 10-th, 50-th and 90-th quantiles, $\{m_{10}, m_{50}, m_{90}\}$. We propose to fit a model of the entire ECDF which is exact in these 3 quantiles. A convenient model for the runtime of an algorithm is a positive random variable which is the product of a high number of small factors, hence the choice of the Log-normal distribution.

A random variable *X* follows the Log-normal distribution $X \sim \text{Log}\mathcal{N}(\mu, \sigma^2)$ if $Y = \ln X$ follows the normal distribution $Y \sim$

Paul Dufossé and Nikolaus Hansen

 $\mathcal{N}(\mu, \sigma^2)$. We are looking for the (estimated) parameters of the log-normal $\hat{\mu}$, $\hat{\sigma}$.

First fit the median:

$$\hat{\mu} = \ln m_{50} \tag{9}$$

Then solve in σ for the repartition function at the 10-th and 90-th quantiles:

$$\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) = c \tag{10}$$

i.e. solve Equation (10) for $\mu = \hat{\mu}$ and $(x, c) \in \{(m_{10}, 0.1), (m_{90}, 0.9)\}$, where erf is the Gauss error function

The solution is:

$$\hat{\sigma}_c = \frac{1}{\sqrt{2}} (\ln x - \mu) [\text{erf}^{-1}(2c - 1)]^{-1}$$
(11)

where erf^{-1} is the inverse function of erf.

We obtain two values $\hat{\sigma}_{10}$, $\hat{\sigma}_{90}$ which are generally different because the 3 data points do not fit perfectly a log-normal distribution).

If we denote by $\Phi(x, \mu, \sigma)$ the cumulative distribution function of a log-normally distributed random variable with parameters (μ, σ), then our final estimate for the distribution of the runtimes, given { m_{10}, m_{50}, m_{90} }, would be:

$$\mathbb{1}_{\{x \le m_{50}\}} \Phi(x, \hat{\mu}, \hat{\sigma}_{10}) + \mathbb{1}_{\{x > m_{50}\}} \Phi(x, \hat{\mu}, \hat{\sigma}_{90})$$
(12)

C STATISTICAL COMPARISON OF ALGORITHMS DATA

Section 5 provides runtime data for many algorithm variants on several problems in the form of ECDF. Here, we test observed runtime differences for statistical significance. For each problem and each pair of algorithms, we compute the *p*-value from the two-sided Wilcoxon-Mann-Whitney test. If an algorithm fails, its runtime is replaced with $+\infty$, such that it's always worse than a successful run, and comparing two failing runs results in equality. We also apply a Bonferroni correction of $\beta = 8$, that is the number of test problems. If algorithm A is better than algorithm B for a given problem with a *p*-value lower than $0.05/\beta$, we present the item in the corresponding entry, see Table 2.

Augmented Lagrangian, penalty techniques and surrogate modeling for constrained optimization with CMA-ES

GECCO '21, July 10-14, 2021, Lille, France

Algorithm	Quadratic Penalty	Linear Penalty	AL single	AL single old settings	AL many old settings	AL many	MM-AL many	MM-AL many Restricted CP, f -evals	MM-AL many Restricted CP, <i>g</i> -evals	AL many Restricted CP
Quadratic Penalty	x	HB -11.7 G6 -5.1 G7 -15.3 G10 -10.7 2.40 -12.9 2.41 -13.0 TR2 -13.0	HB -14.5 G6 -13.3 G7 -16.6 G9 -7.3 G10 -18.9 2.40 -15.8 2.41 -15.4 TR2 -8.7	HB -7.0 G6 -11.2 G7 -8.4 G9 -6.3 G10 -14.0 2.40 -5.3 2.41 -5.6 TR2 -10.5	G6 -17.1 2.40 -3.6 2.41 -2.4 TR2 -12.2	G6 -10.2 TR2 -11.5	HB -12.7 G6 -16.7 G7 -15.5 G9 -15.6 2.40 -11.9 2.41 -12.1 TR2 -17.2	HB -10.1 G6 -14.0 G7 -11.9 G9 -18.7 2.40 -9.0 2.41 -8.9 TR2 -16.7	HB -16.5 G6 -17.1 G7 -17.4 G9 -19.0 G10 -11.7 2.40 -15.7 2.41 -14.5 TR2 -17.2	HB -9.5 G7 -16.7 G9 -8.3 G10 -15.2 2.40 -7.5 2.41 -8.6 TR2 -13.3
Linear Penalty		x	HB -2.4 G6 -11.7 G7 -4.8 G9 -4.0 G10 -18.1 2.40 -3.4 2.41 -3.8	G6 -6.9 G9 -3.3	G6 -17.2	G6 -4.8	G6 -17.1 G9 -14.8 TR2 -17.2	G6 -12.8 G7 -2.5 G9 -18.3 TR2 -12.6	HB -8.9 G6 -17.2 G7 -17.4 G9 -19.0 TR2 -17.1	G7 - 5.7 G9 - 5.3 G10 -3.5
AL single			Х		G6 -15.9		G6 -7.2 G9 -11.0 TR2 -13.8	G9 -17.7 TR2 -7.7	HB -5.1 G6 -12.3 G7 -16.5 G9 -19.0 TR2 -12.6	
AL single old settings		HB -4.7 G7 -2.9 2.40 -5.1 2.41 -9.0	HB -10.6 G7 -5.5 G10 -17.4 2.40 -10.7 2.41 -13.7	х	G6 -16.9		HB -5.9 G6 -12.6 G7 -2.8 G9 -8.4 2.40 -3.5 2.41 -6.5 TR2 -17.0	G7 -4.8 G9 -17.6 TR2 -11.3	HB -16.1 G6 -15.9 G7 -15.0 G9 -18.7 2.40 -9.5 2.41 -12.7 TR2 -16.7	G7 -5.8
AL many old settings	G7 -17.1	HB -10.6 G7 -17.2 2.40 -10.0 2.41 -13.4	HB -15.2 G7 -17.2 G9 -6.0 G10 -9.0 2.40 -14.9 2.41 -16.0	HB -4.5 G7 -17.2 G9 -5.4 2.41 - 2.5	х	G7 -2.5	HB -13.0 G7 -17.2 G9 -15.2 2.40 -9.2 2.41 -12.1 TR2 -17.2	HB -9.0 G7 -17.4 G9 -18.5 2.40 -4.7 2.41 -7.6 TR2 -12.4	HB -17.0 G7 -17.4 G9 -19.0 2.40 -15.4 2.41 -15.3 TR2 -17.1	HB -8.2 G7 -17.2 G9 -6.8 2.40 -3.0 2.41 -6.5
AL many	G7 - 16.9 G10 - 11.4	HB -13.9 G7 -17.2 G10 -12.5 2.40 -14.1 2.41 -15.3	HB -16.5 G6 -4.5 G7 -17.2 G9 -8.8 G10 -18.1 2.40 -16.4 2.41 -16.6	HB -9.2 G7 -17.1 G9 -7.6 G10 -13.0 2.40 -5.3 2.41 -6.7	G6 -17.0 2.40 -3.8 2.41 -2.7	х	HB -15.5 G6 -14.1 G7 -17.2 G9 -16.3 G10 -11.2 2.40 -13.2 2.41 -14.6 TR2 -17.0	HB -12.6 G6 -4.5 G7 -17.4 G9 -18.6 G10 -11.0 2.40 -10.8 2.41 -11.2 TR2 -8.9	HB -17.2 G6 -16.5 G7 -17.4 G9 -19.0 G10 -12.8 2.40 -16.7 2.41 -16.4 TR2 -16.4	HB -12.3 G7 -17.2 G9 -10.4 G10 -13.4 2.40 -8.0 2.41 -10.5
MM-AL many	G10 -12.1	G10 -16.7	HB -3.3 G7 -5.7 G10 -18.9 2.40 -7.4 2.41 -7.4	G10 -17.1	G6 -13.5		x	G7 -2.6 G9 - 16.0	HB -12.3 G6 -3.8 G7 -17.4 G9 -17.4 G10 -16.8 2.40 -5.1 2.41 -4.6	G7 -6.5 G10 -17.1
MM-AL many Restricted CP, <i>f</i> -evals	G10 -8.4	G10 -13.0 2.40 -3.1 2.41 -6.5	HB -5.5 G10 -17.5 2.40 -10.3 2.41 -12.3	G10 -13.5	G6 -8.0		G6 -5.7 2.41 -3.2 TR2 -9.0	х	HB -11.9 G6 -7.8 G10 -13.3 2.40 -8.5 2.41 -11.2 TR2 -7.0	G10 -13.5
MM-AL many Restricted CP, <i>g</i> -evals			G10 -16.7		G6 -5.5				Х	
AL many Restricted CP		HB -2.7 G6 -4.6 2.40 -4.1 2.41 -4.2	HB -8.7 G6 -14.3 G10 -17.3 2.40 -11.0 2.41 -10.0	G6 -12.1	G6 -17.2	G6 - 10.8	HB -3.6 G6 -17.2 G9 -11.6 2.40 -2.6 TR2 -16.8	G6 - 15.3 G9 - 17.8 TR2 - 7.8	HB -15.6 G6 -17.2 G7 -17.4 G9 -19.0 2.40 -9.4 2.41 -7.7 TR2 -16.0	х

Table 2: Comparing the different algorithms' on the proposed benchmark test set. If the test is significant, meaning algorithm in row is better than algorithm in column ($p < 0.05/\beta$) for a test problem, we display the problem name and $\log_{10} p$. If $p < 10^{-3}/\beta$, it is displayed as bold text.