Parameter Tuned CMA-ES on the CEC'15 Expensive Problems

Martin Andersson, Sunith Bandaru, Amos H.C. Ng, Anna Syberfeldt
School of Engineering Science,
University of Skövde, Skövde, Sweden
{martin.andersson,sunith.bandaru,amos.ng,anna.syberfeldt}@his.se

***Abstract*—Evolutionary optimization algorithms have parameters that are used to adapt the search strategy to suit different optimization problems. Selecting the optimal parameter values for a given problem is difficult without *a-priori* knowledge. Experimental studies can provide this knowledge by finding the best parameter values for a specific set of problems. This knowledge can also be constructed into heuristics (rule-of-thumbs) that can adapt the parameters for the problem. The aim of this paper is to assess the heuristics of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimization algorithm. This is accomplished by tuning CMA-ES parameters so as to maximize its performance on the CEC'15 problems, using a bilevel optimization approach that searches for the optimal parameter values. The optimized parameter values are compared against the parameter values suggested by the heuristics. The difference between specialized and generalized parameter values are also investigated.**

I. Introduction

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [1], [2] is a single objective evolutionary optimization algorithm that for certain problems outperforms other evolutionary algorithms, like genetic algorithms, differential evolution and particle swarm optimization [3], [4], [5]. CMA-ES generates new search points by sampling them from a multi-variate normal distribution. A normal distribution is determined by its mean $m \in \mathbb{R}$, standard deviation $\sigma \in \mathbb{R}$ and its covariance matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$. By modifying the covariance matrix the search distribution is made to fit the contour lines of the objective function, thereby increasing the probability of generating good solutions.

The no free lunch theorem [6] states that no optimization algorithm can be better than all other algorithms on all problems. Optimization algorithms try to circumvent this fact by using parameters that can be tweaked to alter its search behavior. Experimental studies can be used to find the appropriate parameter values for a given problem and optimization algorithm. However, this is both difficult and time-consuming, since many experiments are required to obtain reliable results. Another approach is to use existing heuristics or rules of thumb to estimate good parameters for new problems. This is less computationally expensive, since it does not require any experiments. There is, however, no guarantee that the selected parameters will actually work well for that problem.

CMA-ES has heuristics for all of its parameters. They estimate parameter values mostly based on the number of dimensions of the problem. The heuristics are derived from both empirical studies and inherent properties of CMA-ES. They are also designed to be effective over a diverse set of problems. The aim of this paper is to investigate how good these heuristics are at selecting the optimal parameter values for the CEC'15 expensive problems [7]. Table I provides a summary of the included functions. A bilevel optimization approach will be used to search for the parameters that maximizes the performance of CMA-ES on the CEC'15 problems. This will allow for a comparison of the difference in performances between the optimized and default (heuristic suggested) parameter values.

Another important aspect of parameter tuning regards generalized and specialized parameter values. Generalized parameter values are those that are meant to work across many different problems, while specialized parameter values are fine-tuned against a small set of problems. This is important because the optimal parameter values of an optimization algorithm can be

TABLE I. Summary of CEC'15 expensive test problems.

Categories	No	Functions	Related Basic Functions	F_i^*
Unimodal Functions	1	Rotated Bent Cigar Function	Bent Cigar Function	100
	2	Rotated Discus Function	Discus Function	200
Simple Multimodal Functions	3	Shifted and Rotated Weierstrass Function	Weierstrass Function	300
	4	Shifted and Rotated Schwefel's Function	Schwefel's Function	400
	5	Shifted and Rotated Katsuura Function	Katsuura Function	500
	6	Shifted and Rotated HappyCat Function	HappyCat Function	600
	7	Shifted and Rotated HGBat Function	HGBat Function	700
	8	Shifted and Rotated Expanded Griewank's plus Rosenbrock's Function	Griewank's Function Rosenbrock's Function	800
	9	Shifted and Rotated Expanded Scaffer's F6 Function	Expanded Scaffer's F6 Function	900
Hybrid Functions	10	Hybrid Function 1 (N=3)	Schwefel's Function Rastrigin's Function High Conditioned Elliptic Function	1000
	11	Hybrid Function 2 (N=4)	Griewank's Function Weierstrass Function Rosenbrock's Function Scaffer's F6 Function	1100
	12	Hybrid Function 3 (N=5)	Katsuura Function HappyCat Function Griewank's Function Rosenbrock's Function Schwefel's Function Ackley's Function	1200
Composite Functions	13	Composite Function 1 (N=5)	Rosenbrock's Function High Conditioned Elliptic Function Bent Cigar Function Discus Function High Conditioned Elliptic Function	1300
	14	Composite Function 2 (N=3)	Schwefel's Function Rastrigin's Function High Conditioned Elliptic Function	1400
	15	Composite Function 3 (N=5)	HGBat Function Rastrigin's Function Schwefel's Function Weierstrass's Function High Conditioned Elliptic Function	1500

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

quite different between problems and also between specialized and generalized parameter values [8]. Generalized parameter values are often more useful for the practitioner because they are designed to be applicable to a wide range of problems.

This paper analyzes both specialized and generalized CMA-ES parameter values, both in terms of performance and the differences of the optimal parameter values. To find specialized parameter values each problem and dimension is optimized individually, while generalized parameter values are obtained by searching for the optimal parameter values across all problems and dimensions.

It is possible to distinguish three layers in parameter tuning: The application layer, the algorithm (lower) layer and the design (upper) layer [9]. The problem to be solved is located on the application layer and the metaheuristic to solve that problem is on the algorithm layer. On the design layer is the parameter tuner that tests different parameters for the metaheuristic on the algorithm layer. To avoid confusion, the quality of solutions for the problem on the application layer is called *fitness* while the quality of the parameters in the design layer is called *utility* [9].

Parameter tuning can itself be viewed as an optimization problem in which the objective is to find the parameter values that give the best performance on a particular problem or a set of problems. This approach can be referred to as meta-EA [9] or bilevel optimization [10]. In this paper, the objective at the upper level optimization is the same as the CEC'15 expensive problems scoring method. The scoring method is the summation of the mean and median of the best function values over multiple runs of the 15 test problems, for both 10 and 30 dimensions, as shown below:

$$\begin{array}{ll} \underset{\mathbf{p}}{\text{Minimize}} & \sum_{i=1}^{15} mean(f_i(\mathbf{p}))\Big|_{D=10} + \\ & \sum_{i=1}^{15} mean(f_i(\mathbf{p}))\Big|_{D=30} + \\ & \sum_{i=1}^{15} median(f_i(\mathbf{p}))\Big|_{D=10} + \\ & \sum_{i=1}^{15} median(f_i(\mathbf{p}))\Big|_{D=30} \\ \text{where,} & f_i(\mathbf{p}) \text{ is the best function value obtained by} \\ & \text{solving the } i\text{th CEC'15 problem of the following} \\ & \text{form with parameters } \mathbf{p} \\ & \quad \underset{\mathbf{x}}{\text{Minimize}} \quad f(\mathbf{x}) \\ & \quad \text{Subject to} \quad \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \end{array} \tag{1}$$

The algorithmic parameters of the lower-level optimization problem become the variables for the upper-level optimization problem. The objective for each test problem is calculated according to the following equation, where MaxFEs is the maximum number of function evaluations allowed for each problem.

$$f(x) = 0.5 * (f_{MaxFEs} + f_{0.5\times MaxFEs}) \tag{2}$$

The CEC'15 problems are called expensive problems because they only allow for small number of function evaluations, 500 and 1500 evaluations for 10 and 30 dimensions respectively. Most studies use far more function evaluations (in [3] the maximum was set at 10^7) when comparing optimization algorithm performances. The size of the function evaluation budget will most probably affect the optimal parameters. This aspect is however not addressed in this paper.

The rest of the paper is organized as follows. Section II introduces CMA-ES and its parameters. In Section III a description of the experimental design is provided. The experimental results appear in Section IV. The conclusions are summarized in Section V.

II. CMA-ES

There are several variants of CMA-ES. The one used in this paper is the $(\mu/\mu_w, \lambda)$-CMA-ES. Here λ is the population size, μ is the number of selected search points and μ_w indicates that the new search points are weighted when updating the mean.

New search points are generated from a multi-variate normal distribution. They are evaluated and ranked according to their fitness. The best μ of all λ points are weighted and summed to form the new mean. Instead of only using the selection information from a single generation, CMA-ES utilizes the path taken by the population over a number of generations. This is called the evolution path. The covariance matrix is updated by the evolution path and the μ weighted difference vectors of previous and new search points. The step size σ is also updated using an evolution path. Reliably estimating the covariance matrix from a single generation is not always possible. That is why information from previous generations are also added, this is called the rank-μ-update. For a complete description of CMA-ES, see [1]. The CMA-ES implementation used in this paper is based on Hansen's C code [11].

A. Parameters and Heuristics

CMA-ES has strategy parameters that can be used to control its search behavior. This section will provide a short description of them and present the heuristics that are used to set them. The heuristics are also based on those found in Hansen's C code [11]. In the following equations, N refers to the dimension of the problem.

1) λ: The number of new search points generated in each generation.

$$\lambda = 4 + \lfloor 3 * \ln(N) \rfloor \tag{3}$$

2) μ: The number of search points that will form the new mean. Better points are given more significance using weights given by Equation (5).

$$\mu = \left\lfloor \frac{\lambda}{2} \right\rfloor \tag{4}$$

$$w_i = \ln(\mu + 1) - \ln(i); \; w_i = \frac{w_i}{\sum_{i=1}^{\mu} w_i} \tag{5}$$

3) c_σ: The length of the evolution path horizon. This controls the learning rate for the cumulation of the step size.

$$c_\sigma = \frac{\mu_{eff} + 2}{N + \mu_{eff} + 3} \tag{6}$$

$$\mu_{eff} = \left(\sum_{i=1}^{\mu} {w_i}^2 \right)^{-1} \tag{7}$$

4) c_c: The length of the evolution path horizon. This controls the learning rate for the cumulation of the rank-one update of the covariance matrix.

$$c_c = \frac{4}{N + 4} \tag{8}$$

5) c_{cov}: The learning rate for the covariance matrix.

$$c_{cov} = \frac{1}{\mu_{cov}} * t_1 + \left(1 - \frac{1}{\mu_{cov}}\right) * t_2 \tag{9}$$

$$\mu_{cov} = \mu_{eff} \tag{10}$$

$$t_1 = \frac{2}{(N + \sqrt{2})^2} \tag{11}$$

$$t_2 = \min\left(\frac{2 * \mu_{eff} - 1}{(N+2) * (N+2) + \mu_{eff}}, 1 \right) \tag{12}$$

6) $\sigma^{(0)}$: The initial step size. The step size is problem dependent, but the optimum of the optimized function should fall within $m^{(0)} \pm 2\sigma^{(0)}$. With decision variables scaled between $[0, 10]$ and random $m^{(0)}$, the initial step size is chosen to be 2.

$$\sigma^{(0)} = 2 \tag{13}$$

7) d_σ: Dampening for the step size update.

$$d_\sigma = t_1 * \max\left(0.3, 1 - \frac{N}{1e^{-6} + \frac{MaxFEs}{\lambda}} \right) + c_\sigma \tag{14}$$

$$t_1 = 1 + 2 * \max\left(0, \sqrt{\left(\frac{\mu_{eff} - 1}{N + 1} \right)} - 1 \right) \tag{15}$$

III. Experimental Design

Each experiment in this study has a set $\mathbf{t}$ that either include one or all test problems. Experiments with only one test problem find specialized parameter values and experiments with all problems find generalized parameter values. In total, there are 31 experiments, 15 test problems × 2 dimensions for the specialized experiments and 1 experiment for the generalized experiment.

The bilevel optimization approach to parameter tuning only provides a single optimized parameter set $\mathbf{p}^*$ for each experiment. To get a larger sample size to draw conclusions from, each experiment is independently replicated 20 times. The outcome of each replication is the set of parameter values with the best objective value, as measured by Equation (2). Therefore, each experiment produces 20 sets of optimized parameter values. Within each experiment, every evaluated parameter set $\mathbf{p}$ is independently replicated 20 times. The average fitness from these optimization is then used as the utility of that parameter set.

The experiments were run on three Dell PowerEdge R420 servers, each with two Intel Xeon E5-2400 V2 processors for a total of 72 logical cores. The optimization runs were distributed across the servers using an optimization framework written in C++. The framework has the capability of distributing and running independent optimizations in parallel, allowing for an efficient use of all the available computing resources. The experiments took 85 hours to complete.

A. Design Layer

CMA-ES is used to solve the single objective minimization problem, Equation (1), at the design layer. There are two main reasons for choosing CMA-ES. It is known to be an effective single objective optimization algorithm and it has default values for most parameters [3], [4]. λ is set to 10 which is slightly larger than the default. All parameters at the design layer are scaled so they fall in the range $[0, 10]$ and because of that $\sigma^{(0)}$ is set to 2. All other parameters use their default values as described previously.

The maximum iterations allowed at the design layer is 6000, with restarts at iteration 2000 and 4000. Restarts are used to reduce the probability of CMA-ES getting trapped in a local optimum. Initial experiments showed that improvements had plateaued by 2000 iterations, which is why each restart was separated by to that number of iterations.

B. Algorithm Layer

The algorithm (lower) layer is where the utility of a parameter set $\mathbf{p}$ is evaluated. The evaluation is performed by starting instances of CMA-ES against a set $\mathbf{t}$ containing one or more of the CEC'15 problems. The composition of the set $\mathbf{t}$ is different for each experiment. For every problem in the set $\mathbf{t}$ 20 instances of CMA-ES are started with the parameter set that is being evaluated. The utility of the parameter set $\mathbf{p}$ is then calculated as the sum of all mean and median fitnesses obtained by these optimizations, as shown in the objective function in Equation (1).

There are 15 functions included in the CEC'15 expensive problems. Each function is optimized for 10 and 30 dimensions, which are allowed a maximum of 500 and 1500 function evaluations. The decision variables are scaled so that they fall in the range $[0, 10]$; they are originally in the range $[-100, 100]$. The fitness for a test problem is calculated according to Equation (2).

C. Starting Positions

Each experiment replication is assigned a set of 20 (one for each lower level replication) random starting positions for the initial mean. That means that CMA-ES instances within a replication use the same set of starting positions for evaluating any given parameter set $\mathbf{p}$, and that the starting positions are different between replications. Every parameter set $\mathbf{p}$ is evaluated against all starting positions in the assigned set.

By evaluating each parameter set $\mathbf{p}$ against different starting positions, the probability of specifically optimizing the parameters for a particular starting position is reduced. Using the same set of starting positions for all evaluations within a replication reduces the effect of the starting position from the performance of a particular parameter set. An alternative to statically assigning starting positions would be to randomly generate them at the start of each optimization. That approach avoids the issue of sub-optimizing for a particular starting position, but it also makes the comparison between parameters sets unfair since they do not have the same starting conditions.

D. Parameters

The following CMA-ES parameters are tuned in this study.

1) λ: The population size: an integer in the range $[2, 300]$.

2) μ: The number of selected search points as a percentage of the population size: a real-value in the range $[0, 1]$.

3) c_σ: Learning rate for the cumulation of the step size: a real-value in the range $]0, 1]$.

4) c_c: Learning rate for the cumulation of the rank-one update: a real-value in the range $]0, 1]$.

5) c_{cov}: Learning rate for the covariance matrix update: a real-value in the range $[0, 1[$.

6) $\sigma^{(0)}$: The initial step size: a real-value in the range $[0, 10]$.

7) d_σ: Dampening parameter for step size update: a real-value in the range $[0, 10]$.

The parameters of the optimization on the algorithm layer in Equation (1) become variables for the optimization on the design layer. Thus the variable vector $\mathbf{p}$ in Equation (1) is $\mathbf{p} = \{\lambda, \mu, c_\sigma, c_c, c_{cov}, \sigma^{(0)}, d_\sigma\}$.

IV. Experimental Results

The results are divided into two sections, the first section consists of the results of the parameter tuning and it is followed by a comparison of the optimized versus the default parameter values. The parameter tuning experiments use Equation (2) to calculate the fitness, while the replicated generalized and default parameter values use Equation (16).

$$f(x) = f_{MaxFEs} \tag{16}$$

TABLE II. Computational complexity for default ($\hat{T1}$) and optimized ($\hat{T2}$) parameter values.

N	$T0$	$\hat{T1}$	$\hat{T2}$	$\hat{T1}/T0$	$\hat{T2}/T0$
10	0.03	5.9	6.0	196.6	200.0
30	0.03	35.8	39.9	1193.3	1330

TABLE III. Best, median and mean results from 20 parameter tuning experiments for specialized parameters (CMAES-S) and generalized parameters (CMAES-G).

		CMAES-S			CMAES-G			
Func	N	Best	Median	Mean	Best	Median	Mean	Min
1	10	1.077E+07	7.440E+07	3.662E+07	3.820E+07	6.722E+07	6.599E+07	200
	30	4.646E+07	7.284E+07	6.870E+07	7.013E+07	1.689E+08	1.108E+08	200
2	10	5.295E+04	5.880E+04	5.808E+04	7.093E+04	1.021E+05	1.024E+05	400
	30	2.258E+05	2.359E+05	2.363E+05	2.786E+05	2.994E+05	2.953E+05	400
3	10	6.103E+02	6.118E+02	6.120E+02	6.129E+02	6.158E+02	6.157E+02	600
	30	6.318E+02	6.351E+02	6.339E+02	6.472E+02	6.543E+02	6.527E+02	600
4	10	2.602E+03	3.183E+03	3.189E+03	3.437E+03	4.087E+03	4.109E+03	800
	30	7.975E+03	8.750E+03	8.673E+03	9.802E+03	1.235E+04	1.204E+04	800
5	10	1.001E+03	1.001E+03	1.001E+03	1.005E+03	1.006E+03	1.006E+03	1000
	30	1.000E+03	1.001E+03	1.001E+03	1.004E+03	1.008E+03	1.008E+03	1000
6	10	1.201E+03	1.201E+03	1.201E+03	1.201E+03	1.201E+03	1.201E+03	1200
	30	1.201E+03	1.201E+03	1.201E+03	1.201E+03	1.202E+03	1.201E+03	1200
7	10	1.401E+03	1.402E+03	1.401E+03	1.401E+03	1.402E+03	1.402E+03	1400
	30	1.401E+03	1.401E+03	1.401E+03	1.401E+03	1.401E+03	1.401E+03	1400
8	10	1.610E+03	1.613E+03	1.613E+03	1.615E+03	1.947E+03	1.648E+03	1600
	30	1.682E+03	1.834E+03	1.767E+03	1.812E+03	3.884E+04	2.321E+03	1600
9	10	1.808E+03	1.808E+03	1.808E+03	1.808E+03	1.808E+03	1.808E+03	1800
	30	1.827E+03	1.827E+03	1.827E+03	1.828E+03	1.828E+03	1.828E+03	1800
10	10	1.163E+05	1.765E+05	1.744E+05	9.096E+05	1.852E+06	1.770E+06	2000
	30	2.660E+06	3.650E+06	3.631E+06	1.148E+07	1.580E+07	1.473E+07	2000
11	10	2.212E+03	2.212E+03	2.212E+03	2.216E+03	2.219E+03	2.219E+03	2200
	30	2.242E+03	2.247E+03	2.246E+03	2.246E+03	2.260E+03	2.258E+03	2200
12	10	2.708E+03	2.739E+03	2.739E+03	2.852E+03	2.976E+03	2.981E+03	2400
	30	3.313E+03	3.444E+03	3.454E+03	3.763E+03	4.083E+03	4.094E+03	2400
13	10	3.255E+03	3.258E+03	3.258E+03	3.272E+03	3.298E+03	3.300E+03	2600
	30	3.372E+03	3.388E+03	3.384E+03	3.409E+03	3.433E+03	3.426E+03	2600
14	10	3.207E+03	3.209E+03	3.209E+03	3.214E+03	3.217E+03	3.217E+03	2800
	30	3.261E+03	3.267E+03	3.266E+03	3.281E+03	3.300E+03	3.300E+03	2800
15	10	3.718E+03	3.765E+03	3.777E+03	3.813E+03	3.911E+03	3.902E+03	3000
	30	4.264E+03	4.414E+03	4.427E+03	4.704E+03	4.836E+03	4.836E+03	3000

The results of four different variants of CMA-ES are presented in this section. CMAES-S and CMAES-G denote the specialized and generalized parameters experiments respectively. CMAES-R is the best parameter set $\mathbf{p}^*$ from the generalized parameter experiment replicated on a new set of starting positions. The same set of starting position is used to get the results for the default parameters, CMAES-D.

The computational complexity for the default and optimized parameter values are shown in Table II, calculated according to the guidelines given in [7].

A. Parameter Tuning Results

Since the same formula is used for the design layer problem as the scoring method in the CEC'15 expensive problems, the results for the parameter tuning experiments are representative of the final score.

Table III shows the best, median and mean utility from each experiment. The scores, as calculated by the objective function in Equation (1), are shown in Table IV. For CMAES-S the values are the summation of all individual problems. Thus, the difference of CMAES-S and CMAES-G is the performance gain (or loss) of allowing each test problem to use specialized instead of generalized parameters.

TABLE IV. The score as calculated by the objective function in Equation (1). The specialized parameters score (CMAES-S), is the summation of all individual results.

	Best	Worst	Median	Mean	Min
CMAES-S	6.035E+07	3.567E+08	1.095E+08	1.514E+08	48000
CMAES-G	1.379E+08	1.248E+09	1.982E+08	2.543E+08	48000

The specialized parameters are able to improve the performance by 81% for the median and 128% for the best result. This shows that the no free lunch theorem holds for CMA-ES and the CEC'15 problems, because no set of parameters could be found that are optimal across all problems. From the results it is also clear that functions $1, 2, 10$ are the most difficult ones. It is reasonable to assume that those functions will influence the general parameters the most, because of the formulation of the scoring method.

The optimized parameter values from all 20 replications, for both 10 and 30 dimensions, are shown as boxplots in Figures 1-7. The function labeled G in the plots uses the generalized parameters.

1) Figure 1: The median values for $\sigma^{(0)}$ vary between problems. The variation pattern is roughly the same for both 10 and 30 dimensions. However, 30 dimensions have lower values. For 10 dimensions all median values are below 4, while for 30 dimensions the median values are below 2.

2) Figure 2: The median values for μ do not significantly change between problems. The exceptions are functions 9 and 15, which have lower values than the rest. Another observation is that 30 dimensions have in general higher values than 10 dimensions.

3) Figures 3, 4 and 5: No clear patterns can be observed with d_σ, c_σ and c_c. Although the significance of d_σ seems to be higher for 30 dimensions, especially for function 1. More experiments are needed to determine if they do not influence the performance in a significant way or if they have dependencies on other parameters which allow them to have a wide range of optimal values.

4) Figure 6: The median values for c_{ccov} vary between problems. One trend that can be observed is that the values for 30 dimensions are smaller than for 10. Smaller values are clearly preferred for the generalized parameters.

5) Figure 7: In general, optimized λ values are around 10, with exceptions for functions 5, 9 and 15 with 10 dimensions and 2, 5 and 9 with 30 dimensions. Apart from those exceptions, there are no discernible differences between 10 and 30 dimensions.

B. Tuned Parameters vs Default Parameters

Table V shows the parameters with the best utility from each experiment. The row labeled *Generalized* are the best generalized parameters and *Default* are the parameters as suggested by the heuristics. The optimized λ and μ parameters are smaller than the default. This will lead to faster convergence at the cost of reduced global search capability. It is difficult to draw any conclusions about the other optimized parameters as they are similar to the default values.

Tables VII and IX shows the performance of the generalized parameters on the CEC'15 problems, for 10 and 30 dimensions. For comparison, the performance using default values are shown in Table VI and VIII. The fitness values for those experiments are calculated using Equation (16). Note that this is different from the parameter tuning experiments which used Equation (2).

TABLE V. OPTIMIZED PARAMETER VALUES: THE PARAMETER VALUES FROM THE EXPERIMENT REPLICATION WITH THE BEST PERFORMANCE

Func	N	$\sigma^{(0)}$	μ	d_σ	c_σ	c_c	c_{cov}	λ
1	10	1.341E+00	9.994E-02	2.934E+00	4.739E-02	6.693E-08	1.499E-01	5.000E+00
	30	1.337E+00	3.932E-01	5.024E-01	3.554E-01	5.272E-01	7.450E-04	7.000E+00
2	10	8.550E-01	3.555E-01	4.954E+00	4.965E-01	9.793E-01	8.161E-01	1.800E+01
	30	6.286E-01	5.599E-01	1.292E-01	1.344E-01	4.651E-01	2.906E-01	1.500E+01
3	10	1.851E+00	2.724E-01	1.000E+01	2.886E-01	1.217E-05	1.814E-01	9.000E+00
	30	2.824E+00	4.494E-01	2.722E-01	1.058E-01	9.398E-01	3.914E-04	1.300E+01
4	10	7.575E-01	1.383E-01	1.492E+00	3.821E-01	3.565E-04	3.412E-01	1.400E+01
	30	4.324E-01	2.531E-01	1.158E+00	6.300E-01	1.212E-04	6.571E-02	1.100E+01
5	10	9.748E-02	5.291E-01	3.206E-01	6.470E-02	6.863E-01	4.627E-01	1.300E+01
	30	1.346E-01	2.801E-01	8.376E-02	1.149E-02	6.462E-01	2.871E-02	1.400E+01
6	10	1.635E+00	1.935E-01	5.480E+00	6.873E-01	4.618E-06	1.495E-01	6.000E+00
	30	4.682E-01	2.050E-01	9.983E+00	6.617E-01	2.078E-06	4.764E-02	6.000E+00
7	10	2.285E+00	6.032E-02	6.916E+00	6.475E-01	4.428E-06	2.086E-01	8.000E+00
	30	1.551E+00	2.320E-01	4.200E-01	3.779E-01	8.272E-01	1.494E-02	7.000E+00
8	10	3.250E+00	1.399E-02	1.000E+01	9.300E-03	1.212E-08	2.793E-01	1.100E+01
	30	8.221E-01	3.907E-01	9.477E-01	4.005E-01	9.992E-01	8.285E-04	8.000E+00
9	10	2.246E+00	9.116E-02	4.103E+00	2.033E-01	9.590E-01	8.284E-01	3.900E+01
	30	1.873E+00	5.461E-02	5.706E+00	2.136E-01	1.827E-01	3.811E-01	3.500E+01
10	10	1.370E+00	1.339E-01	6.347E+00	4.320E-01	8.358E-01	2.989E-01	6.000E+00
	30	1.086E+00	2.904E-01	3.039E+00	7.439E-01	5.137E-01	1.462E-01	1.000E+01
11	10	2.581E+00	2.417E-01	9.192E+00	9.964E-01	6.081E-01	3.129E-01	8.000E+00
	30	2.575E+00	2.454E-01	3.206E-01	9.489E-02	2.544E-01	2.010E-02	8.000E+00
12	10	2.958E+00	1.494E-01	8.872E+00	5.143E-01	2.276E-01	4.474E-01	1.500E+01
	30	9.861E-01	3.173E-01	4.710E+00	2.676E-01	8.234E-01	2.935E-01	1.500E+01
13	10	4.026E+00	2.875E-01	2.686E-01	5.115E-02	9.365E-01	2.592E-01	8.000E+00
	30	7.501E-01	2.974E-01	8.156E-01	7.556E-01	6.063E-01	2.166E-02	8.000E+00
14	10	1.247E+00	1.997E-01	2.225E+00	3.027E-01	4.124E-01	2.420E-01	9.000E+00
	30	6.248E-01	2.756E-01	2.228E+00	8.839E-01	1.888E-01	1.615E-02	7.000E+00
15	10	2.796E+00	7.519E-03	2.124E+00	3.883E-01	4.969E-01	3.731E-01	3.500E+01
	30	3.394E-01	3.728E-01	9.461E+00	5.925E-01	5.177E-01	8.894E-03	7.000E+00
Generalized		1.336E+00	2.629E-01	5.206E-01	2.777E-01	6.560E-01	2.632E-03	6.000E+00
Default	10	2.000E+00	5.000E-01	1.130E+00	3.299E-01	2.857E-01	3.246E-02	1.000E+01
	20	2.000E+00	5.000E-01	8.942E-01	1.742E-01	1.176E-01	6.573E-03	1.400E+01

TABLE VI. CMAES-D, RESULTS FOR 10D

Func	Best	Worst	Median	Mean	Std
1	4.914E+06	6.702E+08	4.275E+07	9.261E+07	1.479E+08
2	1.524E+04	2.063E+05	4.694E+04	6.232E+04	4.452E+04
3	3.032E+02	3.109E+02	3.063E+02	3.066E+02	2.107E+00
4	1.419E+03	2.563E+03	2.278E+03	2.187E+03	3.059E+02
5	5.019E+02	5.044E+02	5.028E+02	5.029E+02	5.829E-01
6	6.004E+02	6.016E+02	6.007E+02	6.007E+02	2.666E-01
7	7.004E+02	7.038E+02	7.006E+02	7.008E+02	7.462E-01
8	8.045E+02	9.544E+02	8.065E+02	8.145E+02	3.302E+01
9	9.036E+02	9.044E+02	9.042E+02	9.041E+02	2.723E-01
10	1.388E+05	2.954E+06	4.438E+05	7.204E+05	8.054E+05
11	1.106E+03	1.115E+03	1.109E+03	1.108E+03	2.161E+00
12	1.296E+03	1.655E+03	1.466E+03	1.480E+03	1.141E+02
13	1.621E+03	1.727E+03	1.638E+03	1.649E+03	2.835E+01
14	1.596E+03	1.619E+03	1.612E+03	1.610E+03	5.338E+00
15	1.559E+03	2.060E+03	1.952E+03	1.942E+03	1.011E+02

TABLE VII. CMAES-R, RESULTS FOR 10D

Func	Best	Worst	Median	Mean	Std
1	1.106E+03	4.433E+07	1.490E+05	2.486E+06	9.853E+06
2	9.478E+03	6.443E+04	3.756E+04	3.725E+04	1.670E+04
3	3.033E+02	3.104E+02	3.063E+02	3.067E+02	1.841E+00
4	8.827E+02	2.674E+03	2.009E+03	1.912E+03	5.718E+02
5	5.010E+02	5.040E+02	5.030E+02	5.027E+02	8.442E-01
6	6.003E+02	6.008E+02	6.005E+02	6.005E+02	1.253E-01
7	7.002E+02	7.062E+02	7.005E+02	7.009E+02	1.305E+00
8	8.037E+02	3.690E+03	8.053E+02	9.496E+02	6.451E+02
9	9.039E+02	9.046E+02	9.040E+02	9.041E+02	1.997E-01
10	2.680E+04	3.712E+06	4.013E+05	8.762E+05	1.013E+06
11	1.104E+03	1.113E+03	1.107E+03	1.108E+03	2.786E+00
12	1.235E+03	1.699E+03	1.421E+03	1.429E+03	1.194E+02
13	1.618E+03	2.023E+03	1.632E+03	1.657E+03	8.793E+01
14	1.592E+03	1.618E+03	1.602E+03	1.603E+03	6.737E+00
15	1.514E+03	2.084E+03	1.913E+03	1.900E+03	1.438E+02

TABLE VIII. CMAES-D, RESULTS FOR 30D

Func	Best	Worst	Median	Mean	Std
1	2.784E+07	3.736E+08	1.230E+08	1.219E+08	7.569E+07
2	1.021E+05	2.181E+05	1.574E+05	1.504E+05	3.346E+04
3	3.112E+02	3.232E+02	3.182E+02	3.177E+02	3.380E+00
4	7.198E+03	8.982E+03	8.149E+03	8.102E+03	5.238E+02
5	5.029E+02	5.054E+02	5.043E+02	5.043E+02	5.577E-01
6	6.005E+02	6.011E+02	6.008E+02	6.008E+02	1.614E-01
7	7.004E+02	7.014E+02	7.006E+02	7.008E+02	2.996E-01
8	8.191E+02	9.451E+02	8.284E+02	8.426E+02	3.109E+01
9	9.133E+02	9.141E+02	9.139E+02	9.138E+02	2.369E-01
10	4.747E+06	4.865E+07	1.551E+07	2.012E+07	1.284E+07
11	1.122E+03	1.170E+03	1.130E+03	1.132E+03	1.019E+01
12	2.038E+03	2.868E+03	2.408E+03	2.424E+03	2.349E+02
13	1.677E+03	2.034E+03	1.769E+03	1.791E+03	8.348E+01
14	1.628E+03	1.712E+03	1.661E+03	1.665E+03	2.131E+01
15	2.079E+03	2.501E+03	2.335E+03	2.304E+03	1.084E+02

TABLE IX. CMAES-R, RESULTS FOR 30D

Func	Best	Worst	Median	Mean	Std
1	9.691E+03	1.476E+07	2.264E+05	1.289E+06	3.347E+06
2	8.245E+04	2.078E+05	1.420E+05	1.416E+05	3.037E+04
3	3.185E+02	3.327E+02	3.250E+02	3.253E+02	3.919E+00
4	3.209E+03	8.927E+03	5.435E+03	5.773E+03	1.905E+03
5	5.005E+02	5.053E+02	5.043E+02	5.041E+02	1.066E+00
6	6.004E+02	6.009E+02	6.007E+02	6.007E+02	1.510E-01
7	7.003E+02	7.011E+02	7.004E+02	7.006E+02	3.107E-01
8	8.195E+02	8.413E+02	8.259E+02	8.273E+02	6.642E+00
9	9.135E+02	9.143E+02	9.140E+02	9.139E+02	2.178E-01
10	1.221E+06	1.148E+07	3.116E+06	4.122E+06	2.974E+06
11	1.119E+03	1.236E+03	1.123E+03	1.136E+03	3.454E+01
12	1.482E+03	2.503E+03	1.794E+03	1.861E+03	2.499E+02
13	1.671E+03	1.747E+03	1.691E+03	1.693E+03	1.618E+01
14	1.621E+03	1.666E+03	1.633E+03	1.635E+03	1.177E+01
15	1.940E+03	2.526E+03	2.351E+03	2.296E+03	1.710E+02

TABLE X. PERFORMANCE DIFFERENCE OF DEFAULT (CMAES-D) AND OPTIMIZED PARAMETER VALUES (CMAES-R).

Func	N	CMAES-D	CMAES-R	Diff	Min
1	10	1.354E+08	2.635E+06	1.327E+08	200
	30	2.449E+08	1.515E+06	2.434E+08	200
2	10	1.093E+05	7.481E+04	3.445E+04	400
	30	3.078E+05	2.836E+05	2.419E+04	400
3	10	6.128E+02	6.130E+02	-1.848E-01	600
	30	6.359E+02	6.503E+02	-1.440E+01	600
4	10	4.465E+03	3.921E+03	5.442E+02	800
	30	1.625E+04	1.121E+04	5.042E+03	800
5	10	1.006E+03	1.006E+03	1.610E-01	1000
	30	1.009E+03	1.008E+03	1.410E-01	1000
6	10	1.201E+03	1.201E+03	5.121E-01	1200
	30	1.202E+03	1.201E+03	1.711E-01	1200
7	10	1.401E+03	1.401E+03	-3.719E-02	1400
	30	1.401E+03	1.401E+03	3.318E-01	1400
8	10	1.621E+03	1.755E+03	-1.339E+02	1600
	30	1.671E+03	1.653E+03	1.794E+01	1600
9	10	1.808E+03	1.808E+03	1.079E-01	1800
	30	1.828E+03	1.828E+03	-1.997E-01	1800
10	10	1.164E+06	1.277E+06	-1.133E+05	2000
	30	3.563E+07	7.238E+06	2.839E+07	2000
11	10	2.217E+03	2.214E+03	2.681E+00	2200
	30	2.263E+03	2.259E+03	3.276E+00	2200
12	10	2.945E+03	2.850E+03	9.574E+01	2400
	30	4.832E+03	3.654E+03	1.178E+03	2400
13	10	3.287E+03	3.289E+03	-1.686E+00	2600
	30	3.561E+03	3.383E+03	1.774E+02	2600
14	10	3.222E+03	3.205E+03	1.694E+01	2800
	30	3.327E+03	3.268E+03	5.827E+01	2800
15	10	3.894E+03	3.813E+03	8.078E+01	3000
	30	4.639E+03	4.647E+03	-8.247E+00	3000
All		4.176E+08	1.309E+07	4.045E+08	48000

The optimized parameter values are able to improve on the default parameter values in most of the CEC'15 problems, see Table X. Functions 1, 2 and 10 are the ones that did improve the most. For the total score over all problems, the optimized parameters increased the performance by a factor of almost 32.

V. CONCLUSIONS

This paper tuned the parameters of CMA-ES with the aim of maximizing its performance on the CEC'15 expensive problems. A bilevel optimization approach was used to search for both generalized and specialized parameters. The results show that generalized parameters have lower performance than specialized parameters, and that the general parameters can have values that are different from any of the specialized parameters.

The optimized parameter values were also compared against the values suggested by the heuristics, both in terms of performance on CEC'15 problems and how similar they were. Two notable differences have been seen in the λ and μ parameters. Small values for the population size parameter λ leads to fast convergence and large values help in avoiding local optima. Compared to the default parameters, the parameter tuning results show that a smaller λ provide a better trade-off for the CEC'15 problems. Larger values of μ increase the explorative behavior. The default heuristic for μ, Equation (4), from [11] suggested a value of 0.5. The results show that the optimized value of 0.26 is closer to the proposed value of 0.27 in [1]. Even though λ and μ are highlighted here, that does not mean that they are solely responsible for the performance increase. Further work needs to be done to determine the significance of each parameter and how they interact with each other.

No parameters have been found that are optimal across all problems, which is in agreement with the no free lunch theorem. The parameters found in this study are influenced by the CEC'15 scoring method and the fact that a relatively limited function evaluation budget was used. How well these parameters work on other problems with different function evaluations budgets is difficult to estimate, without doing further experiments.

The heuristics are designed to scale the parameters with the dimension of the problem. This aspect of the optimized parameters are not studied in this paper. Experiments that find optimal parameters for different dimensions are needed to address this issue.

REFERENCES

[1] N. Hansen and A. Ostermeier, "Completely Derandomized Self-Adaptation in Evolution Strategies," *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, Jun. 2001.

[2] N. Hansen, "The CMA Evolution Strategy: A Comparing Review," in *Towards a New Evolutionary Computation*, ser. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Jan. 2006, no. 192, pp. 75–102.

[3] N. Hansen and S. Kern, "Evaluating the CMA Evolution Strategy on Multimodal Test Functions," in *Parallel Problem Solving from Nature - PPSN VIII*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2004, no. 3242, pp. 282–291.

[4] N. Hansen, R. Ros, N. Mauny, M. Schoenauer, and A. Auger, "Impacts of invariance in search: When CMA-ES and PSO face ill-conditioned and non-separable problems," *Applied Soft Computing*, vol. 11, no. 8, pp. 5755–5769, Dec. 2011.

[5] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík, "Comparing Results of 31 Algorithms from the Black-box Optimization Benchmarking BBOB-2009," in *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, ser. GECCO '10. New York, NY, USA: ACM, 2010, pp. 1689–1696.

[6] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

[7] Q. Chen, B. Liu, Q. Zhang, J. J. Liang, P. N. Suganthan, and B. Y. Qu, "Problem Definition and Evaluation Criteria for CEC 2015 Special Session and Competition on Bound Constrained Single-Objective Computationally Expensive Numerical Optimization," Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Technical Report, Nanyang Technological University, Singapore, Technical Report, Nov. 2014.

[8] S. K. Smit and A. E. Eiben, "Parameter Tuning of Evolutionary Algorithms: Generalist vs. Specialist," in *Applications of Evolutionary Computation*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2010, no. 6024, pp. 542–551.

[9] A. E. Eiben and S. K. Smit, "Parameter tuning for configuring and analyzing evolutionary algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 19–31, Mar. 2011.

[10] A. Sinha, P. Malo, P. Xu, and K. Deb, "A Bilevel Optimization Approach to Automated Parameter Tuning," in *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation*, ser. GECCO '14. New York, NY, USA: ACM, 2014, pp. 847–854.

[11] N. Hansen, "CMA-ES code in C," 2014. [Online]. Available: https://github.com/cma-es/c-cmaes

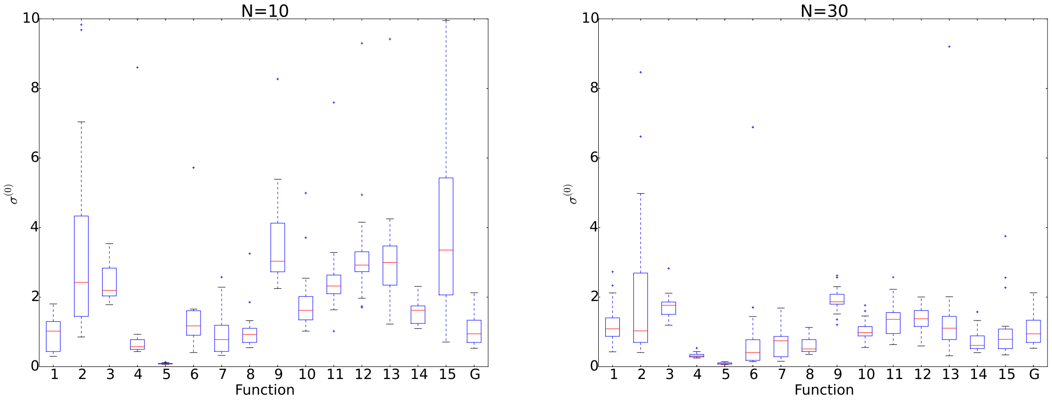

Fig. 1. Parameter tuning results for parameter $\sigma^{(0)}$.

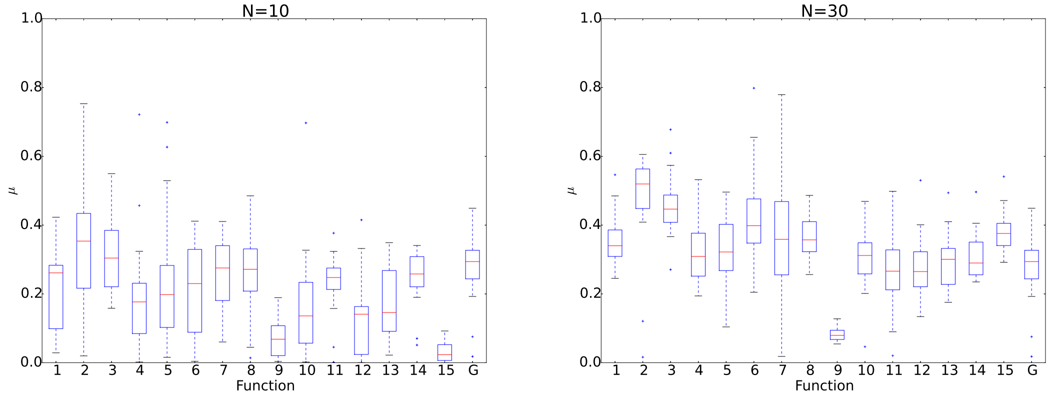

Fig. 2. Parameter tuning results for parameter μ.

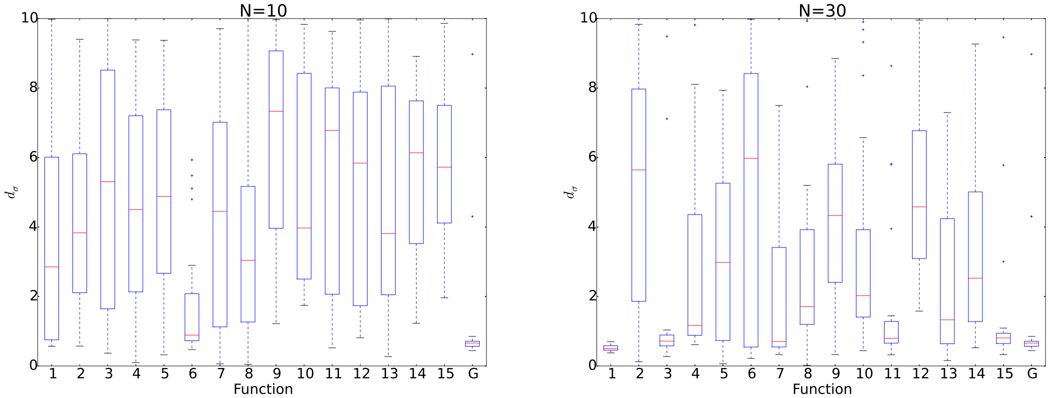

Fig. 3. Parameter tuning results for parameter d_σ.

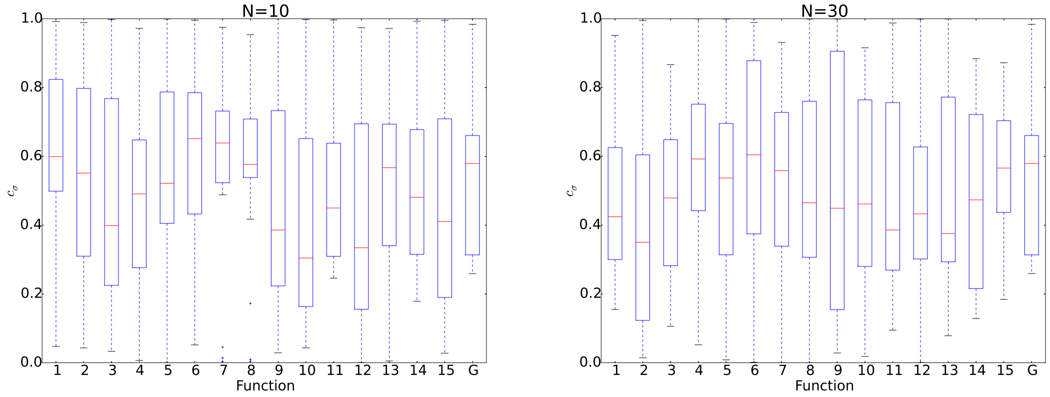

Fig. 4. Parameter tuning results for parameter c_σ.

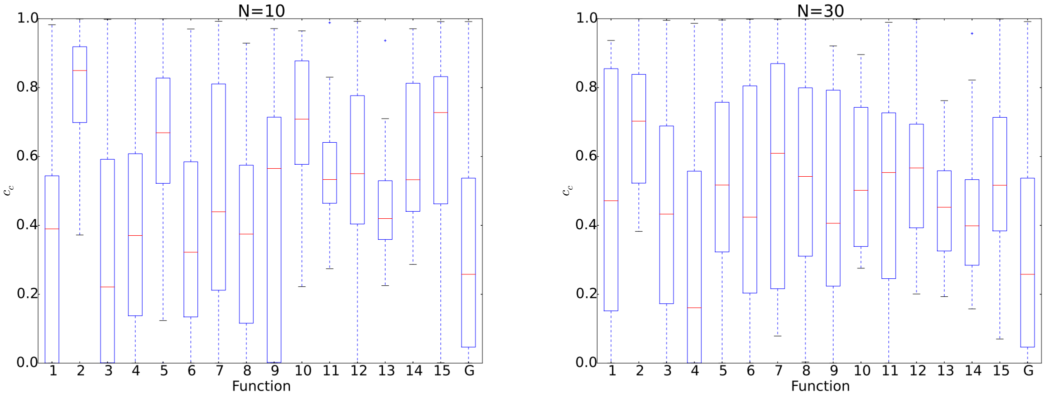

Fig. 5. Parameter tuning results for parameter c_{c}.

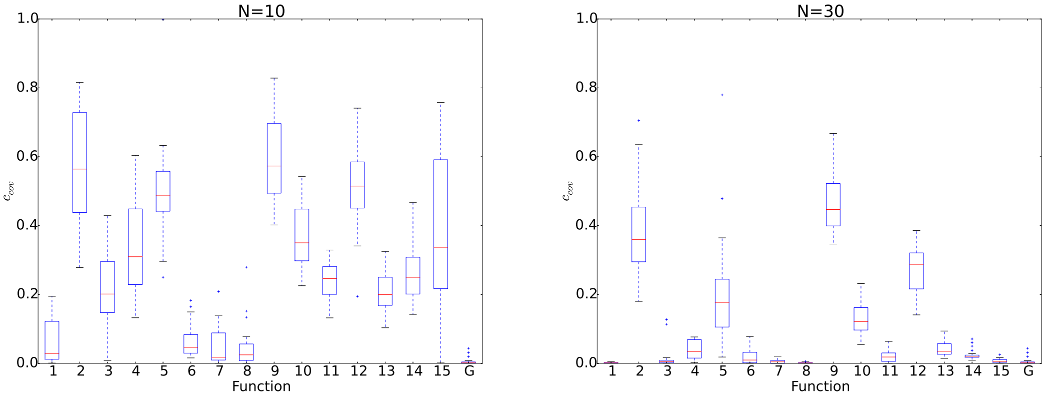

Fig. 6. Parameter tuning results for parameter c_{cov}.

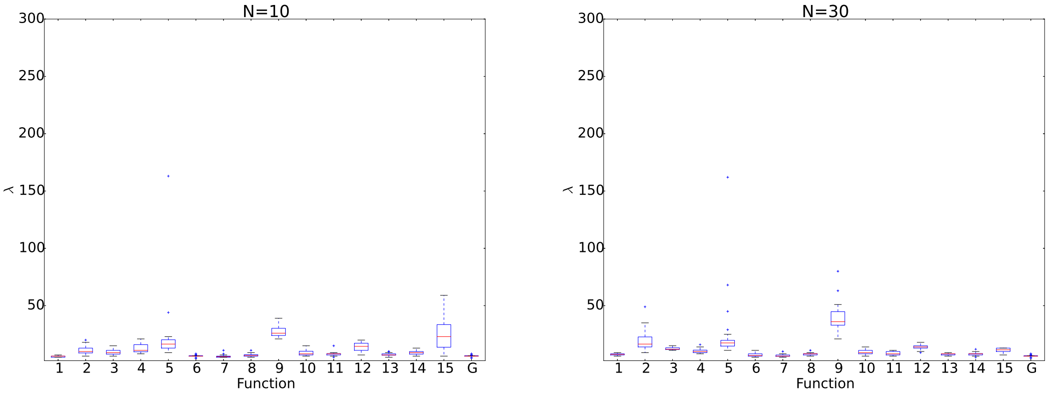

Fig. 7. Parameter tuning results for parameter λ.