Evolution strategies for robust optimization

Kruisselbrink, J.W.

Author: Kruisselbrink, Johannes Willem
Title: Evolution strategies for robust optimization
Date: 2012-05-10

Chapter 6

A Study on Noise Handling Schemes

In Chapter 5, three adaptive averaging techniques are identified as suitable and promising techniques for noise handling in the context of the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES. These techniques are: poset based adaptive averaging (PUH, Section 5.4.3), rank-change based adaptive averaging (UH, Section 5.4.5), and rank-inversions based adaptive averaging (IUH, Section 5.4.6). All three are based on the same inter-generation adaptive averaging framework and use the same noise treatment approach of explicit averaging (or resampling). The difference between them is that they are based on different uncertainty quantification methods.

In this chapter, we aim to study these approaches in more depth. In particular, we will aim to answer the following questions: 1) At what rate should the sample size grow when using adaptive resampling for optimization of noisy objective functions? 2) What parameter settings are appropriate for the adaptive averaging techniques considered in this work? 3) Does using the adaptive averaging techniques considered in this chapter yield better results than using explicit or implicit averaging?

This chapter is structured as follows: Section 6.1 presents a theoretical study on the ideal growth of the sample size on a simple test problem to gain insight in the dynamics of an optimally functioning adaptive averaging scheme. Section 6.2 presents the results of an empirical study that aims to gain insight into the algorithmic parameters of the adaptive averaging techniques. Section 6.3 presents the results of an empirical study in which the adaptive averaging techniques are compared to the standard techniques of explicit and implicit averaging. Section 6.4 closes with a summary and discussion.

6.1 The Growth Rate of the Sample Size

The inter-generation adaptive averaging techniques considered in this chapter all use a multiplicative update rule for adapting the number of samples when the uncertainty level is too high. In literature, the increment factor for updating the sample size is regarded as an algorithmic parameter that is to be tuned manually. However, two questions that remain are:

should the update rule indeed be exponential and, if so, at what rate should it ideally grow? This section studies how the sample size should optimally develop within an Evolution Strategy, the $(\mu/\mu_I, \lambda)$-ES, on a simple artificial test problem, the noisy sphere problem. This section builds forth on the results obtained by Arnold and Beyer [AB02].

The motivation behind using an adaptive sample size for resampling strategies is to adapt the number of samples used for evaluating each individual such that it is sufficient for the Evolutionary Algorithm to progress. On the other hand, it should not be too high, because then samples will be wasted. This way of formulating the problem of adaptive averaging suggests that, when an Evolution Strategy is in a certain state of the evolution process, there is an optimal sample size. Ideally, an adaptive averaging scheme is able to follow this optimal sample size over the course of evolution.

6.1.1 The Progress Rate of the $(\mu/\mu_I, \lambda)$-ES on the Noisy Sphere

In the analysis of Evolution Strategies, the *progress rate* is an often used performance measure. The progress rate, denoted φ, is defined as the expected distance covered by the centroid of the population towards the location of the optimum within one generation. That is, when R denotes the distance to the optimum location of the centroid of the parents, and when r denotes the distance to the optimum location of the centroid of the selected offspring, then the progress rate reads $\varphi = \mathbf{E}[R - r]$. The progress rate can be used to determine the *efficiency per evaluation* when dividing it by the number of evaluations used for the current generation. By optimizing the efficiency per evaluation with respect to the sample size, we can derive the optimal sample size for an Evolution Strategy at a given stage in the optimization.

The noisy sphere problem is given by the function

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 + \sigma_\epsilon z_\epsilon \to \min \,, \; z_\epsilon \sim \mathcal{N}(0,1), \tag{6.1}$$

mapping n-dimensional vectors $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$ to a noisy objective function value with Gaussian noise with standard deviation σ_ϵ. For this model, Arnold and Beyer [AB02] derived for the $(\mu/\mu_I, \lambda)$-ES the normalized progress rate

$$\varphi^*_{\mu/\mu,\lambda} \simeq \frac{c_{\mu/\mu,\lambda}\sigma^* \left(1 + \sigma^{*2}/(2\mu n)\right)}{\sqrt{1+\sigma^{*2}/(\mu n)}\sqrt{1+\upsilon^2+\sigma^{*2}/(2n)}} - n\left[\sqrt{1+\frac{\sigma^{*2}}{\mu n}} - 1\right], \tag{6.2}$$

where $c_{\mu/\mu,\lambda}$ is the $(\mu/\mu_I, \lambda)$-progress coefficient and φ^* is the normalized progress rate

$$\varphi^* = \varphi \frac{n}{R}. \tag{6.3}$$

Furthermore, σ^*_ϵ is a normalized version of σ_ϵ, according to normalization

$$\sigma^*_\epsilon = \sigma_\epsilon \frac{n}{2R^2}, \tag{6.4}$$

σ^* is a normalized version of the stepsize parameter σ, reading

$$\sigma^* = \sigma \frac{n}{R}, \tag{6.5}$$

and υ is the noise-to-signal ratio, defined as

$$\upsilon = \frac{\sigma_\epsilon^*}{\sigma^*}. \tag{6.6}$$

6.1.2 The Efficiency of Resampling

Resampling with m samples reduces the error of the objective function approximations by a factor of $\sqrt{m}$, that is, when using m samples, we have an effective approximation error of

$$\hat{\sigma}_\epsilon = \sigma_\epsilon / \sqrt{m}. \tag{6.7}$$

Hence, resampling increases the progress rate φ^* (see Eq. 6.2), however, at the cost of requiring more evaluations. In order to determine the efficiency, the progress rate should be divided by the number of evaluations used for the current generation, i.e., by λm.

When letting $\varphi_{\hat{\sigma}_\epsilon}$ denote the progress rate for noise factor $\hat{\sigma_\epsilon}$, we can state the efficiency η_m of a certain sample size m as

$$\eta_m = \frac{\varphi_{\hat{\sigma}_\epsilon}}{\lambda m}. \tag{6.8}$$

For the $(\mu/\mu_I, \lambda)$-ES on the noisy sphere problem, when substituting Eq. 6.2 into this equation and using $\upsilon = \sigma_\epsilon^*/(\sqrt{m}\sigma^*)$ (i.e., using the effective approximation error due to resampling), this yields a normalized efficiency

$$\begin{aligned}
\eta_m^* &= \frac{\frac{c_{\mu/\mu,\lambda}\sigma^*\left(1+\sigma^{*2}/(2\mu n)\right)}{\sqrt{1+\sigma^{*2}/(\mu n)}\sqrt{1+\sigma_\epsilon^{*2}/(m\sigma^{*2})+\sigma^{*2}/(2n)}} - n\left[\sqrt{1+\frac{\sigma^{*2}}{\mu n}}-1\right]}{\lambda m} && (6.9)\\
&= \frac{c_{\mu/\mu,\lambda}\sigma^*\left(1+\sigma^{*2}/(2\mu n)\right)}{\lambda m\sqrt{1+\sigma^{*2}/(\mu n)}\sqrt{1+\sigma_\epsilon^{*2}/(m\sigma^{*2})+\sigma^{*2}/(2n)}} \\
&\quad -\frac{n\left[\sqrt{1+\frac{\sigma^{*2}}{\mu n}}-1\right]}{\lambda m}. && (6.10)
\end{aligned}$$

6.1.3 The Optimal Sample Size

The optimal sample size for Evolution Strategies at a given stage of the optimization can be determined by maximizing the efficiency with respect to m. That is,

$$m_{\text{opt}} = \text{argmax}_m \eta_m^*. \tag{6.11}$$

When omitting the fact that in practice m can only assume positive integer values, this can be done by solving

$$\frac{\partial \eta_m^*}{\partial m} = 0. \tag{6.12}$$

In order to do so, we apply the substitutions

$$\begin{aligned} c_1 &= c_{\mu/\mu,\lambda}\,\sigma^* \left(\frac{\sigma^{*2}}{2\,\mu\,n}+1\right), && (6.13)\\ c_2 &= \lambda\sqrt{\frac{\sigma^{*2}}{\mu\,n}+1}, && (6.14)\\ c_3 &= 1+\frac{\sigma^{*2}}{2n}, && (6.15)\\ c_4 &= \sigma_\epsilon^{*2}/\sigma^{*2}, && (6.16)\\ c_5 &= \frac{n\left[\sqrt{1+\frac{\sigma^{*2}}{\mu n}}-1\right]}{\lambda}. && (6.17) \end{aligned}$$

This yields

$$\eta_m^* = \frac{c_1}{mc_2\sqrt{c_3+c_4/m}} - \frac{c_5}{m}, \tag{6.18}$$

and

$$\frac{\partial\eta_m^*}{\partial m} = \frac{2\,c_2\,c_5\,m\left(\frac{c_3\,m+c_4}{m}\right)^{3/2} - 2\,c_1\,c_3\,m - c_1\,c_4}{2\,c_2\,m^3\left(\frac{c_3\,m+c_4}{m}\right)^{3/2}}. \tag{6.19}$$

Hence, to solve $\frac{\partial\eta_m^*}{\partial m}=0$ for $m \geq 1$, we need to solve

$$2\,c_2\,c_5\,m\left(\frac{c_3\,m+c_4}{m}\right)^{\frac{3}{2}} - 2\,c_1\,c_3\,m - c_1\,c_4 = 0. \tag{6.20}$$

This equation has only one real solution, namely

$$\begin{aligned} m_{\text{opt}} &= -\frac{c_4(3c_2^2c_3c_5^2-c_1^2)}{3c_3(c_2^2c_3c_5^2-c_1^2)} - \frac{(c_1^6c_3^3c_4^3+36c_1^4c_2^2c_3^4c_4^3c_5^2+27c_1^2c_2^4c_3^5c_4^3c_5^4+c_6)^{1/3}}{6c_3^2(c_2^2c_3c_5^2-c_1^2)} && (6.21)\\ &\quad+\frac{(-4c_1^4c_3^2c_4^2-60c_1^2c_2^2c_3^3c_4^2c_5^2)}{24c_3^2(c_2^2c_3c_5^2-c_1^2)(c_1^6c_3^3c_4^3+36c_1^4c_2^2c_3^4c_4^3c_5^2+27c_1^2c_2^4c_3^5c_4^3c_5^4+c_6)^{1/3}}, && (6.22) \end{aligned}$$

with

$$c_6 = 3\sqrt{3}\sqrt{c_1^{10}c_2^2c_3^7c_4^6c_5^2+25c_1^8c_2^4c_3^8c_4^6c_5^4-53c_1^6c_2^6c_3^9c_4^6c_5^6+27c_1^4c_2^8c_3^{10}c_4^6c_5^8}. \tag{6.23}$$

As can be seen, this expression is quite involved and hard to simplify any further. Yet, it is an exact derivation with respect to Eq. 6.20. Moreover, when assuming that σ^* is constant (which is realistic when assuming a properly functioning adaptation mechanism of the stepsize) and using the substitutions Eq. 6.17 and Eq. 6.4, it can be shown that

$$m_{\text{opt}} \propto \Theta\left(\frac{1}{R^4}\right). \tag{6.24}$$

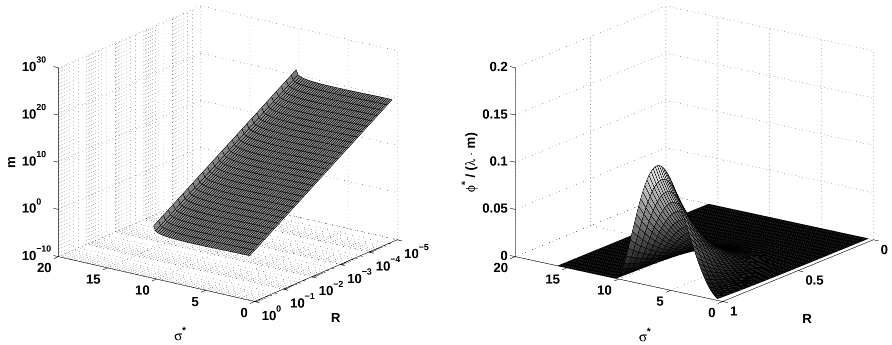


Figure 6.1: The derived optimal sample size (left) and the efficiency $\eta^* = \varphi^*/(\lambda m)$ obtained for the optimal sample size (right) for the $(\mu/\mu_I, \lambda)$-ES for different values of R and σ^*. These results are obtained with $\mu = 8$, $\lambda = 32$, $\sigma_\epsilon = 1$, $n = 10$.

That is, the sample size should grow quartically with the inverse distance to the optimum. Intuitively, this result is plausible, because the resampling reduces the noise with a factor of $\sqrt{m}$ and for the sphere problem the local fitness differences decrease quadratically with R.

In order to get a picture of the development of m_{opt} for different values of R and σ^*, Figure 6.1 shows the derived optimal sample size and the efficiency $\eta^* = \varphi^*/(\lambda m)$ obtained for the optimal sample size for the $(\mu/\mu_I, \lambda)$-ES for different values of R and σ^*. These results are obtained with $\mu = 8$, $\lambda = 32$, $\sigma_\epsilon = 1$, $n = 10$. The slope of the log-log plot of m_{opt} in the direction of R is approximately 4, which is an empirical indicator to support the result of Eq. 6.24. Furthermore, we see that even with an optimally adapted sample size, the efficiency drops quickly for decreasing R (which can largely be attributed to the fast growth of m_{opt}), and seems optimal for $\sigma^* \approx 5$.

6.1.4 The Minimally Required Sample Size

Similarly, a lower bound for m can be derived. That is, the minimally required number of samples to maintain positive progress. This requires solving $\eta^*_m = 0$ for m. For this we obtain

$$m_{\text{least}} = -\frac{c_2^2 c_4 c_5^2}{c_2^2 c_3 c_5^2 - c_1^2} \tag{6.25}$$

$$= \frac{-{\sigma_\epsilon}^2 n^2 \left(\sqrt{\frac{\sigma n^2}{\mu R} + 1} - 1\right)^2 \left(\frac{\sigma^2 n}{\mu R^2} + 1\right)}{4\sigma^2 \left(n^2 \left(\sqrt{\frac{\sigma n^2}{\mu R} + 1} - 1\right)^2 \left(\frac{\sigma^2 n}{2R^2} + 1\right)\left(\frac{\sigma^2 n}{\mu R^2} + 1\right) - \frac{\log\left(\frac{\lambda}{\mu}\right)\sigma^2 n^2 \left(\frac{\sigma^2 n}{2\mu R^2} + 1\right)^2}{R^2}\right) R^2}. \tag{6.26}$$

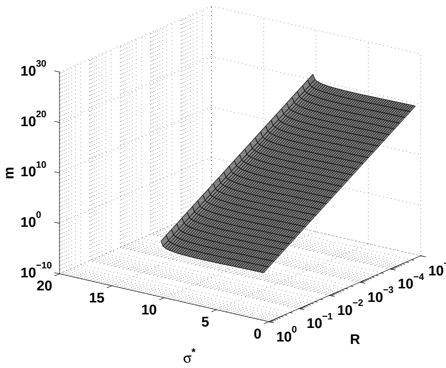


Figure 6.2: The derived minimally required number of samples for the $(\mu/\mu_I, \lambda)$-ES for different values of R and σ^*. These results are obtained with $\mu = 8$, $\lambda = 32$, $\sigma_\epsilon = 1$, $n = 10$.

When assuming that σ^* is constant then we obtain

$$m_{\text{least}} = -\frac{{\sigma_\epsilon}^2\left(\frac{\sigma^{*2}}{\mu n}+1\right)n^4\left(\sqrt{\frac{\sigma^* n}{\mu}+1}-1\right)^2}{4\sigma^{*2}\left(\left(\frac{\sigma^{*2}}{2n}+1\right)\left(\frac{\sigma^{*2}}{\mu n}+1\right)n^2\left(\sqrt{\frac{\sigma^* n}{\mu}+1}-1\right)^2 - c_{\mu/\mu,\lambda}{}^2\sigma^{*2}\left(\frac{\sigma^{*2}}{2\mu n}+1\right)^2\right)R^4}. \tag{6.27}$$

Hence, also for m_{least} we can conclude that, when σ^* is constant,

$$m_{\text{least}} \propto \Theta\left(\frac{1}{R^4}\right). \tag{6.28}$$

Hence, the sample size should grow at least quartically with the inverse distance to the optimum to maintain positive progress. In order to get an impression of the development for different R and σ^*, Figure 6.2 shows the derived minimal required number of samples for the $(\mu/\mu_I, \lambda)$-ES (see, Eq. 6.26) for different values of R and σ^*. These results are obtained with $\mu = 8$, $\lambda = 32$, $\sigma_\epsilon = 1$, $n = 10$. Comparing Figure 6.2 to Figure 6.1, we observe practically the same results. This is not surprising, given that $\frac{1}{R^4}$ is the dominating term of both m_{opt} and m_{least} when n, μ, λ, and σ^* are relatively small.

6.1.5 The Growth Rate of the Sample Size

We consider $\varphi^* = (n\varphi)/R$ and take the optimistic assumption that $\varphi*$ is constant (obtained when the stepsize adaptation mechanism functions optimally). Let R_t denote the distance to the optimum of the centroid of the population at time t. We can express the expected distance to the optimum $\mathbf{E}[R_{t+1}]$ of the centroid of the next generation, at time $t+1$ in terms of φ. I.e.,

$$\mathbf{E}[R_{t+1}] = R_t - \varphi = R_t - \frac{R_t\varphi^*}{n} = R_t - kR_t, \tag{6.29}$$

where $0 < k < 1$ is some constant. We can derive the *generation-wise* limit behavior as

$$\lim_{R_t \to 0} \frac{\mathbf{E}[R_{t+1}]}{R_t} = \lim_{R_t \to 0} \frac{R_t - kR_t}{R_t} = 1 - k < 1. \tag{6.30}$$

Hence, the convergence behavior is linear with respect to the number of generations.

When considering the progress per evaluation using adaptive resampling, we use the efficiency, defined as $\eta_m = \varphi/(\lambda m)$. Considering an optimally tuned adaptive resampling mechanism with $m = k_m/R^4$, k_m being some constant. Still assuming that $\varphi*$ is constant, we can use $m = k_m/R^4$ and obtain

$$\eta_m = \frac{\varphi}{\lambda(k_m/R^4)} = \frac{R^5\varphi^*}{n\lambda k_m} = kR^5. \tag{6.31}$$

where $0 < k < 1$ is some constant. We can express the expected distance to the optimum $\mathbf{E}[R_{t+1}]$ of the centroid of the next *evaluation*, at time $t+1$, in terms of η as

$$\mathbf{E}[R_{t+1}] = R_t - \eta_m = R_t - kR_t^5. \tag{6.32}$$

For this, we can derive the *evaluation-wise* limit behavior as

$$\lim_{R_t \to 0} \frac{\mathbf{E}[R_{t+1}]}{R_t} = \lim_{R_t \to 0} \frac{R_t - kR_t^5}{R_t} = 1. \tag{6.33}$$

Hence, for an optimally adapted sample size, the convergence rate of a $(\mu/\mu_I, \lambda)$-ES on the noisy sphere is sublinear with respect to the number of objective function evaluations.

Finally, using $\mathbf{E}[R_{t+1}] = (1-k)R_t$, $m_t = c/R_t$, and $m_{t+1} = k/\mathbf{E}[R_{t+1}]^4$, we obtain the following growth of the sample size when assuming linear convergence of R with respect to the number of generations:

$$\lim_{R_t \to 0} \frac{m_{t+1}}{m_t} = \lim_{R_t \to 0} \frac{R_t^4}{\mathbf{E}[R_{t+1}]^4} = \lim_{R_t \to 0} \frac{R_t^4}{((1-k)R_t)^4} = \frac{1}{(1-k)^4} > 1. \tag{6.34}$$

In conclusion, the sample size should indeed grow exponentially with respect to the number of generations to achieve a generation-wise linear convergence rate.

6.2 Tuning the Adaptive Averaging Methods

The previous analysis confirmed that the sample size within an adaptive resampling scheme should grow exponentially when generation-wise linear convergence is desired. However, the rate at which it should grow is not yet determined and also the uncertainty threshold for the adaptive averaging techniques remains to be set. To gain insight into how these parameters should be set, we perform an empirical study on each of the adaptive averaging techniques (the PUH, the UH, and the IUH scheme) incorporated in the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES. We consider different instances of these schemes, varying the uncertainty thresholds δ or θ and the resampling factor α.

The experiments are performed on the 10-dimensional noisy sphere problem and a large evaluation budget of $100,000$ function evaluations is considered. For the uncertainty handling schemes, we consider for the uncertainty thresholds δ or θ the values $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ and for the uncertainty treatment parameter we consider the values $\alpha \in \{1.1, 1.3, 1.5, 1.7, 1.9\}$. This yields 25 combinations of settings for each scheme for both the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES. Each run for each instance is repeated 10 times.

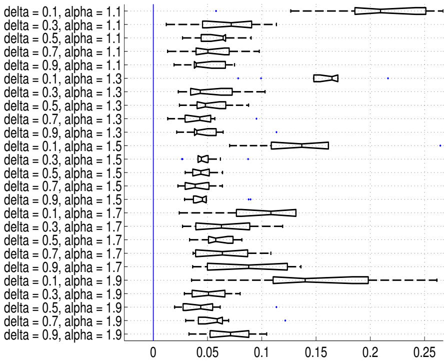
The results of the experiments are shown in Section 6.2.1 and Section 6.2.2 for the PUH-scheme, Section 6.2.3 and Section 6.2.4 show the results of the UH-scheme, and Section 6.2.5 and Section 6.2.6 show the results of the IUH-scheme. For each scheme, the boxplot shows the final solution quality after $100,000$ objective function evaluations, which is computed a posteriori using the noise-free signal function as measure of the expected objective function. The line in the boxplot indicates the objective function value of the optimal solution. The other plot shows the convergence dynamics in terms of the median of the real (noise-free) objective function value of the best individual of the current population versus the number of evaluations. The table summarizes the statistics of the final solution quality. The best instance is determined based on the rank sum.

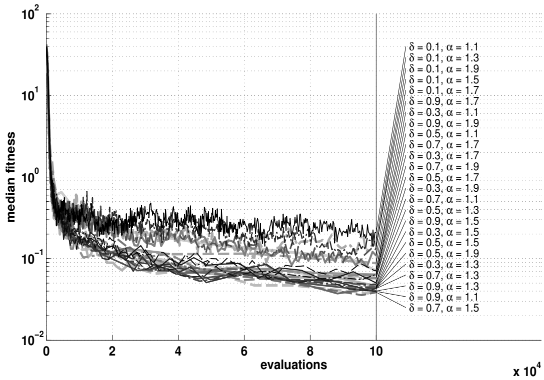
The results show that all adaptive resampling schemes are fairly robust with respect to different settings of δ/θ and α. Except for a few outlier-settings, such as $\delta = 0.1$ for the PUH approach or $\theta = 0.1$ for the UH approach, most of the tested settings yield comparable results. The optimal settings that can be derived from these results differ for each uncertainty quantification scheme and also for the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES. The results of these experiments are summarized in Table 6.1.

	$(5/2_{DI}, 35)$-σ**SA-ES**	**CMA-ES**
PUH	$\delta = 0.7$, $\alpha = 1.5$	$\delta = 0.9$, $\alpha = 1.3$
UH	$\theta = 0.9$, $\alpha = 1.1$	$\theta = 0.9$, $\alpha = 1.9$
IUH	$\theta = 0.3$, $\alpha = 1.9$	$\theta = 0.1$, $\alpha = 1.3$

Table 6.1: The optimal settings (with an approximate error of ± 0.1) of the uncertainty threshold δ/θ and the resampling rate α to achieve the best convergence accuracy on a budget of $100,000$ function evaluations.

6.2.1 Results Tuning PUH-$(5/2_{DI}, 35)$-σSA-ES on the Noisy Sphere

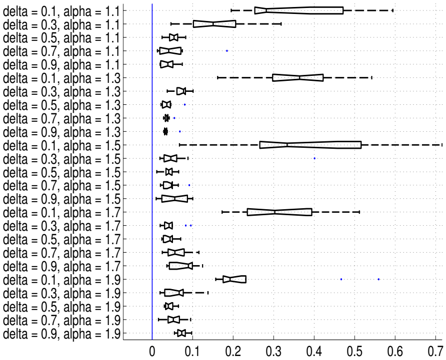


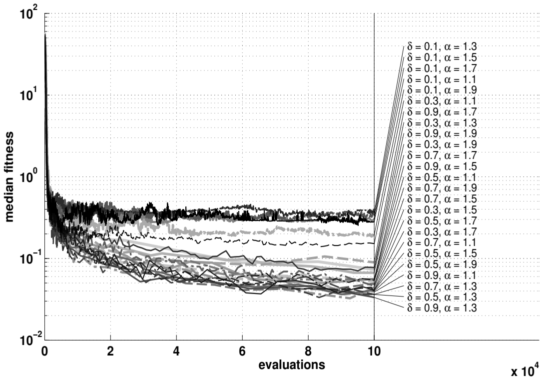


NOISY SPHERE PROBLEM

	Mean	Std	Med	$\sum$#	#
delta = 0.1, alpha = 1.1	5.29	16.12	0.21	2235	25
delta = 0.3, alpha = 1.1	0.07	0.03	0.07	1312	17
delta = 0.5, alpha = 1.1	4.19	13.08	0.06	1302	16
delta = 0.7, alpha = 1.1	0.05	0.02	0.05	1034	12
delta = 0.9, alpha = 1.1	0.05	0.02	0.04	852	6
delta = 0.1, alpha = 1.3	0.17	0.07	0.16	2200	24
delta = 0.3, alpha = 1.3	0.05	0.03	0.04	895	9
delta = 0.5, alpha = 1.3	0.05	0.02	0.05	1016	11
delta = 0.7, alpha = 1.3	0.04	0.02	0.04	752	2
delta = 0.9, alpha = 1.3	0.05	0.03	0.04	858	7
delta = 0.1, alpha = 1.5	0.15	0.07	0.14	2133	23
delta = 0.3, alpha = 1.5	0.05	0.02	0.04	829	5
delta = 0.5, alpha = 1.5	0.04	0.01	0.04	783	3
delta = 0.7, alpha = 1.5	**0.04**	**0.01**	**0.04**	**654**	**1**
delta = 0.9, alpha = 1.5	0.05	0.02	0.05	893	8
delta = 0.1, alpha = 1.7	0.14	0.12	0.11	1839	21
delta = 0.3, alpha = 1.7	0.07	0.03	0.06	1237	14
delta = 0.5, alpha = 1.7	0.06	0.02	0.06	1249	15
delta = 0.7, alpha = 1.7	4.37	13.61	0.06	1357	18
delta = 0.9, alpha = 1.7	0.09	0.04	0.09	1552	20
delta = 0.1, alpha = 1.9	0.17	0.09	0.14	2038	22
delta = 0.3, alpha = 1.9	0.05	0.02	0.05	988	10
delta = 0.5, alpha = 1.9	0.05	0.03	0.04	809	4
delta = 0.7, alpha = 1.9	0.06	0.03	0.06	1106	13
delta = 0.9, alpha = 1.9	4.65	14.48	0.07	1452	19

6.2.2 Results Tuning PUH-CMA-ES on the Noisy Sphere

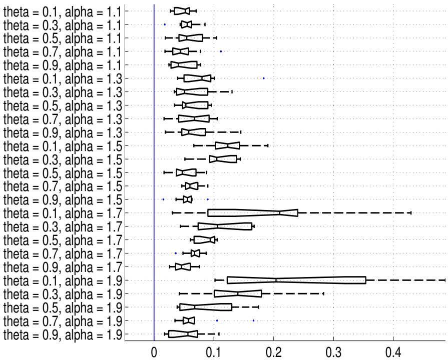


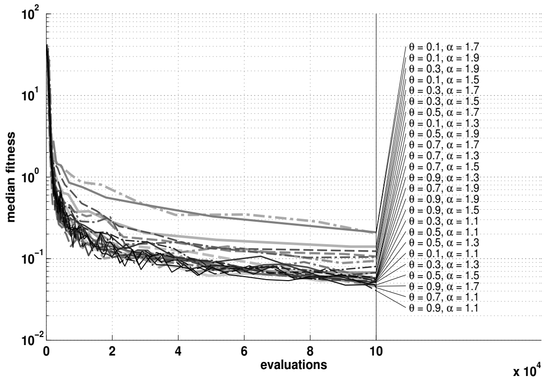


NOISY SPHERE PROBLEM

	Mean	Std	Med	$\sum$#	#
delta = 0.1, alpha = 1.1	0.35	0.14	0.28	2260	24
delta = 0.3, alpha = 1.1	0.16	0.09	0.15	1843	20
delta = 0.5, alpha = 1.1	0.05	0.02	0.05	1044	14
delta = 0.7, alpha = 1.1	0.05	0.05	0.04	839	8
delta = 0.9, alpha = 1.1	0.04	0.02	0.04	707	4
delta = 0.1, alpha = 1.3	0.36	0.12	0.36	2274	25
delta = 0.3, alpha = 1.3	0.07	0.02	0.08	1456	17
delta = 0.5, alpha = 1.3	0.04	0.02	0.04	643	3
delta = 0.7, alpha = 1.3	0.04	0.01	0.04	641	2
delta = 0.9, alpha = 1.3	**0.04**	**0.01**	**0.03**	**589**	**1**
delta = 0.1, alpha = 1.5	0.41	0.25	0.33	2231	22
delta = 0.3, alpha = 1.5	0.08	0.11	0.05	998	11
delta = 0.5, alpha = 1.5	0.04	0.01	0.04	757	5
delta = 0.7, alpha = 1.5	0.05	0.02	0.05	852	9
delta = 0.9, alpha = 1.5	0.06	0.03	0.06	1035	12
delta = 0.1, alpha = 1.7	0.32	0.11	0.30	2231	23
delta = 0.3, alpha = 1.7	0.05	0.02	0.04	855	10
delta = 0.5, alpha = 1.7	0.04	0.02	0.04	801	6
delta = 0.7, alpha = 1.7	0.06	0.03	0.06	1160	15
delta = 0.9, alpha = 1.7	0.08	0.03	0.09	1482	18
delta = 0.1, alpha = 1.9	0.25	0.14	0.19	2112	21
delta = 0.3, alpha = 1.9	0.07	0.04	0.07	1184	16
delta = 0.5, alpha = 1.9	0.04	0.01	0.04	824	7
delta = 0.7, alpha = 1.9	0.05	0.02	0.05	1042	13
delta = 0.9, alpha = 1.9	0.07	0.01	0.07	1515	19

6.2.3 Results Tuning UH-$(5/2_{DI}, 35)$-σSA-ES on the Noisy Sphere

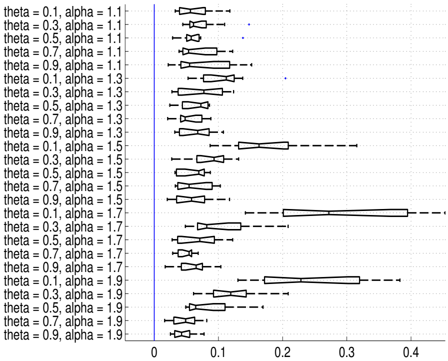


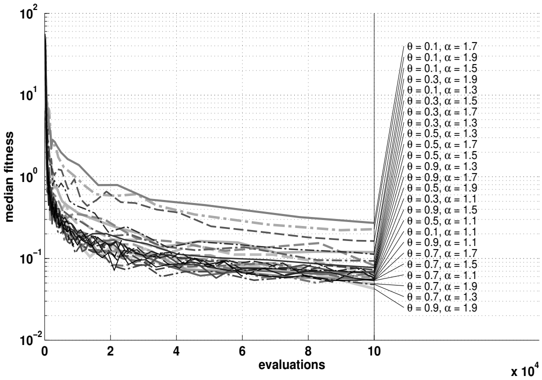


NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
theta = 0.1, alpha = 1.1	2.58	8.00	0.05	872	6
theta = 0.3, alpha = 1.1	0.06	0.02	0.06	944	8
theta = 0.5, alpha = 1.1	4.56	14.26	0.05	996	9
theta = 0.7, alpha = 1.1	0.05	0.03	0.04	730	3
theta = 0.9, alpha = 1.1	**0.05**	**0.02**	**0.04**	**659**	**1**
theta = 0.1, alpha = 1.3	0.08	0.04	0.08	1373	18
theta = 0.3, alpha = 1.3	4.40	13.73	0.05	1102	12
theta = 0.5, alpha = 1.3	0.06	0.02	0.05	1020	10
theta = 0.7, alpha = 1.3	3.01	9.31	0.07	1151	14
theta = 0.9, alpha = 1.3	0.07	0.04	0.06	1084	11
theta = 0.1, alpha = 1.5	0.12	0.04	0.12	1946	22
theta = 0.3, alpha = 1.5	4.72	14.58	0.10	1880	21
theta = 0.5, alpha = 1.5	0.05	0.02	0.05	799	4
theta = 0.7, alpha = 1.5	0.06	0.02	0.06	1154	15
theta = 0.9, alpha = 1.5	0.05	0.02	0.06	901	7
theta = 0.1, alpha = 1.7	0.20	0.12	0.21	1974	24
theta = 0.3, alpha = 1.7	3.38	10.36	0.11	1747	20
theta = 0.5, alpha = 1.7	2.81	8.62	0.09	1685	19
theta = 0.7, alpha = 1.7	0.07	0.01	0.07	1202	16
theta = 0.9, alpha = 1.7	0.05	0.02	0.05	691	2
theta = 0.1, alpha = 1.9	0.25	0.14	0.20	2235	25
theta = 0.3, alpha = 1.9	0.15	0.07	0.14	1946	23
theta = 0.5, alpha = 1.9	4.76	14.82	0.07	1292	17
theta = 0.7, alpha = 1.9	0.07	0.04	0.06	1132	13
theta = 0.9, alpha = 1.9	0.05	0.03	0.06	860	5

6.2.4 Results Tuning UH-CMA-ES on the Noisy Sphere

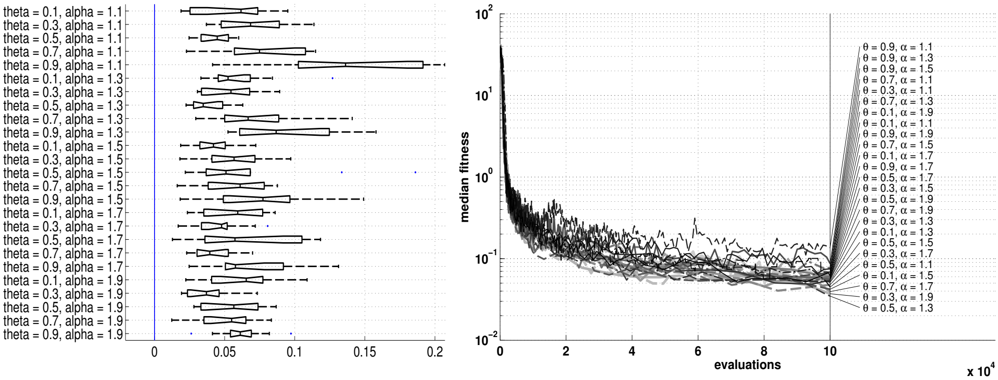




NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
theta = 0.1, alpha = 1.1	0.06	0.03	0.06	994	9
theta = 0.3, alpha = 1.1	0.07	0.03	0.06	1207	16
theta = 0.5, alpha = 1.1	0.06	0.03	0.06	958	6
theta = 0.7, alpha = 1.1	0.07	0.03	0.05	1046	11
theta = 0.9, alpha = 1.1	0.08	0.05	0.06	1118	15
theta = 0.1, alpha = 1.3	0.11	0.04	0.11	1694	21
theta = 0.3, alpha = 1.3	0.17	0.31	0.08	1265	17
theta = 0.5, alpha = 1.3	0.07	0.02	0.07	1101	14
theta = 0.7, alpha = 1.3	0.05	0.02	0.05	827	4
theta = 0.9, alpha = 1.3	0.07	0.03	0.07	1097	13
theta = 0.1, alpha = 1.5	0.18	0.07	0.16	2167	23
theta = 0.3, alpha = 1.5	0.09	0.04	0.09	1435	19
theta = 0.5, alpha = 1.5	0.06	0.02	0.07	978	7
theta = 0.7, alpha = 1.5	0.06	0.03	0.05	998	10
theta = 0.9, alpha = 1.5	0.06	0.03	0.06	912	5
theta = 0.1, alpha = 1.7	0.34	0.24	0.27	2355	25
theta = 0.3, alpha = 1.7	0.10	0.05	0.08	1573	20
theta = 0.5, alpha = 1.7	0.07	0.03	0.07	1081	12
theta = 0.7, alpha = 1.7	0.05	0.01	0.05	715	3
theta = 0.9, alpha = 1.7	0.06	0.03	0.07	979	8
theta = 0.1, alpha = 1.9	0.27	0.14	0.23	2314	24
theta = 0.3, alpha = 1.9	0.12	0.05	0.12	1875	22
theta = 0.5, alpha = 1.9	0.26	0.56	0.06	1385	18
theta = 0.7, alpha = 1.9	0.05	0.02	0.05	693	2
theta = 0.9, alpha = 1.9	**0.05**	**0.02**	**0.04**	**608**	**1**

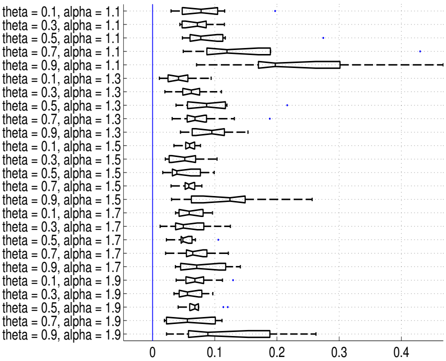
6.2.5 Results Tuning IUH-$(5/2_{DI}, 35)$-σSA-ES on the Noisy Sphere

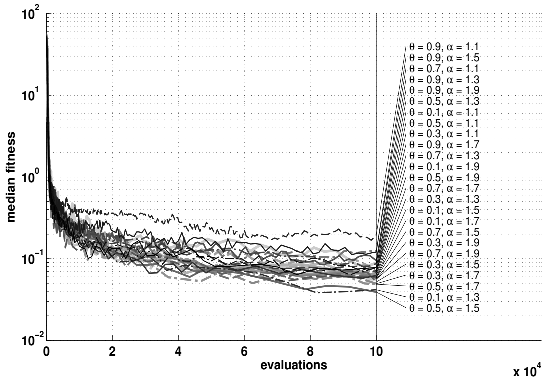


NOISY SPHERE PROBLEM

	Mean	Std	Med	$\sum$#	#
theta = 0.1, alpha = 1.1	0.05	0.03	0.06	1104	7
theta = 0.3, alpha = 1.1	3.17	9.82	0.07	1517	21
theta = 0.5, alpha = 1.1	3.42	10.68	0.04	918	5
theta = 0.7, alpha = 1.1	0.08	0.03	0.07	1610	23
theta = 0.9, alpha = 1.1	0.14	0.06	0.14	2149	25
theta = 0.1, alpha = 1.3	0.06	0.03	0.05	1243	12
theta = 0.3, alpha = 1.3	0.06	0.02	0.05	1168	9
theta = 0.5, alpha = 1.3	0.04	0.01	0.03	673	2
theta = 0.7, alpha = 1.3	0.07	0.04	0.07	1508	20
theta = 0.9, alpha = 1.3	0.09	0.04	0.09	1864	24
theta = 0.1, alpha = 1.5	0.04	0.02	0.04	833	4
theta = 0.3, alpha = 1.5	4.26	13.31	0.06	1288	16
theta = 0.5, alpha = 1.5	0.07	0.05	0.05	1211	10
theta = 0.7, alpha = 1.5	0.06	0.02	0.06	1258	13
theta = 0.9, alpha = 1.5	0.08	0.04	0.08	1572	22
theta = 0.1, alpha = 1.7	0.06	0.02	0.06	1239	11
theta = 0.3, alpha = 1.7	0.05	0.02	0.05	950	6
theta = 0.5, alpha = 1.7	3.51	10.91	0.06	1336	17
theta = 0.7, alpha = 1.7	0.04	0.01	0.04	744	3
theta = 0.9, alpha = 1.7	0.07	0.03	0.06	1455	19
theta = 0.1, alpha = 1.9	0.06	0.03	0.07	1272	14
theta = 0.3, alpha = 1.9	**0.04**	**0.02**	**0.04**	**661**	**1**
theta = 0.5, alpha = 1.9	3.35	10.43	0.06	1277	15
theta = 0.7, alpha = 1.9	3.80	11.85	0.06	1149	8
theta = 0.9, alpha = 1.9	0.06	0.02	0.06	1376	18

6.2.6 Results Tuning IUH-CMA-ES on the Noisy Sphere





NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
theta = 0.1, alpha = 1.1	0.08	0.05	0.08	1367	18
theta = 0.3, alpha = 1.1	0.07	0.02	0.07	1275	13
theta = 0.5, alpha = 1.1	0.10	0.07	0.08	1524	19
theta = 0.7, alpha = 1.1	0.15	0.11	0.12	1947	24
theta = 0.9, alpha = 1.1	0.23	0.13	0.20	2241	25
theta = 0.1, alpha = 1.3	**0.04**	**0.02**	**0.04**	**578**	**1**
theta = 0.3, alpha = 1.3	0.06	0.03	0.06	1099	11
theta = 0.5, alpha = 1.3	0.10	0.05	0.09	1532	20
theta = 0.7, alpha = 1.3	0.08	0.05	0.07	1348	17
theta = 0.9, alpha = 1.3	0.14	0.18	0.10	1610	22
theta = 0.1, alpha = 1.5	0.06	0.01	0.06	1006	8
theta = 0.3, alpha = 1.5	0.05	0.03	0.05	807	3
theta = 0.5, alpha = 1.5	0.05	0.03	0.04	734	2
theta = 0.7, alpha = 1.5	0.06	0.01	0.06	985	6
theta = 0.9, alpha = 1.5	0.12	0.07	0.12	1704	23
theta = 0.1, alpha = 1.7	0.06	0.02	0.06	1043	9
theta = 0.3, alpha = 1.7	0.06	0.03	0.05	935	5
theta = 0.5, alpha = 1.7	0.06	0.02	0.05	842	4
theta = 0.7, alpha = 1.7	0.07	0.03	0.06	1211	12
theta = 0.9, alpha = 1.7	0.12	0.14	0.07	1322	15
theta = 0.1, alpha = 1.9	0.07	0.03	0.07	1295	14
theta = 0.3, alpha = 1.9	0.10	0.15	0.06	1075	10
theta = 0.5, alpha = 1.9	0.07	0.02	0.07	1335	16
theta = 0.7, alpha = 1.9	0.06	0.04	0.06	994	7
theta = 0.9, alpha = 1.9	0.15	0.15	0.09	1566	21

6.3 Adaptive Versus Non-Adaptive Averaging

Given the tuned adaptive averaging techniques, the final question that we aim to answer is whether adaptive averaging is better than explicit resampling, implicit averaging, or even a canonical implementation. For the algorithmic schemes considered in this work we will attempt to answer this question based on empirical comparison on a number of artificial test problems in the context of Gaussian additive noise.

Table 6.2 shows the general setup adopted in the experiments. Table 6.3 shows the set of test problems used in the comparison. This set is partially based on the test problems used in [SHL$^+$05, HFRA09b, HFRA09a, HFRA10]. Full descriptions of these test problems are given in Appendix A. Table 6.4 shows the noise handling schemes that are considered for comparison. For implicit and explicit averaging we consider the optimal sample size or population size for each test problem, therewith aiming to compare the adaptive averaging techniques to optimally tuned non-adaptive schemes. For this, Section 6.3.1 and Section 6.3.2 present the results of an empirical study to find for each test problem the optimal sample size or population size. In Section 6.3.3 these results are used for the full empirical comparison.

General experimental settings	
Search space dimension size	$n = 10$
Evaluation budget per run	10,000
Runs per algorithmic scheme	100
Performance indicators	Final solution quality w.r.t. the underlying signal function (mean, std, median) over all runs, and rank sum for ranking of the algorithmic schemes

Table 6.2: The general experimental setup.

Test problem	**Properties of the underlying signal function**		
Noisy Sphere Problem	unimodal	separable	well-conditioned
Noisy Ellipsoid Problem	unimodal	non-separable	ill-conditioned
Noisy Step Ellipsoid Problem	unimodal	non-separable	ill-conditioned
Noisy Rosenbrock Problem	unimodal	non-separable	well-conditioned
Noisy Ackley Problem	multimodal	separable	well-conditioned
Noisy Griewank Problem	multimodal	separable	well-conditioned
Noisy Rastrigin Problem	multimodal	separable	well-conditioned
Noisy Schaffer's F7 Problem	multimodal	separable	well-conditioned
Noisy Branke's Multipeak Problem	multimodal	separable	well-conditioned
Noisy Keane's Bump Problem	multimodal	non-separable	well-conditioned

Table 6.3: The test problems used for empirical comparison.

Noise handling schemes used for empirical comparison	
Canonical	A canonical $(5/2_{DI}, 35)$-σSA-ES and CMA-ES.
MEM	Explicit resampling, optimally tuned for each test problem.
MPM	Implicit averaging, optimally tuned for each test problem.
PUH	The poset-based adaptive averaging method.
UH	The rank-based adaptive averaging method.
IUH	The inversions-based adaptive averaging method.

Table 6.4: The techniques considered in the empirical study on noise handling schemes.

6.3.1 The Optimal Sample Size for Explicit Averaging

This experiment is done in order to determine, for each test problem, the optimal sample size for explicit averaging. Different instances of the MEM-$(5/2_{DI}, 35)$-σSA-ES and the MEM-CMA-ES are considered with varying sample sizes: $m = 2, 4, \ldots, 16$. These sample sizes are compared on the test problems listed in Table 6.3 using the experimental setup shown in Table 6.2. The results of these experiments are shown in the tables and figures of Section 6.3.1.1 and Section 6.3.1.2 for the MEM-$(5/2_{DI}, 35)$-σSA-ES and the MEM-CMA-ES respectively.

The results show the trade-off between taking too few samples, leading to early stagnation, and too many samples, leading to slow convergence. In between lies an optimal sample size for the considered evaluation budget of $10,000$ function evaluations. For the explicit averaging schemes for each of the test problems with respect to the general experimental setup the optimal sample sizes lie, with an approximate error of ± 1, at the values shown in Table 6.5. From this table we see that the sample sizes for the MEM-$(5/2_{DI}, 35)$-σSA-ES are generally low and for the MEM-CMA-ES it seems that slightly higher sample sizes should be used.

Test problem	MEM-$(5, 35)$-σSA-ES	MEM-CMA-ES
Noisy Sphere Problem	m = 4	m = 12
Noisy Ellipsoid Problem	m = 4	m = 8
Noisy Step Ellipsoid Problem	m = 2	m = 2
Noisy Rosenbrock Problem	m = 2	m = 4
Noisy Ackley Problem	m = 4	m = 6
Noisy Griewank Problem	m = 6	m = 14
Noisy Rastrigin Problem	m = 2	m = 2
Noisy Schaffer's F7 Problem	m = 2	m = 4
Noisy Branke's Multipeak Problem	m = 4	m = 2
Noisy Keane's Bump Problem	m = 4	m = 2

Table 6.5: The optimal sample size for the MEM approach with an approximate error of ± 1 to achieve best convergence accuracy on a budget of $10,000$ function evaluations.

6.3.1.1 Results MEM-$(5/2_{DI}, 35)$-σSA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
m = 2	2.65	9.46	0.37	27616	3
m = 4	**1.12**	**5.45**	**0.31**	**21548**	**1**
m = 6	1.30	6.54	0.33	22462	2
m = 8	2.77	8.84	0.41	31286	4
m = 10	2.85	8.96	0.54	40452	5
m = 12	2.52	5.56	0.88	50776	6
m = 14	7.33	12.70	1.86	60611	7
m = 16	5.61	8.47	2.77	65649	8

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.95	4.00	0.85	22374	2
m = 4	**1.86**	**4.32**	**0.81**	**20584**	**1**
m = 6	3.09	5.93	0.97	27054	3
m = 8	2.63	3.71	1.37	33075	4
m = 10	4.68	5.38	2.23	46163	5
m = 12	5.54	5.33	3.91	52517	6
m = 14	6.72	6.13	4.58	56617	7
m = 16	8.87	7.30	5.90	62016	8

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**1.18**	**3.77**	**0.00**	**13133**	**1**
m = 4	1.97	4.52	1.00	23326	2
m = 6	1.19	2.56	1.00	24521	3
m = 8	3.28	5.48	1.00	37849	4
m = 10	3.67	4.35	2.00	45768	5
m = 12	5.39	5.32	4.00	54037	6
m = 14	6.12	5.05	4.00	57872	7
m = 16	8.39	6.02	7.00	63894	8

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**8.63**	**1.10**	**8.74**	**15838**	**1**
m = 4	8.79	1.15	8.91	17670	2
m = 6	9.11	1.00	9.10	20403	3
m = 8	12.04	8.18	10.16	31062	4
m = 10	19.52	14.37	14.22	45578	5
m = 12	38.13	30.13	27.73	57724	6
m = 14	53.66	46.84	39.92	62745	7
m = 16	87.22	75.31	66.17	69380	8

NOISY ACKLEY PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.18	1.99	0.60	18224	2
m = 4	**1.13**	**1.93**	**0.66**	**18101**	**1**
m = 6	1.23	1.54	0.77	22770	3
m = 8	2.23	2.01	1.53	38075	4
m = 10	2.69	1.65	2.27	45781	5
m = 12	3.24	1.54	3.09	52847	6
m = 14	3.82	1.25	3.58	59927	7
m = 16	4.43	1.42	4.01	64675	8

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.29	0.40	1.20	44960	6
m = 4	1.27	0.44	1.15	37310	4
m = 6	**1.19**	**0.27**	**1.14**	**33226**	**1**
m = 8	1.18	0.19	1.13	33330	2
m = 10	1.22	0.32	1.15	36458	3
m = 12	1.23	0.29	1.17	39513	5
m = 14	1.25	0.23	1.20	45903	7
m = 16	1.36	0.38	1.21	49700	8

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**27.79**	**37.93**	**12.06**	**16549**	**1**
m = 4	49.94	42.88	31.42	26768	2
m = 6	66.15	33.99	70.44	32571	3
m = 8	82.83	22.06	78.51	41122	4
m = 10	90.65	23.09	87.51	47240	5
m = 12	94.16	19.57	93.40	50716	7
m = 14	93.36	19.08	89.77	50053	6
m = 16	100.06	20.33	100.72	55381	8

NOISY SCHAFFERS F7 PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**6.11**	**13.90**	**2.40**	**9634**	**1**
m = 4	11.05	16.89	4.72	20150	2
m = 6	18.86	17.13	12.19	31284	3
m = 8	22.65	13.39	18.31	37013	4
m = 10	32.46	13.82	26.99	49204	5
m = 12	37.37	14.42	34.08	54709	6
m = 14	39.45	12.61	35.54	57584	7
m = 16	42.26	11.18	41.26	60822	8

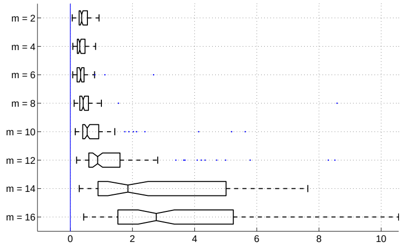
NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	0.37	0.10	0.34	30149	3
m = 4	**0.35**	**0.06**	**0.33**	**23717**	**1**
m = 6	0.35	0.07	0.33	27505	2
m = 8	0.38	0.09	0.34	34868	4
m = 10	0.41	0.09	0.37	44727	5
m = 12	0.42	0.09	0.38	47615	6
m = 14	0.44	0.09	0.42	53656	7
m = 16	0.46	0.09	0.44	58163	8

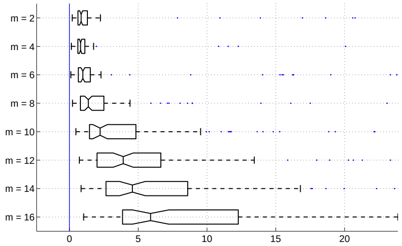
NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	∑#	#
m = 2	-0.59	0.14	-0.64	17807	2
m = 4	**-0.60**	**0.07**	**-0.61**	**17351**	**1**
m = 6	-0.55	0.10	-0.56	26485	3
m = 8	-0.48	0.11	-0.52	37184	4
m = 10	-0.43	0.11	-0.45	46214	5
m = 12	-0.39	0.09	-0.39	53539	6
m = 14	-0.35	0.08	-0.35	59281	7
m = 16	-0.33	0.09	-0.33	62539	8

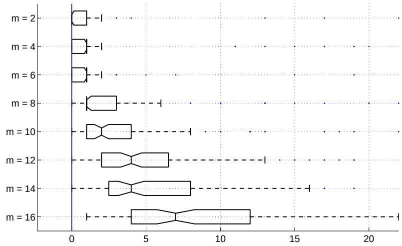
NOISY SPHERE PROBLEM



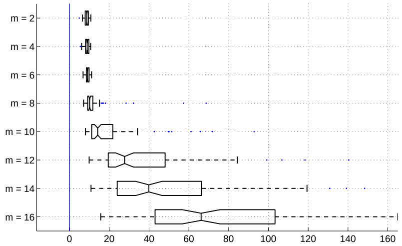
NOISY ELLIPSOID PROBLEM



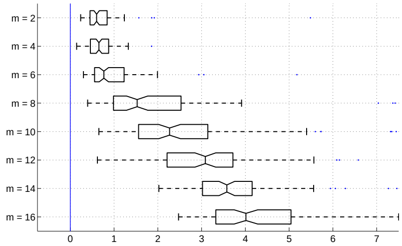
NOISY STEP ELLIPSOID PROBLEM



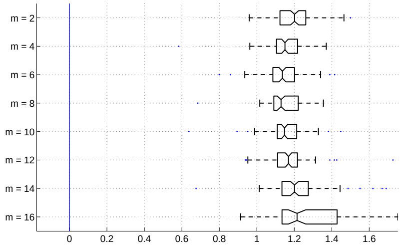
NOISY ROSENBROCK PROBLEM



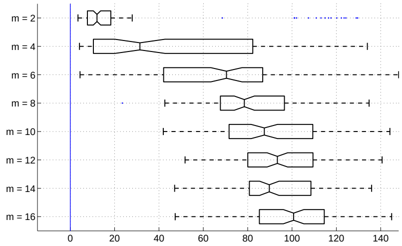
NOISY ACKLEY PROBLEM



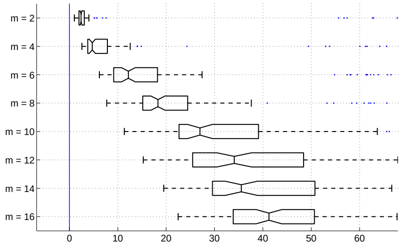
NOISY GRIEWANK PROBLEM



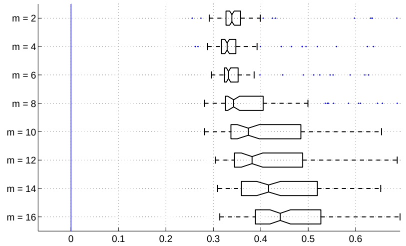
NOISY RASTRIGIN PROBLEM



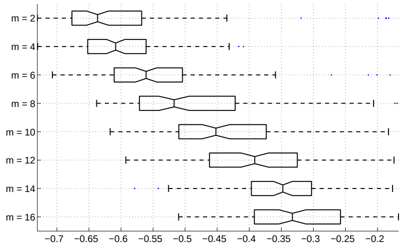
NOISY SCHAFFER'S F7 PROBLEM



NOISY BRANKE'S MULTIPEAK PROBLEM



NOISY KEANE'S BUMP PROBLEM



6.3.1.2 Results MEM-CMA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
m = 2	0.44	0.18	0.42	61210	8
m = 4	0.32	0.14	0.30	49041	7
m = 6	0.85	6.02	0.23	37108	6
m = 8	2.46	9.91	0.20	36067	5
m = 10	0.71	4.74	0.22	34633	3
m = 12	**1.36**	**7.88**	**0.21**	**32802**	**1**
m = 14	1.06	5.81	0.21	33597	2
m = 16	1.95	9.80	0.21	35942	4

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.31	1.70	1.07	50014	8
m = 4	1.14	2.53	0.86	38460	5
m = 6	1.55	5.31	0.70	33460	2
m = 8	**1.80**	**4.81**	**0.67**	**31769**	**1**
m = 10	1.87	4.37	0.75	36178	3
m = 12	2.35	7.13	0.78	37697	4
m = 14	3.51	9.52	0.88	44445	6
m = 16	3.86	7.93	1.11	48377	7

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**1.27**	**3.44**	**0.00**	**22837**	**1**
m = 4	1.05	1.18	1.00	29963	3
m = 6	1.50	3.97	1.00	29819	2
m = 8	2.19	8.57	1.00	35177	4
m = 10	2.73	6.74	1.00	45646	5
m = 12	2.11	4.29	1.00	46145	6
m = 14	3.36	6.47	1.00	53768	7
m = 16	5.04	8.48	2.00	57045	8

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	38.99	261.44	9.63	31838	2
m = 4	**71.43**	**408.26**	**9.25**	**28336**	**1**
m = 6	86.79	404.63	9.52	32392	3
m = 8	217.35	639.96	9.78	38770	5
m = 10	210.64	636.02	9.54	36780	4
m = 12	283.76	832.69	9.98	43746	6
m = 14	310.96	766.21	12.08	50537	7
m = 16	431.76	1015.12	20.58	58001	8

NOISY ACKLEY PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.10	0.68	0.86	56409	8
m = 4	0.90	1.23	0.53	43022	6
m = 6	**0.69**	**1.14**	**0.39**	**29784**	**1**
m = 8	0.90	1.80	0.42	30780	2
m = 10	0.78	1.49	0.41	31160	3
m = 12	0.63	0.80	0.47	34489	4
m = 14	0.72	0.61	0.53	40276	5
m = 16	1.24	1.31	0.81	54480	7

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	1.08	0.32	1.17	46503	7
m = 4	1.11	0.18	1.14	47223	8
m = 6	1.09	0.20	1.13	45640	6
m = 8	1.14	0.32	1.10	40570	5
m = 10	1.08	0.13	1.10	35794	4
m = 12	1.13	0.23	1.09	35078	3
m = 14	**1.09**	**0.08**	**1.09**	**34535**	**1**
m = 16	1.15	0.30	1.08	35057	2

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**16.09**	**14.75**	**13.54**	**17294**	**1**
m = 4	21.95	25.06	14.75	21551	2
m = 6	26.91	25.97	16.50	24969	3
m = 8	46.89	36.19	28.91	37678	4
m = 10	64.43	35.16	67.65	48504	5
m = 12	71.29	28.52	77.92	53043	6
m = 14	81.88	22.42	82.71	59232	8
m = 16	81.21	21.02	79.88	58129	7

NOISY SCHAFFERS F7 PROBLEM

	Mean	Std	Med	∑#	#
m = 2	5.06	6.53	3.72	20228	3
m = 4	**5.37**	**8.95**	**3.30**	**19009**	**1**
m = 6	4.62	3.40	3.30	20120	2
m = 8	9.31	12.27	5.99	33921	4
m = 10	12.20	11.78	8.91	44711	5
m = 12	16.42	13.37	12.14	54832	6
m = 14	19.27	13.63	15.17	60625	7
m = 16	21.14	8.90	19.39	66954	8

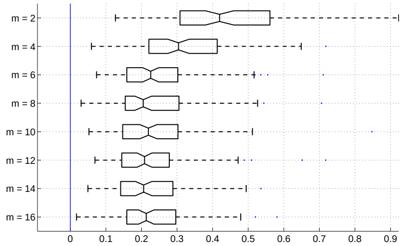
NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**0.36**	**0.11**	**0.33**	**31570**	**1**
m = 4	0.40	0.16	0.33	33160	2
m = 6	0.44	0.18	0.33	36489	3
m = 8	0.44	0.16	0.34	37489	4
m = 10	0.45	0.17	0.35	39659	5
m = 12	0.48	0.16	0.47	44608	6
m = 14	0.50	0.15	0.54	48942	8
m = 16	0.50	0.15	0.53	48483	7

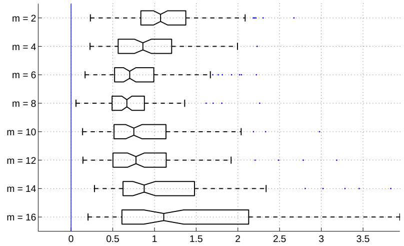
NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	∑#	#
m = 2	**-0.42**	**0.17**	**-0.47**	**27358**	**1**
m = 4	-0.41	0.16	-0.44	27921	2
m = 6	-0.37	0.16	-0.35	34292	3
m = 8	-0.34	0.15	-0.33	37736	4
m = 10	-0.29	0.13	-0.27	45434	5
m = 12	-0.27	0.11	-0.26	48101	7
m = 14	-0.28	0.10	-0.28	45571	6
m = 16	-0.24	0.09	-0.23	53987	8

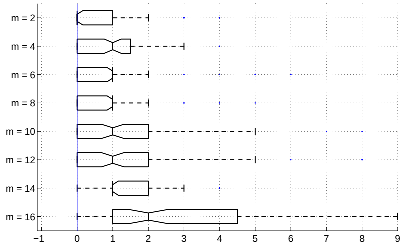
NOISY SPHERE PROBLEM



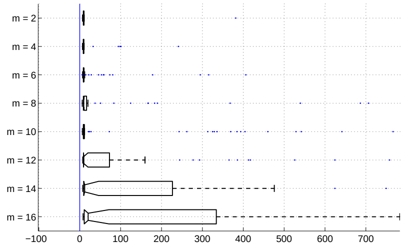
NOISY ELLIPSOID PROBLEM



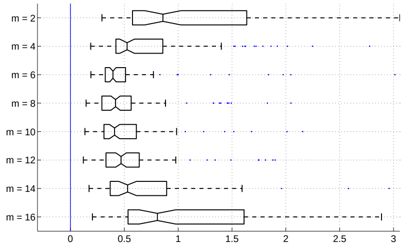
NOISY STEP ELLIPSOID PROBLEM



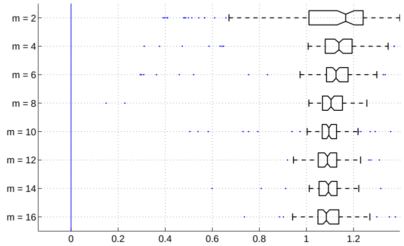
NOISY ROSENBROCK PROBLEM



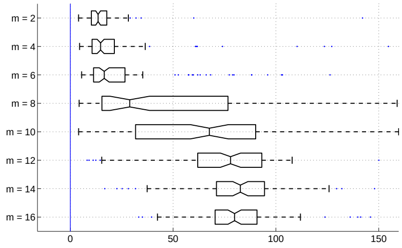
NOISY ACKLEY PROBLEM



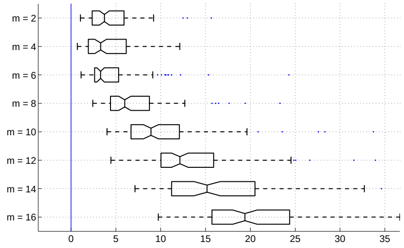
NOISY GRIEWANK PROBLEM



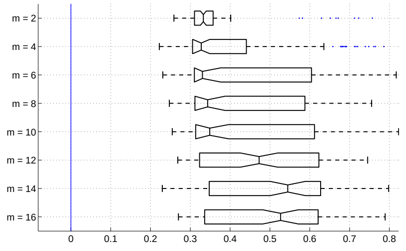
NOISY RASTRIGIN PROBLEM



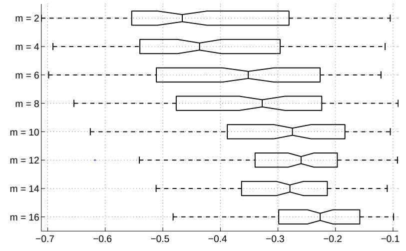
NOISY SCHAFFER'S F7 PROBLEM



NOISY BRANKE'S MULTIPEAK PROBLEM



NOISY KEANE'S BUMP PROBLEM



6.3.2 The Optimal Sample Size for Implicit Averaging

This experiment is done in order to determine, for each test problem, the optimal population size for implicit averaging. Different instances of the $(\mu/2_{DI}, \lambda)$-σSA-ES and the CMA-ES with varying population sizes are considered. For the $(\mu/2_{DI}, \lambda)$-σSA-ES: $(5, 35)$, $(10, 70)$, $(20, 140)$, $(30, 210)$, $(40, 280)$, $(50, 350)$, $(60, 420)$. For the CMA-ES: $(5, 10)$, $(10, 20)$, $(20, 40)$, $(30, 60)$, $(40, 80)$, $(50, 100)$, $(60, 120)$, $(70, 140)$. These different population sizes are compared on the test problems listed in Table 6.3 using the experimental setup as displayed in Table 6.2. The results of these experiments are shown in the tables and figures of Section 6.3.2.1 and Section 6.3.2.2 for the MPM-$(\mu/2_{DI}, \lambda)$-σSA-ES and the MPM-CMA-ES respectively.

Similar to the results of explicit averaging, the results show the trade-off that exists between taking too small population sizes that leads to early stagnation and too high population sizes that leads to slow convergence. In between lies an optimal population size for the considered evaluation budget of $10,000$ function evaluations. Based on these results, we can conclude that for the implicit averaging schemes, the MPM-$(\mu/2_{DI}, \lambda)$-σSA-ES and the MPM-CMA-ES, given the experimental setup, the optimal sample sizes lie at the values shown in Table 6.6. When taking α as the scaling factor of the default population sizes for both schemes, the error of these results is ± 10 for μ, maintaining the default ratios between μ and λ. From Table 6.6 we observe that for the MPM-$(\mu/2_{DI}, \lambda)$-σSA-ES, the implicit averaging factor is generally low and for the CMA-ES, slightly higher factors are optimal. This observation is similar to what is observed in the experiments of Section 6.3.1 in the comparison of different sample size for explicit averaging.

Test problem	MPM-$(\mu/2_{DI}, \lambda)$-σSA-ES	MPM-CMA-ES
Noisy Sphere Problem	(10,70)	(50,100)
Noisy Ellipsoid Problem	(20,140)	(40,80)
Noisy Step Ellipsoid Problem	(20,140)	(30,60)
Noisy Rosenbrock Problem	(20,140)	(40,80)
Noisy Ackley Problem	(10,70)	(40,80)
Noisy Griewank Problem	(20,140)	(40,80)
Noisy Rastrigin Problem	(20,140)	(30,60)
Noisy Schaffer's F7 Problem	(10,70)	(20,40)
Noisy Branke's Multipeak Problem	(20,140)	(30,60)
Noisy Keane's Bump Problem	(10,70)	(30,60)

Table 6.6: For each multi-population scheme for each test problem the optimal population size to achieve best convergence accuracy on a budget of $10,000$ function evaluations. When taking α as the scaling factor of the default population sizes for both schemes, the error of these results is ± 10 for μ, maintaining the default ratios between μ and λ.

6.3.2.1 Results MPM-$(\mu/2_{DI}, \lambda)$-σSA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	2.80	8.95	0.59	40598	5
(10,70)	**0.41**	**0.19**	**0.37**	**23013**	**1**
(20,140)	0.42	0.23	0.35	23713	2
(30,210)	0.52	0.23	0.51	33364	3
(40,280)	0.59	0.30	0.52	36746	4
(50,350)	0.68	0.31	0.65	44069	6
(60,420)	0.87	0.34	0.84	55173	7
(70,490)	1.14	0.47	1.09	63724	8

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	1.88	3.64	1.02	31179	4
(10,70)	1.40	3.05	0.94	25903	2
(20,140)	**1.07**	**1.12**	**0.82**	**25236**	**1**
(30,210)	1.09	0.57	0.97	28911	3
(40,280)	1.49	0.69	1.41	41165	5
(50,350)	1.74	0.74	1.73	47970	6
(60,420)	2.17	0.96	2.11	55671	7
(70,490)	2.85	1.23	2.56	64365	8

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	1.91	5.25	1.00	23880	3
(10,70)	1.16	3.44	0.00	23053	2
(20,140)	**0.38**	**0.58**	**0.00**	**21802**	**1**
(30,210)	0.58	0.73	0.00	30540	4
(40,280)	1.00	0.92	1.00	41283	5
(50,350)	1.46	1.04	1.00	51105	6
(60,420)	2.12	1.20	2.00	61149	7
(70,490)	2.68	1.38	3.00	67588	8

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	9.36	6.42	8.70	26691	3
(10,70)	8.55	0.99	8.57	22443	2
(20,140)	**8.54**	**0.75**	**8.53**	**21091**	**1**
(30,210)	8.91	0.80	8.84	28526	4
(40,280)	9.38	0.83	9.20	36718	5
(50,350)	10.17	1.00	10.19	48650	6
(60,420)	12.00	1.46	12.17	64043	7
(70,490)	14.10	2.19	13.98	72238	8

NOISY ACKLEY PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	1.98	2.76	0.85	24915	3
(10,70)	**1.27**	**2.06**	**0.65**	**15895**	**1**
(20,140)	1.03	0.80	0.87	18688	2
(30,210)	1.72	0.51	1.71	32300	4
(40,280)	2.66	0.72	2.67	45007	5
(50,350)	3.28	0.61	3.27	54289	6
(60,420)	3.85	0.58	3.85	63504	7
(70,490)	4.04	0.66	4.01	65802	8

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	1.40	0.46	1.28	46229	7
(10,70)	1.21	0.11	1.20	32629	3
(20,140)	**1.19**	**0.11**	**1.18**	**28874**	**1**
(30,210)	1.21	0.10	1.19	31370	2
(40,280)	1.26	0.11	1.25	42578	5
(50,350)	1.26	0.13	1.25	41984	4
(60,420)	1.28	0.15	1.27	45009	6
(70,490)	1.33	0.16	1.32	51727	8

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	25.66	36.95	11.27	33009	4
(10,70)	15.12	29.17	5.75	20476	2
(20,140)	**7.16**	**16.56**	**3.94**	**10749**	**1**
(30,210)	20.52	19.30	12.32	30308	3
(40,280)	44.92	12.00	46.97	48249	5
(50,350)	51.65	9.00	51.81	54729	6
(60,420)	56.01	10.28	56.51	59848	7
(70,490)	58.29	7.27	57.89	63032	8

NOISY SCHAFFERS F7 PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	5.12	11.47	2.28	15876	2
(10,70)	**2.68**	**5.17**	**2.08**	**10141**	**1**
(20,140)	3.44	0.65	3.45	22667	3
(30,210)	6.22	1.20	6.23	34166	4
(40,280)	10.26	1.69	10.26	44882	5
(50,350)	14.33	2.23	13.99	55410	6
(60,420)	17.77	2.44	17.73	64592	7
(70,490)	21.17	2.68	21.10	72666	8

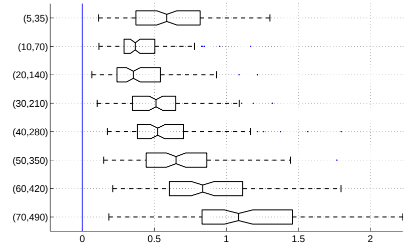
NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	0.38	0.09	0.35	31124	4
(10,70)	0.36	0.06	0.34	25428	2
(20,140)	**0.34**	**0.02**	**0.34**	**20688**	**1**
(30,210)	0.36	0.03	0.35	29870	3
(40,280)	0.38	0.05	0.37	40128	5
(50,350)	0.41	0.06	0.40	52021	6
(60,420)	0.43	0.06	0.42	59024	7
(70,490)	0.45	0.06	0.45	62117	8

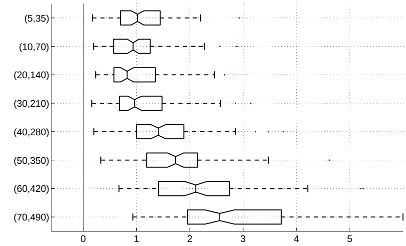
NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	∑#	#
(5,35)	-0.53	0.19	-0.59	37524	4
(10,70)	**-0.67**	**0.07**	**-0.68**	**14122**	**1**
(20,140)	-0.66	0.07	-0.67	16270	2
(30,210)	-0.61	0.09	-0.62	28437	3
(40,280)	-0.54	0.09	-0.56	42853	5
(50,350)	-0.48	0.07	-0.49	53171	6
(60,420)	-0.42	0.07	-0.42	61586	7
(70,490)	-0.39	0.06	-0.39	66437	8

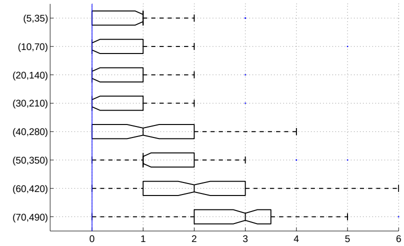
NOISY SPHERE PROBLEM



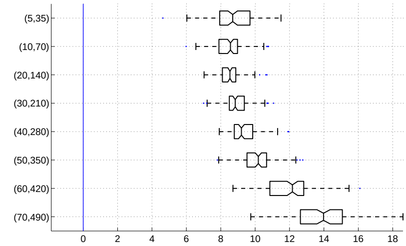
NOISY ELLIPSOID PROBLEM



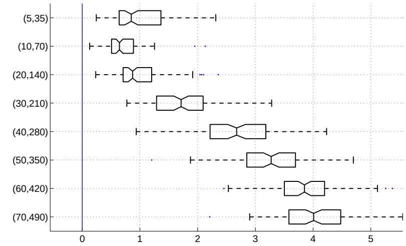
NOISY STEP ELLIPSOID PROBLEM



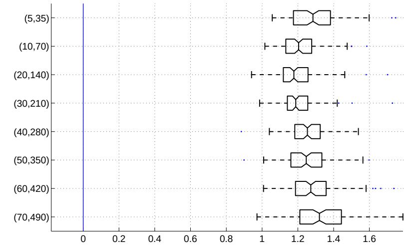
NOISY ROSENBROCK PROBLEM



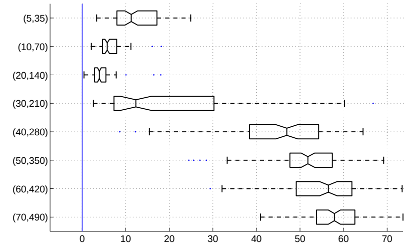
NOISY ACKLEY PROBLEM



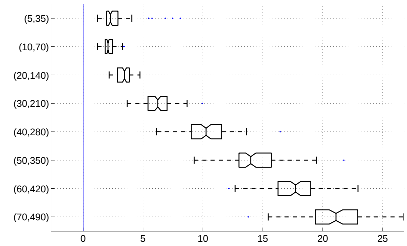
NOISY GRIEWANK PROBLEM



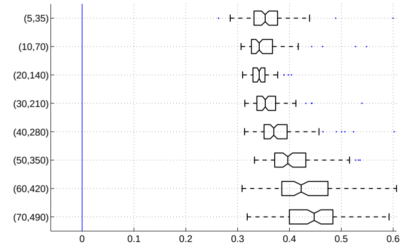
NOISY RASTRIGIN PROBLEM



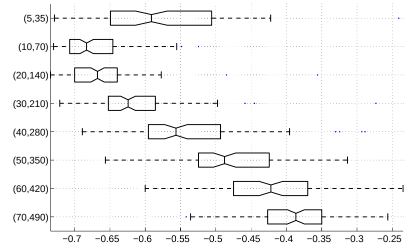
NOISY SCHAFFER'S F7 PROBLEM



NOISY BRANKE'S MULTIPEAK PROBLEM



NOISY KEANE'S BUMP PROBLEM



6.3.2.2 Results MPM-CMA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	0.64	0.30	0.57	72163	8
(10,20)	0.32	0.13	0.30	58550	7
(20,40)	0.41	2.24	0.16	39097	6
(30,60)	0.15	0.08	0.13	30397	3
(40,80)	1.30	5.95	0.13	31001	4
(50,100)	**0.13**	**0.07**	**0.12**	**27863**	**1**
(60,120)	0.14	0.09	0.13	29096	2
(70,140)	0.16	0.11	0.13	32233	5

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	1.88	4.43	1.29	67558	8
(10,20)	0.98	0.47	0.83	61041	7
(20,40)	1.24	3.51	0.45	42527	6
(30,60)	0.50	1.43	0.27	27724	2
(40,80)	**0.35**	**0.24**	**0.26**	**26619**	**1**
(50,100)	0.40	0.29	0.34	31036	4
(60,120)	0.42	0.44	0.31	29705	3
(70,140)	0.49	0.58	0.38	34190	5

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	1.25	1.34	1.00	42626	6
(10,20)	0.78	1.15	0.00	33268	2
(20,40)	0.93	2.27	0.00	35117	3
(30,60)	**0.66**	**2.56**	**0.00**	**30781**	**1**
(40,80)	0.39	0.94	0.00	35860	4
(50,100)	0.35	0.95	0.00	40772	5
(60,120)	0.19	0.60	0.00	46332	7
(70,140)	0.33	0.77	0.00	55644	8

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	72.01	616.49	9.76	54698	8
(10,20)	9.26	0.77	9.31	45337	6
(20,40)	36.26	201.84	9.07	40242	4
(30,60)	14.37	56.67	8.65	29441	2
(40,80)	**8.54**	**0.63**	**8.50**	**24091**	**1**
(50,100)	8.91	0.95	8.80	32494	3
(60,120)	9.16	0.78	9.07	40896	5
(70,140)	10.03	1.79	9.57	53201	7

NOISY ACKLEY PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	1.98	1.17	1.89	71112	8
(10,20)	0.74	0.89	0.53	50352	6
(20,40)	0.60	1.56	0.29	27570	3
(30,60)	0.56	1.50	0.25	23687	2
(40,80)	**0.27**	**0.10**	**0.25**	**20831**	**1**
(50,100)	0.40	0.52	0.31	29462	4
(60,120)	0.72	1.13	0.47	45430	5
(70,140)	0.73	0.58	0.55	51956	7

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	1.27	0.18	1.27	64831	8
(10,20)	1.10	0.20	1.14	54192	7
(20,40)	1.02	0.20	1.07	34586	5
(30,60)	1.05	0.17	1.06	33716	4
(40,80)	**1.05**	**0.08**	**1.05**	**31175**	**1**
(50,100)	1.08	0.17	1.05	32149	2
(60,120)	1.06	0.06	1.06	32849	3
(70,140)	1.10	0.20	1.06	36902	6

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	14.92	5.02	14.79	56521	7
(10,20)	7.56	2.83	6.99	41094	5
(20,40)	7.32	16.49	4.31	27918	3
(30,60)	**6.12**	**14.81**	**3.29**	**20923**	**1**
(40,80)	9.16	18.17	3.66	25880	2
(50,100)	13.06	16.23	4.54	35720	4
(60,120)	19.77	17.00	13.59	49587	6
(70,140)	30.88	20.04	30.56	62757	8

NOISY SCHAFFERS F7 PROBLEM

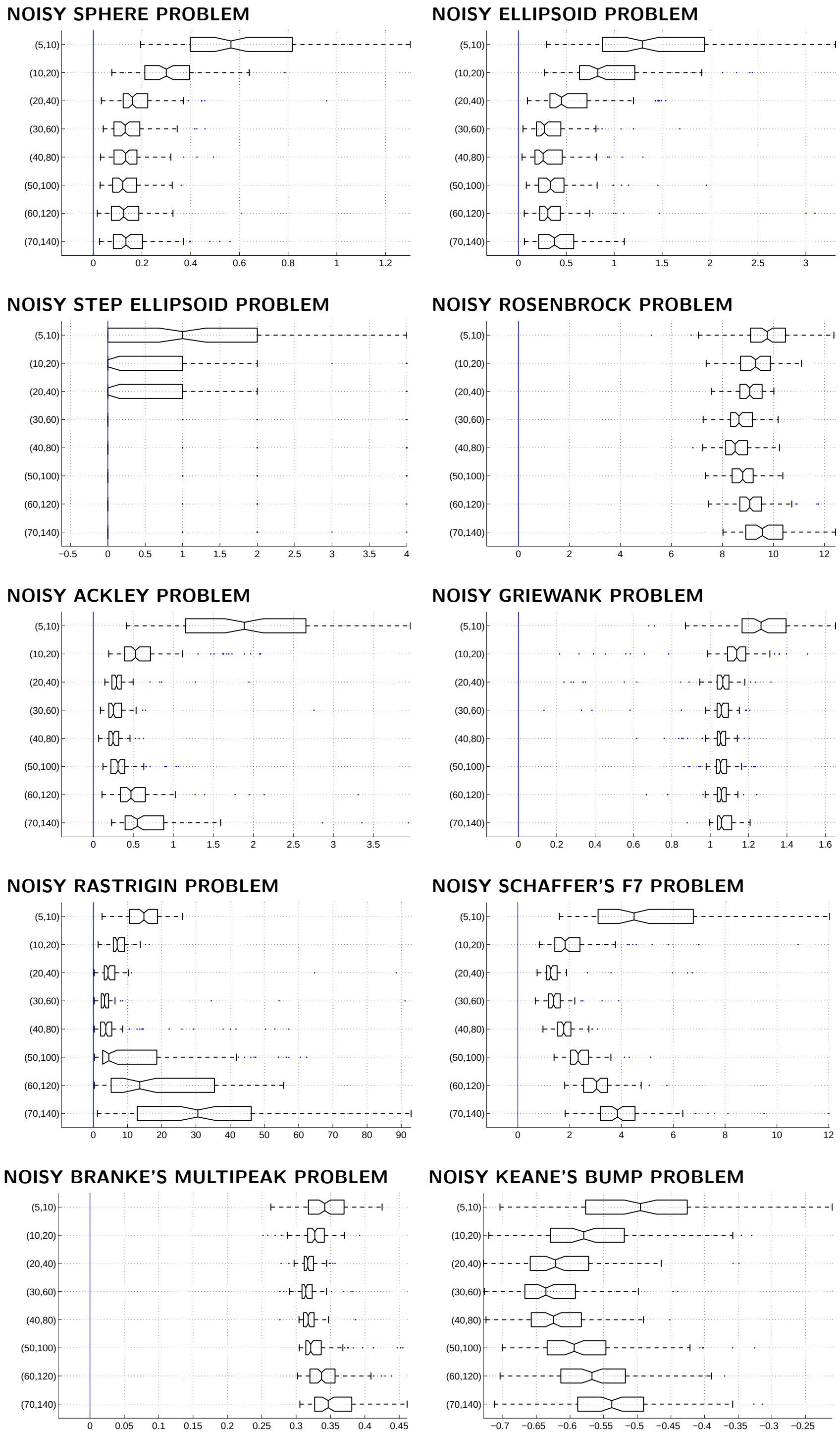
	Mean	Std	Med	∑#	#
(5,10)	5.91	6.55	4.49	64893	8
(10,20)	2.17	1.35	1.82	31858	4
(20,40)	**2.83**	**8.60**	**1.27**	**16479**	**1**
(30,60)	2.02	5.77	1.38	17063	2
(40,80)	2.58	5.96	1.77	28836	3
(50,100)	2.42	0.60	2.33	42748	5
(60,120)	3.09	0.70	3.05	54625	6
(70,140)	4.12	1.51	3.84	63898	7

NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	0.37	0.11	0.34	49302	7
(10,20)	0.34	0.06	0.33	40695	5
(20,40)	0.33	0.07	0.32	30137	3
(30,60)	**0.33**	**0.06**	**0.31**	**25950**	**1**
(40,80)	0.32	0.04	0.32	30013	2
(50,100)	0.33	0.03	0.32	38527	4
(60,120)	0.34	0.04	0.34	48637	6
(70,140)	0.36	0.05	0.35	57139	8

NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	∑#	#
(5,10)	-0.48	0.13	-0.50	57303	8
(10,20)	-0.54	0.15	-0.58	43049	5
(20,40)	-0.60	0.10	-0.62	30184	3
(30,60)	**-0.62**	**0.08**	**-0.64**	**25179**	**1**
(40,80)	-0.62	0.05	-0.62	28369	2
(50,100)	-0.58	0.08	-0.59	38621	4
(60,120)	-0.56	0.07	-0.57	45886	6
(70,140)	-0.54	0.08	-0.54	51809	7


NOISY SPHERE PROBLEM
NOISY ELLIPSOID PROBLEM
NOISY STEP ELLIPSOID PROBLEM
NOISY ROSENBROCK PROBLEM
NOISY ACKLEY PROBLEM
NOISY GRIEWANK PROBLEM
NOISY RASTRIGIN PROBLEM
NOISY SCHAFFER'S F7 PROBLEM
NOISY BRANKE'S MULTIPEAK PROBLEM
NOISY KEANE'S BUMP PROBLEM
(5,10)
(10,20)
(20,40)
(30,60)
(40,80)
(50,100)
(60,120)
(70,140)
0 0.2 0.4 0.6 0.8 1 1.2
0 0.5 1 1.5 2 2.5 3
−0.5 0 0.5 1 1.5 2 2.5 3 3.5 4
0 2 4 6 8 10 12
0 0.5 1 1.5 2 2.5 3 3.5
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
0 10 20 30 40 50 60 70 80 90
0 2 4 6 8 10 12
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45
−0.7 −0.65 −0.6 −0.55 −0.5 −0.45 −0.4 −0.35 −0.3 −0.25

6.3.3 Comparison Adaptive versus Non-Adaptive

Lastly we run an empirical comparative study of the best instances of the five noise handling schemes incorporated in the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES, compared against the canonical instance of both schemes on the complete set of test problems (see Table 6.3). Each instance uses the optimal settings as found in the previous sections (see Table 6.7). For the MEM and the MPM schemes, this varies for each test problem. The question that we aim to answer with this empirical comparison is: how do the adaptive averaging techniques compare to optimally tuned MEM and MPM schemes?

	$(5/2_{DI}, 35)$-σ**SA-ES**	**CMA-ES**
Canonical	default	default
MEM	see Table 6.5	see Table 6.5
MPM	see Table 6.6	see Table 6.6
PUH	$\delta = 0.9$, $\alpha = 1.3$	$\delta = 0.5$, $\alpha = 1.7$
UH	$\theta = 0.9$, $\alpha = 1.1$	$\theta = 0.9$, $\alpha = 1.5$
IUH	$\theta = 0.1$, $\alpha = 1.7$	$\theta = 0.3$, $\alpha = 1.5$

Table 6.7: Algorithm settings comparison noise handling techniques.

Section 6.3.3.1 and Section 6.3.3.2 show the results for the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES respectively. Table 6.8 shows the combined rank scores of the compared schemes.

	$(5/2_{DI}, 35)$-σ**SA-ES**	**CMA-ES**
Canonical	354,792	384,091
MEM	296,002	291,950
MPM	289,257	**135,653**
PUH	335,450	344,457
UH	265,871	329,444
IUH	**261,628**	317,405

Table 6.8: Combined rank sums of the full comparison of noise handling techniques on the full set of test problems.

For the $(5/2_{DI}, 35)$-σSA-ES, the statistics tables and the boxplots show that overall, all schemes yield comparable results except for a few outliers. On the sphere problem, the adaptive averaging schemes clearly outperform the static schemes, which also holds for the Griewank problem and Branke's multipeak problem. On the Rastriging problem, the implicit averaging scheme is clearly better. A remarkable negative result is observed for the PUH scheme on the Ackley problem, Schaffer's f7 problem, and the Keane bump problem, where it is clearly worse than all other schemes. This can be attributed to a very slow convergence rate that is due to a fast increase of the sample size. The canonical instantiation of the $(5/2_{DI}, 35)$-σSA-ES is in most, but not all cases outperformed by the other schemes.

For the CMA-ES, the implicit averaging scheme shows to be a clear winner, ranking first on all test problems. Although implicit averaging already showed to be more suitable for (weighted) intermediate recombination in Section 5.3.3, the gain is remarkable. Hence, for the CMA-ES, using larger population sizes for noisy objective functions is clearly beneficial. No clear winner can be identified when comparing the adaptive averaging schemes against the explicit resampling scheme. However, Table 6.8 shows a slightly better combined rank sum for the MEM approach. Of the adaptive averaging techniques, the PUH scheme is the worst choice and the IUH seems to be marginally better than the UH scheme. Also here, the canonical instantiation is generally outperformed by the other schemes.

To summarize, a well-tuned static noise handling scheme can yield competitive or better results than adaptive averaging scheme. Increasing the population size is, especially for the CMA-ES, a promising strategy for noise handling. Among the adaptive averaging schemes, the UH and the IUH scheme yield better results than the PUH scheme, and the difference between the UH and the IUH scheme is marginal. Finally, the results confirm that the noise handling schemes generally yield better results than the canonical instantiations of the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES.

6.3.3.1 Results $(5/2_{DI}, 35)$-σSA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	2.39	9.45	0.48	45077	6
MEMopt	2.09	7.86	0.28	33477	4
MPMopt	1.03	4.08	0.39	41957	5
PUH	**1.97**	**8.02**	**0.17**	**18400**	**1**
UH	1.27	6.04	0.20	21950	3
IUH	1.23	6.01	0.18	19439	2

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	1.60	2.62	1.07	37491	6
MEMopt	2.39	4.94	0.83	31699	4
MPMopt	1.14	1.64	0.89	31829	5
PUH	2.19	5.27	0.86	31632	3
UH	1.51	4.00	0.72	25089	2
IUH	**1.62**	**4.26**	**0.64**	**22560**	**1**

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	**1.46**	**4.49**	**0.50**	**20917**	**1**
MEMopt	1.32	3.23	1.00	26712	3
MPMopt	0.48	0.63	0.00	24501	2
PUH	1.60	3.69	1.00	36378	5
UH	1.58	4.64	0.00	33421	4
IUH	1.85	5.28	0.00	38371	6

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	8.93	1.16	8.90	34190	6
MEMopt	8.72	1.13	8.71	30623	5
MPMopt	8.58	0.73	8.51	27934	2
PUH	9.48	6.97	8.65	29283	3
UH	8.64	1.01	8.66	30395	4
IUH	**9.16**	**6.68**	**8.52**	**27875**	**1**

NOISY ACKLEY PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	1.52	2.34	0.77	33316	5
MEMopt	0.95	1.50	0.58	25200	3
MPMopt	1.14	1.80	0.70	29322	4
PUH	2.17	1.68	1.79	48382	6
UH	1.59	2.71	0.49	22154	2
IUH	**1.06**	**1.93**	**0.53**	**21926**	**1**

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	1.32	0.35	1.23	44692	6
MEMopt	1.16	0.21	1.12	29809	4
MPMopt	1.22	0.12	1.21	42350	5
PUH	**1.10**	**0.24**	**1.08**	**19199**	**1**
UH	1.14	0.35	1.09	21755	2
IUH	1.18	0.40	1.08	22495	3

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	23.63	35.54	10.56	32775	2
MEMopt	26.91	39.33	10.52	33178	4
MPMopt	**9.54**	**21.66**	**4.26**	**11411**	**1**
PUH	25.59	35.93	11.53	35138	6
UH	24.80	36.01	10.59	33095	3
IUH	33.85	44.41	11.50	34703	5

NOISY SCHAFFERS F7 PROBLEM

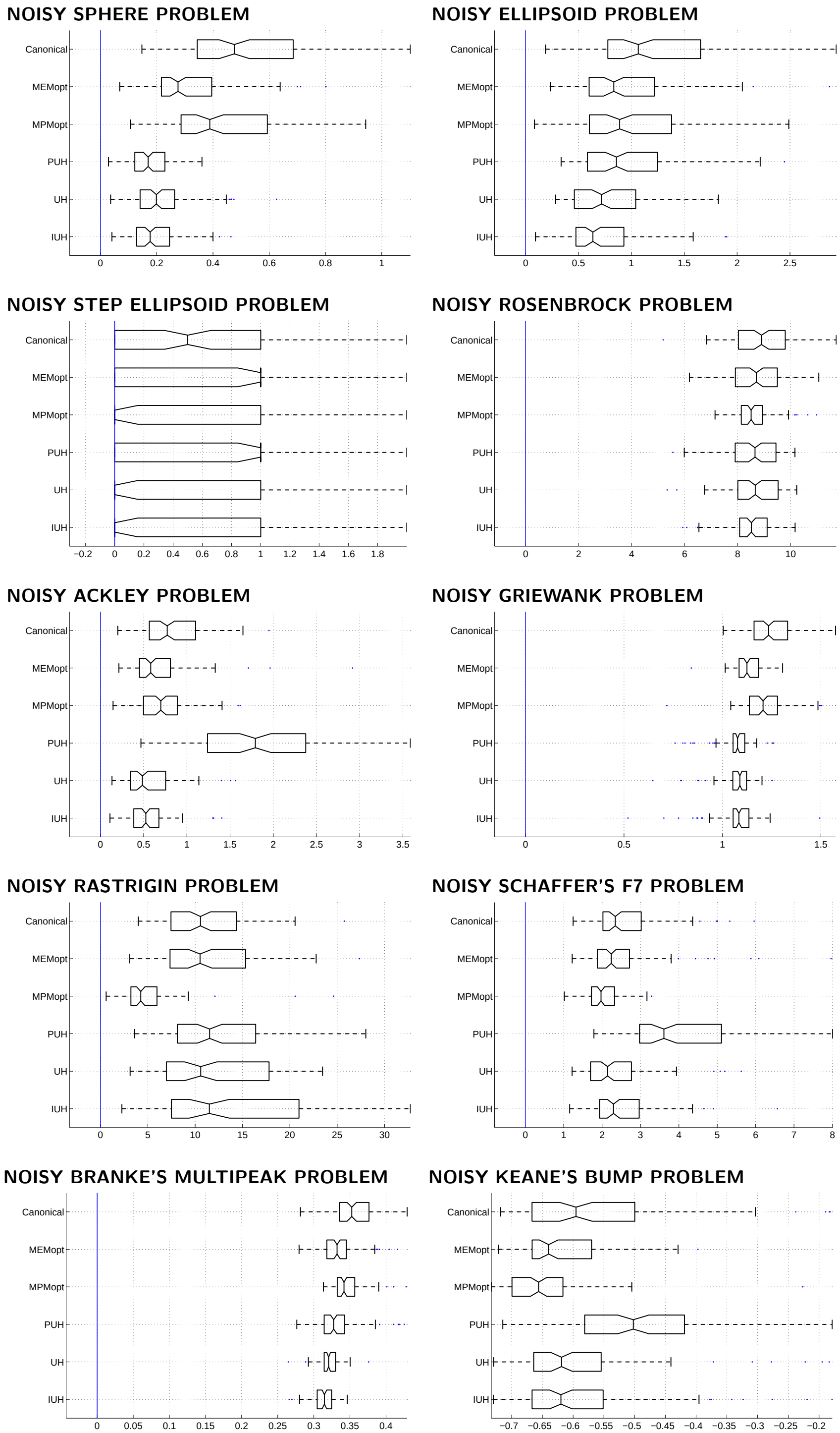
	Mean	Std	Med	$\sum$#	#
Canonical	7.88	17.14	2.34	30599	5
MEMopt	7.49	16.26	2.24	28243	3
MPMopt	**2.02**	**0.42**	**1.97**	**19458**	**1**
PUH	10.41	17.84	3.61	45587	6
UH	10.05	20.42	2.14	26159	2
IUH	10.15	20.34	2.30	30254	4

NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	0.39	0.12	0.35	42964	6
MEMopt	0.34	0.06	0.33	30442	4
MPMopt	0.35	0.02	0.34	40114	5
PUH	0.35	0.07	0.33	28137	3
UH	0.34	0.08	0.32	22750	2
IUH	**0.32**	**0.06**	**0.31**	**15893**	**1**

NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	$\sum$#	#
Canonical	-0.53	0.20	-0.60	32771	5
MEMopt	-0.61	0.09	-0.64	26619	2
MPMopt	**-0.63**	**0.12**	**-0.66**	**20381**	**1**
PUH	-0.49	0.13	-0.50	43314	6
UH	-0.57	0.15	-0.62	29103	4
IUH	-0.59	0.12	-0.62	28112	3


NOISY SPHERE PROBLEM
NOISY ELLIPSOID PROBLEM
NOISY STEP ELLIPSOID PROBLEM
NOISY ROSENBROCK PROBLEM
NOISY ACKLEY PROBLEM
NOISY GRIEWANK PROBLEM
NOISY RASTRIGIN PROBLEM
NOISY SCHAFFER'S F7 PROBLEM
NOISY BRANKE'S MULTIPEAK PROBLEM
NOISY KEANE'S BUMP PROBLEM
Canonical
MEMopt
MPMopt
PUH
UH
IUH

6.3.3.2 Results CMA-ES

NOISY SPHERE PROBLEM

	Mean	Std	Med	∑#	#
Canonical	0.58	0.24	0.54	52492	6
MEMopt	2.03	8.76	0.20	32436	5
MPMopt	**0.14**	**0.09**	**0.12**	**19068**	**1**
PUH	0.16	0.08	0.15	22805	2
UH	0.18	0.07	0.16	26880	4
IUH	0.79	6.19	0.17	26619	3

NOISY ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
Canonical	1.96	3.72	1.30	46425	6
MEMopt	1.86	4.45	0.83	33230	5
MPMopt	**0.35**	**0.25**	**0.28**	**10562**	**1**
PUH	1.37	4.36	0.69	28530	2
UH	1.24	2.78	0.77	31731	4
IUH	0.98	1.42	0.70	29822	3

NOISY STEP ELLIPSOID PROBLEM

	Mean	Std	Med	∑#	#
Canonical	1.17	1.03	1.00	27668	3
MEMopt	1.11	2.02	1.00	25590	2
MPMopt	**0.30**	**0.73**	**0.00**	**16698**	**1**
PUH	1.88	7.00	1.00	32718	4
UH	1.90	5.55	1.00	38420	5
IUH	1.20	2.37	1.00	39206	6

NOISY ROSENBROCK PROBLEM

	Mean	Std	Med	∑#	#
Canonical	10.08	6.11	9.83	38624	6
MEMopt	56.56	327.76	9.25	31418	3
MPMopt	**8.64**	**0.63**	**8.67**	**21359**	**1**
PUH	56.58	402.82	9.34	31974	5
UH	66.85	352.85	9.22	31785	4
IUH	8.73	0.99	8.75	25140	2

NOISY ACKLEY PROBLEM

	Mean	Std	Med	∑#	#
Canonical	2.14	1.21	2.11	43224	5
MEMopt	0.80	1.57	0.43	17300	2
MPMopt	**0.35**	**0.41**	**0.28**	**8159**	**1**
PUH	2.19	1.27	2.03	44090	6
UH	1.55	1.33	1.28	34218	4
IUH	1.45	1.35	1.15	33309	3

NOISY GRIEWANK PROBLEM

	Mean	Std	Med	∑#	#
Canonical	1.25	0.26	1.23	47773	6
MEMopt	1.11	0.18	1.10	31302	5
MPMopt	**1.05**	**0.07**	**1.05**	**19048**	**1**
PUH	1.12	0.28	1.08	27034	2
UH	1.08	0.06	1.08	27569	3
IUH	1.08	0.07	1.08	27574	4

NOISY RASTRIGIN PROBLEM

	Mean	Std	Med	∑#	#
Canonical	15.69	12.86	13.94	35374	5
MEMopt	15.10	11.95	13.36	34359	4
MPMopt	**4.61**	**7.33**	**3.25**	**7810**	**1**
PUH	16.25	14.13	14.20	36903	6
UH	14.62	8.53	13.37	33823	3
IUH	14.70	14.38	12.32	32031	2

NOISY SCHAFFERS F7 PROBLEM

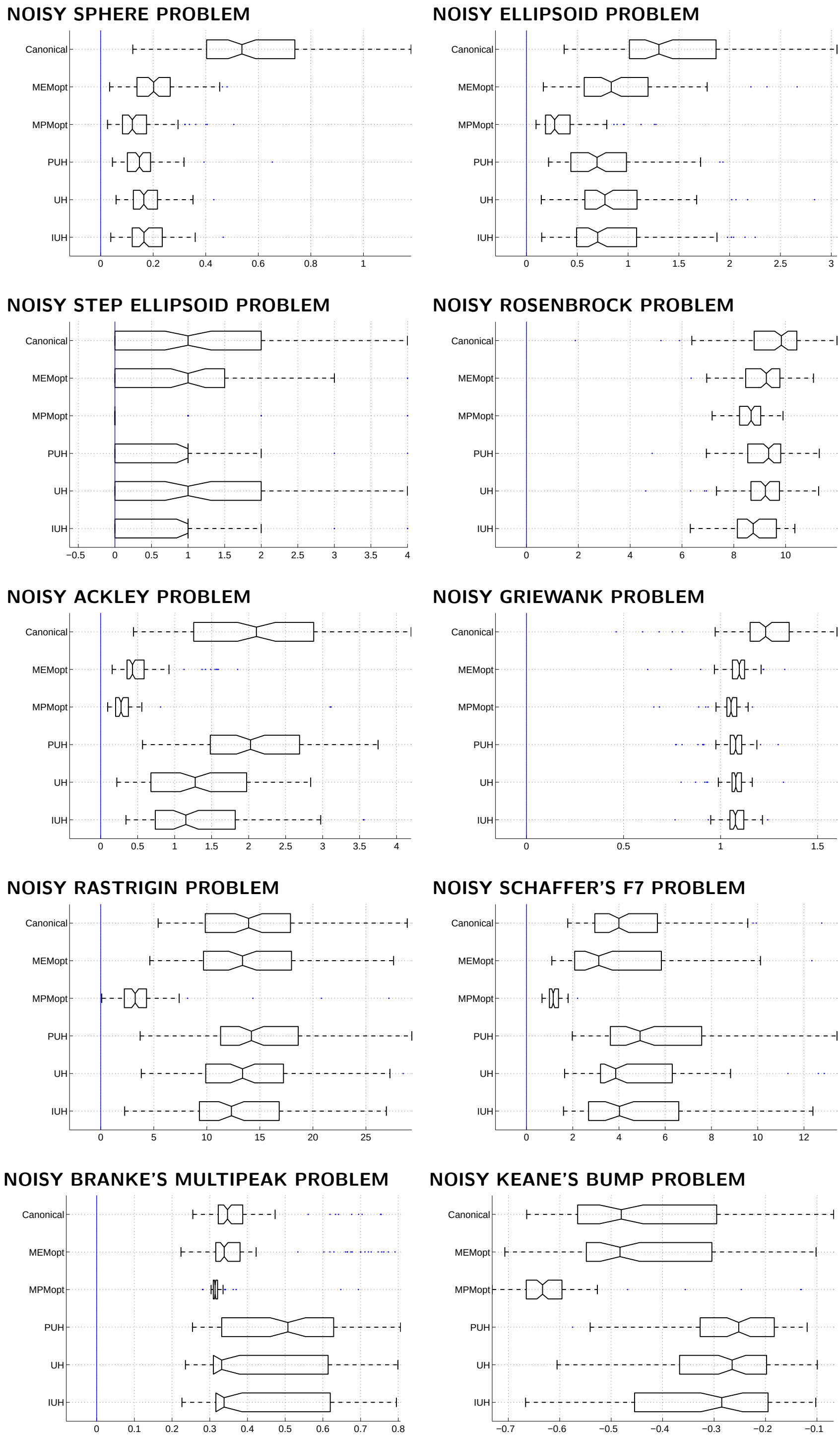
	Mean	Std	Med	∑#	#
Canonical	5.32	6.62	4.00	34220	3
MEMopt	4.78	6.55	3.12	29294	2
MPMopt	**2.96**	**10.21**	**1.15**	**6829**	**1**
PUH	6.25	6.34	4.91	39970	6
UH	5.04	2.91	3.86	35461	5
IUH	5.64	7.15	4.02	34526	4

NOISY BRANKE MULTIPEAK PROBLEM

	Mean	Std	Med	∑#	#
Canonical	0.39	0.12	0.35	32235	5
MEMopt	0.40	0.15	0.34	30623	3
MPMopt	**0.32**	**0.05**	**0.31**	**17366**	**1**
PUH	0.49	0.15	0.51	38577	6
UH	0.42	0.16	0.33	29348	2
IUH	0.44	0.16	0.34	32151	4

NOISY KEANE BUMP PROBLEM

	Mean	Std	Med	∑#	#
Canonical	-0.44	0.17	-0.48	26056	2
MEMopt	-0.43	0.16	-0.48	26398	3
MPMopt	**-0.62**	**0.10**	**-0.63**	**8754**	**1**
PUH	-0.27	0.10	-0.25	41856	6
UH	-0.28	0.11	-0.26	40209	5
IUH	-0.32	0.15	-0.29	37027	4


NOISY SPHERE PROBLEM
NOISY ELLIPSOID PROBLEM
NOISY STEP ELLIPSOID PROBLEM
NOISY ROSENBROCK PROBLEM
NOISY ACKLEY PROBLEM
NOISY GRIEWANK PROBLEM
NOISY RASTRIGIN PROBLEM
NOISY SCHAFFER'S F7 PROBLEM
NOISY BRANKE'S MULTIPEAK PROBLEM
NOISY KEANE'S BUMP PROBLEM
Canonical
MEMopt
MPMopt
PUH
UH
IUH

6.4 Summary and Discussion

In this chapter, the technique of adaptive averaging was studied in more detail, particularly focusing on poset based adaptive averaging (PUH), rank-change based adaptive averaging (UH), and rank-inversions based adaptive averaging or (IUH).

A theoretical study is presented on the growth rate of the sample size in case of an optimally adapted sample size for the noisy sphere problem. For the $(\mu/\mu_I, \lambda)$-ES on the noisy sphere, it is shown that the sample size must grow quartically with respect to the distance to the optimum to keep positive or optimal progress. Consequently, to achieve linear convergence over the number of generations, the sample size must grow exponentially. Hence, a multiplicative update rule should be used for the sample size within inter-generation adaptive resampling methods.

Furthermore, optimal settings have been derived for the adaptive averaging techniques for the $(5/2_{DI}, 35)$-σSA-ES and the CMA-ES, based on an empirical study on the noisy sphere problem (see Table 6.1). The adaptive averaging techniques seem to be quite robust for different settings of the uncertainty threshold δ/θ and the growth rate of the sample size α.

Finally, from the empirical study in the last part of this chapter we conclude that the adaptive averaging techniques yield results that are comparable to optimally tuned static noise handling techniques. That is, except for the implicit averaging scheme in case of the CMA-ES, which clearly outperforms the other schemes. Hence, in terms of sampling efficiency (see Section 5.6) the adaptive averaging techniques do not yield a gain by starting out with a few samples for each individual and gradually increase the sample budget. The exponential growth required to achieve positive progress possibly negates the gain in earlier generations. On the other hand, advantages of using adaptive averaging over non-adaptive techniques are that they do not require the a priori setting of a sample size or population size and that they allow for arbitrary convergence accuracy. The newly introduced algorithmic parameters (the uncertainty threshold δ/θ and the growth rate of the sample size α) are more robust than the sample size (i.e., weak parameter reduction is achieved). Of the adaptive averaging schemes, the rank-change based adaptive averaging scheme (UH) and the inversions-based adaptive averaging scheme (IUH) provide the most promising results.

In conclusion, for solving noisy optimization problems, well-tuned implicit or explicit averaging techniques are simple but effective ways to counter the effects of noise. The rank-change based and rank-inversions based adaptive averaging technique can yield results comparable to well-tuned static noise handling schemes and are therefore well-suited alternatives when it is not possible to tune the static noise handling approaches.

As a future direction, in the context of global intermediate recombination, instead of increasing the sample size, it might be interesting to consider adaptive averaging techniques based on increasing the population size.