A Surrogate Model Assisted (1+1)-ES with Increased Exploitation of the Model

Jingyun Yang and Dirk V. Arnold
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada B3H 4R2
Jingyun.Yang@dal.ca|dirk@cs.dal.ca

ABSTRACT

Surrogate models in black-box optimization can be exploited to different degrees. At one end of the spectrum, they can be used to provide inexpensive but inaccurate assessments of the quality of candidate solutions generated by the black-box optimization algorithm. At the other end, optimization of the surrogate model function can be used in the process of generating those candidate solutions themselves. The latter approach more fully exploits the model, but may be more prone to fall prey to systematic model error. This paper examines the effect of the degree of exploitation of the surrogate model in the context of a simple $(1 + 1)$-ES from two perspectives. First, we analytically derive the potential gain from more fully exploiting surrogate models by using a spherically symmetric test function and a simple model for the error resulting from the use of surrogate models. We then observe the effects of increased exploitation in an evolution strategy employing local Gaussian process surrogate models applied to a range of test problems. We find that the gain resulting from more fully exploiting surrogate models can be considerable primarily on less than perfectly conditioned problems.

CCS CONCEPTS

•Mathematics of computing → Bio-inspired optimization;
•Computing methodologies → Continuous space search;

KEYWORDS

Stochastic black-box optimization; evolution strategy; surrogate modelling; Gaussian process

ACM Reference format:
Jingyun Yang and Dirk V. Arnold. 2019. A Surrogate Model Assisted (1+1)-ES with Increased Exploitation of the Model. In *Proceedings of GECCO '19, Prague, Czech Republic, July 13-17, 2019,* 7 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Surrogate modelling techniques are commonly used when solving black-box optimization problems where the evaluation of the objective function is expensive. Models are built based on information gained through evaluation of the objective function in previous iterations. These models can then be used as inaccurate but inexpensive surrogates for the true objective function. Jin [6] and Loshchilov [11] provide surveys covering the use of surrogate modelling techniques in evolutionary computation.

Black-box optimization algorithms that employ surrogate models need to weigh the savings in cost from using those models against the potentially poor steps made as a result of their inaccuracy. Surrogate model assisted algorithms differ in the degree to which they exploit those models. One approach is to leave the generation of candidate solutions under the control of the black-box optimization algorithm. The surrogate model is used in place of the true objective function in order to avoid the costly evaluation of likely poor candidate solutions. An example of an algorithm that implements this approach is the Local Meta-Model Covariance Matrix Adaptation Evolution Strategy (lmm-CMA-ES) by Kern et al. [8]. A second approach is to involve the surrogate model in the generation of candidate solutions themselves. The surrogate model function is optimized in each iteration, and the optimizer of the surrogate model is subsequently evaluated using the true objective function. This approach is employed for example in the Gaussian Process Optimization Procedure (GPOP) by Büche et al. [4]. Arguably, the latter approach more fully exploits the surrogate models. At the same time, it may be more prone to make poor steps due to surrogate model error.

Implications of design decisions when incorporating surrogate models in black box optimization algorithms are not well understood. As a step toward understanding the consequences of different degrees of exploitation of surrogate models, we consider the performance of a surrogate model assisted $(1 + 1)$-ES[1]. The $(1 + 1)$-ES provides a baseline that has been well established since the early work of Rechenberg [14], and a large body of research concerning the performance of other evolution strategies relative to that of the $(1 + 1)$-ES exists [3]. Kayhani and Arnold [7] have studied the performance of a $(1 + 1)$-ES that employs surrogate models in order to decide whether an offspring candidate solution should be evaluated by the true objective function. They both analytically consider the performance of the algorithm on spherically symmetric test functions by using a simple model for the error resulting from the use of surrogate models, and they experimentally evaluate it on a wider range of test functions. We use their algorithm, but generalize the approach to the generation of candidate solutions to potentially more fully exploit surrogate models. Using their error model, we compare the potential speed-up resulting from more fully exploiting surrogate models relative to that of the strategy

GECCO '19, Prague, Czech Republic

978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

[1] See Hansen et al. [5] for evolution strategy terminology.

that minimally exploits the models. We then experimentally consider a wider range of test functions and find that while the benefits predicted under the simple error model cannot be fully realized, the gain from increased exploitation of the surrogate model can nonetheless be significant, especially for less than perfectly well conditioned problems.

The remainder of this paper is organized as follows. Section 2 reviews related work and introduces a surrogate model assisted (1+1)-ES with variable exploitation of the model. Section 3 studies the performance of that algorithm on spherically symmetric objective functions by assuming a simple model for the error resulting from the use of surrogate models. Section 4 employs Gaussian process surrogate models and applies the algorithm to a wider range of test functions. Section 5 concludes with a brief summary and proposed future work.

2 BACKGROUND

Many approaches that employ surrogate models in evolutionary algorithms have been proposed. Jin [6] and Loshchilov [11] provide comprehensive overviews. Much recent work has focused on surrogate model assisted variants of covariance matrix adaptation evolution strategies (CMA-ES); see for example Loshchilov et al. [12] and Pitra et al. [13]. Our goal here is to study the impact of the degree of exploitation of the surrogate model on the performance of evolution strategies, and is thus orthogonal to the use of covariance matrix adaptation.

The value of surrogate modelling approaches is usually quantified through the notion of speed-up, which is defined as the number of objective function evaluations required to solve a given optimization problem using some black box optimization algorithm that does not employ surrogate models, divided by the number of objective function evaluations required by the corresponding algorithm that does use surrogate models. That is, it is assumed that the cost of surrogate modelling is negligible compared to the cost of evaluating the objective function.

A straightforward use of surrogate models has been proposed by Kern et al. [8]. Their algorithm in each iteration generates $\lambda > 1$ candidate solutions, employs a surrogate model that is based on information gained in previous iterations to generate estimates of their relative fidelities, and then uses the true objective function to evaluate those candidate solutions that score best under the model. The ratio of candidate solutions evaluated using the objective function is adapted based on whether the true objective function values that are obtained are in conflict with the ranking of the candidate solutions according to the surrogate model. As in each iteration at least one candidate solution is evaluated using the true objective function, the maximum speed-up compared to the strategy that uses the objective function to evaluate all λ candidate solutions is λ, but will be lower if the accuracy of the surrogate models is poor. Kern et al. [8] report experimentally observed speed-ups by factors between two and eight on unimodal functions, including Schwefel's ellipsoid and Rosenbrock's function, where dimensions range from two to sixteen.

Notice that candidate solutions in the algorithm by Kern et al. [8] are generated by the evolution strategy, and are thus random and unbiased. An algorithm that more fully exploits surrogate models is the Gaussian Process Optimization Procedure by Büche et al. [4]. In that algorithm, candidate solutions are obtained by in each iteration attempting to determine an optimal solution to the surrogate model function. That solution is then evaluated using the true objective function. Büche et al. [4] report speed-ups by a factor between four and five compared to CMA-ES on quadratic sphere functions and on Schwefel's ellipsoid, and smaller speed-ups on Rosenbrock's function.

Input:
- candidate solution $\mathbf{x} \in \mathbb{R}^n$
- step size parameter $\sigma \in \mathbb{R}$, $\sigma > 0$

Generate step vector $\mathbf{z} \in \mathbb{R}^n$ and let $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$.
Evaluate $\mathbf{y}$ using the surrogate model, yielding $f_\epsilon(\mathbf{y})$.
if $f_\epsilon(\mathbf{y}) \geq f(\mathbf{x})$ **then**
    Let $\sigma \leftarrow \sigma\, \mathrm{e}^{-c_1/D}$.
else
    Evaluate $\mathbf{y}$ using the objective function, yielding $f(\mathbf{y})$.
    Update the surrogate model.
    if $f(\mathbf{y}) \geq f(\mathbf{x})$ **then**
        Let $\sigma \leftarrow \sigma\, \mathrm{e}^{-c_2/D}$.
    else
        Let $\mathbf{x} \leftarrow \mathbf{y}$ and $\sigma \leftarrow \sigma\, \mathrm{e}^{c_3/D}$.
    end if
end if

Figure 1: Single iteration of the surrogate model assisted (1+1)-ES.

We argue that it is meaningful to determine speed-up values relative to a well established baseline, and that in the context of evolution strategies, the (1 + 1)-ES forms a natural baseline. A surrogate model assisted variant of the (1+1)-ES has been proposed by Kramer [10]. That algorithm uses a surrogate model to evaluate the offspring candidate solution. If the value obtained suggests that the candidate solution is superior to the parental candidate solution from t iterations prior, then the objective function is used to establish its true quality. The 1/5th rule [14] is used for adaptation of the step size.

Kayhani and Arnold [7] consider that same algorithm for $t = 0$, but with a novel form of step size adaptation derived from the implementation of the 1/5th rule due to Kern et al. [9]. Assuming that the task at hand is minimization of function $f : \mathbb{R}^n \to \mathbb{R}$, a single iteration of that algorithm is detailed in Fig. 1. Parental candidate solution $\mathbf{x} \in \mathbb{R}^n$ and step size parameter $\sigma \in \mathbb{R}$ form the state of the algorithm. Step vector $\mathbf{z} \in \mathbb{R}^n$ is generated by sampling from a multivariate Gaussian distribution with zero mean and unit covariance. The resulting offspring candidate solution $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$ is then evaluated using the surrogate model function $f_\epsilon(\cdot)$. If the value obtained suggests that the offspring candidate solution is inferior to the parent, then the step size is reduced. Otherwise, the offspring candidate solution is evaluated using the objective function, and it replaces the parent if and only if it is superior. The step size is increased in that case, and it is reduced if the parent prevails. Parameter $D > 0$ controls the speed with which the step size

parameter can be adapted. Values for the coefficients $c_1, c_2, c_3 > 0$ are derived from an analysis of the performance of the algorithm on spherically symmetric test functions, using a simple model for surrogate model error to be detailed in Section 3. Employing Gaussian process surrogate models, speed-ups between 1.6 and 3.5 relative to the (1 + 1)-ES without surrogate model assistance can be observed on several unimodal test functions.

The algorithm considered in this paper is identical to that considered by Kayhani and Arnold [7], save for the generation of step vectors in Line 1 of Fig. 1. Rather than sampling $\mathbf{z}$ from a zero mean Gaussian distribution, we sample $\lambda \geq 1$ trial step vectors $\mathbf{z}_i \in \mathbb{R}^n$, $i = 1, 2, \ldots, \lambda$, from a Gaussian distribution with zero mean and unit covariance. We then compute fitness estimates $f_e(\mathbf{x} + \sigma\mathbf{z}_i)$ using the surrogate model and use $k;\lambda$ to refer to the index of the kth smallest of the λ values observed. Finally, we compute the average

$$\mathbf{z} = \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbf{z}_{k;\lambda} \tag{1}$$

of the $\mu \geq 1$ seemingly best trial steps as the step to be used in Line 1 of Fig. 1. As this approach to computing step vectors amounts to the same process as the generation of a step in the $(\mu/\mu, \lambda)$-ES, we refer to the resulting algorithm as the surrogate model assisted (1 + 1)-ES with $(\mu/\mu, \lambda)$ preselection. Notice that the algorithm makes no more than one true objective function evaluation per iteration, independent of λ. Also notice that the algorithm considered by Kayhani and Arnold [7] is included as the special case that $\mu = \lambda = 1$. Choosing larger values for λ more fully exploits the surrogate model in that the model is used to evaluate a larger number of points. Letting $\lambda \to \infty$, $f_\epsilon(\mathbf{x} + \sigma\mathbf{z}_{1;\lambda})$ will converge almost surely to the optimum of the surrogate model function. Practically, due to the inaccuracy of the surrogate models, we expect the gain from increasing λ to level off, and values of $\mu > 1$ to be useful.

3 ANALYSIS

This section studies the performance of the algorithm presented in Section 2 by considering spherically symmetric test function $f : \mathbb{R}^n \to \mathbb{R}$ with $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$, using the simple model for the surrogate model error proposed by Kayhani and Arnold [7]. It generalizes the analysis presented there by considering step vectors from distributions other than a Gaussian distribution. Specifically, we consider the case that the step in Line 1 of the algorithm in Fig. 1 is generated by $(\mu/\mu, \lambda)$ preselection.

As Kayhani and Arnold [7], in this section we assume that the surrogate model error (i.e., the difference $f(\mathbf{y}) - f_\epsilon(\mathbf{y})$ between the true objective function value of a candidate solution and its estimated value according to the surrogate model) can be modelled by an independent Gaussian random variable with mean zero and variance σ_ϵ^2, and that that variance is the same for all candidate solutions generated within an iteration. The standard deviation σ_ϵ is a measure for the quality of the surrogate model in that more accurate models correspond to smaller values of σ_ϵ. We expect the quality of this simple model to deteriorate as λ increases as especially the assumption of the independence of samples is unrealistic for deterministic surrogate models.

As proposed by Rechenberg [14], step vector $\mathbf{z}$ can be decomposed into a component $\mathbf{z}_1$ in the direction $-\nabla f(\mathbf{x})$ of the negative of the gradient of the objective and a component $\mathbf{z}_{2\ldots n}$ orthogonal to the gradient direction. Due to the assumptions made with regard to the surrogate model error, the distribution of both components can be obtained from earlier work on the performance of the $(\mu/\mu, \lambda)$-ES applied to noisy spherically symmetric functions [1, 2]. Component $\mathbf{z}_{2\ldots n}$ is of random direction in the $(n-1)$-dimensional subspace orthogonal to the gradient. Its squared length for large dimensions is governed by

$$\lim_{n\to\infty} \frac{\|\mathbf{z}_{2\ldots n}\|^2}{n} = \frac{1}{\mu}\,. \tag{2}$$

The component $\mathbf{z}_1$ of step vector $\mathbf{z}$ in direction of the negative gradient has a signed length with a probability density that cannot be given in closed form. However, expressions for the cumulants of the distribution can be obtained, allowing to approximate the density by using the first terms in a Gram-Charlier expansion (see [15])

$$p_1(z) = \frac{1}{\sqrt{2\pi\kappa_2}} \exp\left(-\frac{(z-\kappa_1)^2}{2\kappa_2}\right) \left[1 + \frac{\gamma_1}{3!}\mathrm{He}_3\left(\frac{z-\kappa_1}{\sqrt{\kappa_2}}\right) + \ldots\right], \tag{3}$$

where κ_k denotes the kth cumulant, $\mathrm{He}_k(\cdot)$ is the kth Hermite polynomial, and $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ is the skewness of the distribution. Expressions for the first three cumulants are derived in [1, 2] and reproduced in the Appendix.

We refer to $\delta(\mathbf{z}) = n(f(\mathbf{x}) - f(\mathbf{x} + \sigma\mathbf{z}))/(2R^2)$, where $R = \|\mathbf{x}\|$, as the normalized fitness advantage associated with step vector $\mathbf{z}$ generated through $(\mu/\mu, \lambda)$ preselection. Introducing normalized step size $\sigma^* = n\sigma/R$, it follows

$$\begin{aligned}\delta(\mathbf{z}) &= \frac{n}{2R^2}\left(\mathbf{x}^{\mathrm{T}}\mathbf{x} - (\mathbf{x}+\sigma\mathbf{z})^{\mathrm{T}}(\mathbf{x}+\sigma\mathbf{z})\right)\\ &= \frac{n}{2R^2}\left(-2\sigma\mathbf{x}^{\mathrm{T}}\mathbf{z} - \sigma^2\|\mathbf{z}\|^2\right)\\ &\overset{n\to\infty}{=} \sigma^* z_1 - \frac{\sigma^{*2}}{2\mu}\,,\end{aligned} \tag{4}$$

where $\overset{n\to\infty}{=}$ denotes convergence in distribution, $z_1 = -\mathbf{x}^{\mathrm{T}}\mathbf{z}/R$ is the signed length of the component of the step in the negative gradient direction, and the presence of μ in the denominator results from Eq. (2) and signifies the presence of genetic repair [3]. Moreover, introducing $\sigma_\epsilon^* = n\sigma_\epsilon/(2R^2)$, the estimated normalized fitness advantage associated with $\mathbf{z}$ (i.e., the normalized fitness advantage estimated by using the surrogate model to evaluate $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$) is $\delta_\epsilon(\mathbf{z}) = \delta(\mathbf{z}) + \sigma_\epsilon^* z_\epsilon$, where z_ϵ is a random variable with Gaussian distribution with zero mean and unit variance.

The algorithm in Fig. 1 evaluates $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$ using the objective function if and only if the estimated fitness advantage associated with $\mathbf{z}$ is positive. The probability of using the objective function to

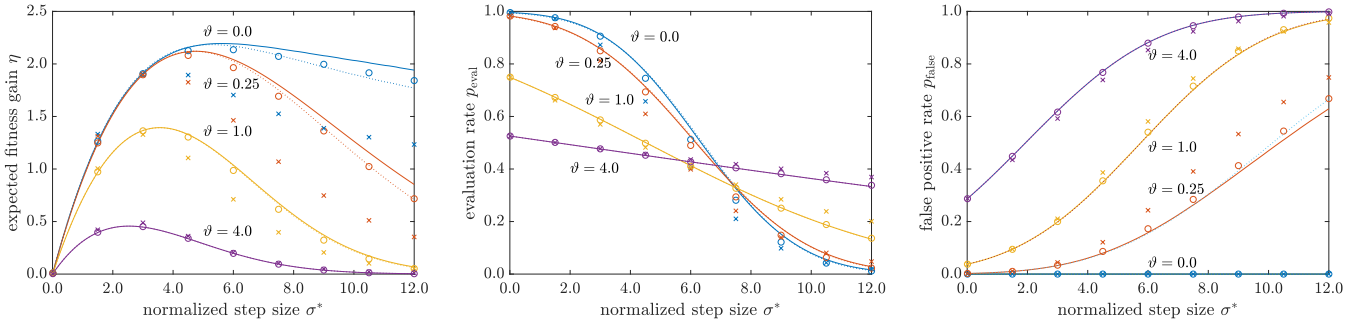

Figure 2: Expected single step behaviour of the surrogate model assisted $(1 + 1)$-ES with $(3/3, 10)$ preselection and unbiased Gaussian surrogate model error. The dashed and solid lines represent results obtained analytically in the limit $n \to \infty$ by considering Gram-Charlier approximations with cumulants up to the second and third included, respectively. The dots represent measurements made in runs of the algorithm with fixed normalized step size with $n = 100$ (○) and $n = 10$ (×).

evaluate a step generated using $(\mu/\mu, \lambda)$ preselection is thus

$$\begin{aligned} p_{\text{eval}} &= \text{Prob}\left[\delta_\epsilon(\mathbf{z}) > 0\right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-(\sigma^* z - \sigma^{*2}/(2\mu))/\sigma_\epsilon^*}^{\infty} \mathrm{e}^{-\frac{1}{2}y^2} p_1(z)\,\mathrm{d}y\,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} p_1(z)\,\Phi\left(\frac{\sigma^* z - \sigma^{*2}/(2\mu)}{\sigma_\epsilon^*}\right)\mathrm{d}z\,, \end{aligned} \tag{5}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the Gaussian distribution with zero mean and unit variance and $p_1(\cdot)$ is the probability density function given in Eq. (3).

Similarly, the probability of using the objective function to evaluate a step generated by $(\mu/\mu, \lambda)$ preselection that has an associated negative fitness advantage is

$$\begin{aligned} p_{\text{false}} &= \text{Prob}\left[\delta(\mathbf{z}) < 0 \,|\, \delta_\epsilon(\mathbf{z}) > 0\right] \\ &= \frac{\text{Prob}[\delta(\mathbf{z}) < 0 \wedge \delta_\epsilon(\mathbf{z}) > 0]}{\text{Prob}[\delta_\epsilon(\mathbf{z}) > 0]} \\ &= \frac{1}{\sqrt{2\pi} p_{\text{eval}}} \int_{-\infty}^{\sigma^*/(2\mu)} \int_{-(\sigma^* z - \sigma^{*2}/(2\mu))/\sigma_\epsilon^*}^{\infty} \mathrm{e}^{-\frac{1}{2}y^2} p_1(z)\,\mathrm{d}y\,\mathrm{d}z \\ &= \frac{1}{p_{\text{eval}}} \int_{-\infty}^{\sigma^*/(2\mu)} p_1(z)\,\Phi\left(\frac{\sigma^* z - \sigma^{*2}/(2\mu)}{\sigma_\epsilon^*}\right)\mathrm{d}z \end{aligned} \tag{6}$$

as $\delta(\mathbf{z}) < 0$ is equivalent to $z_1 < \sigma^*/(2\mu)$. We refer to this probability as the false positive rate.

Finally, the expected value of the normalized change in objective function value

$$\Delta = \begin{cases} \delta(\mathbf{z}) & \text{if } \delta_\epsilon(\mathbf{z}) > 0 \text{ and } \delta(\mathbf{z}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

from one iteration of the surrogate model assisted $(1 + 1)$-ES to the next can be computed as

$$\begin{aligned} \mathrm{E}[\Delta] &= \frac{1}{\sqrt{2\pi}} \int_{\sigma^*/(2\mu)}^{\infty} \int_{-(\sigma^* z - \sigma^{*2}/(2\mu))/\sigma_\epsilon^*}^{\infty} \left(\sigma^* z - \frac{\sigma^{*2}}{2\mu}\right) \\ &\qquad \mathrm{e}^{-\frac{1}{2}y^2} p_1(z)\,\mathrm{d}y\,\mathrm{d}z \\ &= \int_{\sigma^*/(2\mu)}^{\infty} \left(\sigma^* z - \frac{\sigma^{*2}}{2\mu}\right) p_1(z)\,\Phi\left(\frac{\sigma^* z - \sigma^{*2}/2\mu}{\sigma_\epsilon^*}\right)\mathrm{d}z\,. \end{aligned} \tag{7}$$

With Eq. (3), Eqs. (5), (6), and (7) can be used to numerically compute the evaluation rate, the false positive rate, and the expected normalized change in objective function value of the strategy.

All of the equations thus derived consider only a single iteration of the surrogate model assisted $(1 + 1)$-ES. However, if a step size adaptation mechanism and surrogate modelling approach are in place such that the distributions of the normalized step size σ^* and the normalized surrogate model error strength σ_ϵ^* are invariant across iterations, then the algorithm converges linearly in expectation with dimension-normalized rate of convergence

$$\begin{aligned} c &= -\frac{n}{2}\,\mathrm{E}\left[\log\left(\frac{f(\mathbf{x}_{t+1})}{f(\mathbf{x}_t)}\right)\right] \\ &= -\frac{n}{2}\,\mathrm{E}\left[\log\left(1 - \frac{2\Delta}{n}\right)\right], \end{aligned} \tag{8}$$

where subscripts denote iteration number. As the rate of convergence does not account for computational cost and costs are incurred only in those iterations where a call to the objective function is made, we use $\eta = c/p_{\text{eval}}$ (normalized rate of convergence per objective function call) as performance measure and refer to it as the expected fitness gain. For $n \to \infty$ the logarithm in Eq. (8) can be linearized and the expected fitness gain is simply $\eta = \mathrm{E}[\Delta]/p_{\text{eval}}$.

We define noise-to-signal ratio $\vartheta = \sigma_\epsilon^*/\sigma^*$ as a measure for the accuracy of the surrogate model relative to the step size of the algorithm. We run experiments with the surrogate model assisted $(1 + 1)$-ES as given in Fig. 1, but with Gaussian distributed surrogate model error in place of a true surrogate model and with the normalized step size set to a fixed value rather than being adapted. Figure **??** plots the expected fitness gain η, the evaluation rate p_{eval}, and the false positive rate p_{false} against the normalized step size. The dots represent data measured in runs of the algorithm with $3/3, 10)$ preselection in dimensions $n \in \{10, 100\}$ that have been averaged over 10^6 iterations. The dashed and solid lines represent analytically obtained results using cumulants up to the second (dashed lines) or third (solid lines). It can be seen that consideration of the skewness of the distribution has a minor impact on the results. Deviaions between experimental measurements and analytical predictions decrease with increasing dimension.

It can be seen from Fig. 2 that for given noise-to-signal ratio, te evaluation rate of the algorithm decreases with increasing step

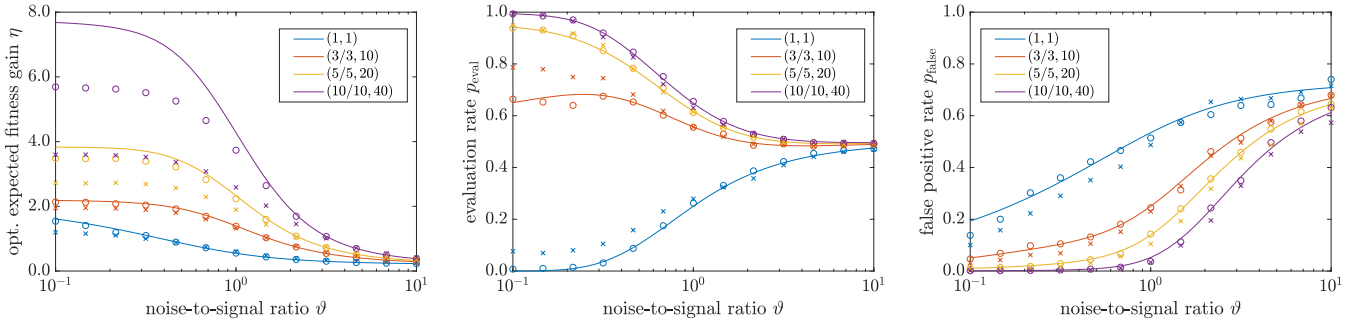

Figure 3: Optimal normalized step size and corresponding values of expected fitness gain, evaluation rate, and false positive rate plotted against the noise-to-signal ratio.

size while the false positive rate increases. Unsurprisingly, false positive rates increase with increasing surrogate model error.

4 EXPERIMENTS

- The number of objective function evaluations that the (1 + 1)-ES wihout surrogate model assistance required to solve the test problems ranges from the hundreads into the tens of thousands.

5 CONCLUSIONS

- The use of covariance matrix adaptation is orthogonal to what is studied here.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] D. V. Arnold. 2002. *Noisy Optimization with Evolution Strategies.* Kluwer Academic Publishers.
[2] D. V. Arnold and H.-G. Beyer. 2005. Expected sample moments of concomitants of selected order statistics. *Statistics and Computing* 15, 3 (2005), 241–250.
[3] H.-G. Beyer. 2001. *The Theory of Evolution Strategies.* Springer Verlag.
[4] D. Büche, N. N. Schraudolph, and P. Koumoutsakos. 2005. Accelerating evolutionary algorithms with Gaussian process fitness function models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 35, 2 (2005), 183–194.
[5] N. Hansen, D. V. Arnold, and A. Auger. 2015. Evolution strategies. In *Springer Handbook of Computational Intelligence*, J. Kacprzyk and W. Pedrycz (Eds.). Springer Verlag, 871–898.
[6] Y. Jin. 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm and Evolutionary Computation* 1, 2 (2011), 61–70.
[7] A. Kayhani and D. V. Arnold. 2018. Design of a surrogate model assisted (1 + 1)-ES. In *Parallel Problem Solving from Nature — PPSN XV*, A. Auger and others (Eds.). Springer Verlag, 16–28.
[8] S. Kern, N. Hansen, and P. Koumoutsakos. 2006. Local meta-models for optimization using evolution strategies. In *Parallel Problem Solving from Nature — PPSN IX*, T. P. Runarsson and others (Eds.). Springer Verlag, 939–948.
[9] S. Kern, S. D. Müller, N. Hansen, D. Büche, J. Ocenasek, and P. Koumoutsakos. 2004. Learning probability distributions in continuous evolutionary algorithms — A comparative review. *Natural Computing* 3, 1 (2004), 77–112.
[10] O. Kramer. 2016. *Machine Learning for Evolution Strategies.* Springer Verlag.
[11] I. Loshchilov. 2013. *Surrogate-Assisted Evolutionary Algorithms.* Ph.D. Dissertation. Université Paris Sud – Paris XI.
[12] I. Loshchilov, M. Schoenauer, and M. Sebag. 2013. Intensive surrogate model exploitation in self-adaptive surrogate-assisted CMA-ES. In *Genetic and Evolutionary Computation Conference — GECCO 2013.* ACM Press, 439–446.
[13] Z. Pitra, L. Bajer, J. Repický, and M. Holena. 2017. Overview of surrogate-model versions of covariance matrix adaptation evolution strategy. In *Genetic and Evolutionary Computation Conference Companion.* ACM Press, 1622–1629.
[14] I. Rechenberg. 1973. *Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Friedrich Frommann Verlag.
[15] A. Stuart and J. K. Ord. 1994. *Kendall's Advanced Theory of Statistics* (sixth ed.). Vol. I: Distribution Theory. Wiley.

APPENDIX

Cumulants of the distribution of the signed length of the component of the step vector in the direction of the negative gradient of the objective function resulting from $(\mu/\mu, \lambda)$ preselection can be obtained by computing expectations of concomitants of order statistics of the Gaussian distribution. Arnold and Beyer [1, 2] derive the following expressions for the first three cumulants:

$$\kappa_1 = a h^{1,0}_{\mu,\lambda}$$

$$\kappa_2 = \frac{1}{\mu}\left(1 + a^2 h^{1,1}_{\mu,\lambda}\right) + \frac{\mu-1}{\mu} a^2 h^{2,0}_{\mu,\lambda} - \kappa_1^2$$

$$\kappa_3 = \frac{1}{\mu^2}\left(3h^{1,0}_{\mu,\lambda} + a^2 h^{1,2}_{\mu,\lambda}\right) + 3\frac{\mu-1}{\mu^2} a \left(h^{1,0}_{\mu,\lambda} + a^2 h^{2,1}_{\mu,\lambda}\right) + \frac{(\mu-1)(\mu-2)}{\mu^2} a^3 h^{3,0}_{\mu,\lambda} - \frac{3}{\mu}\kappa_1 \left(1 + a^2 h^{1,1}_{\mu,\lambda}\right) - 3\frac{\mu-1}{\mu}\kappa_1 a^2 h^{2,0}_{\mu,\lambda} + 2\kappa_1^3$$

where $a = 1/\sqrt{1+\vartheta^2}$ with noise-to-signal ratio ϑ as defined in Section 3, and

$$h^{i,k}_{\mu,\lambda} = \frac{\lambda-\mu}{\sqrt{2\pi}}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} [\phi(x)]^i [\Phi(x)]^{\lambda-\mu-1} [1-\Phi(x)]^{\mu-i} \mathrm{d}x$$

with $\phi(\cdot)$ and $\Phi(\cdot)$ being the probability density and cumulative distribution functions of the univariate Gaussian distribution with mean zero and unit variance, respectively.

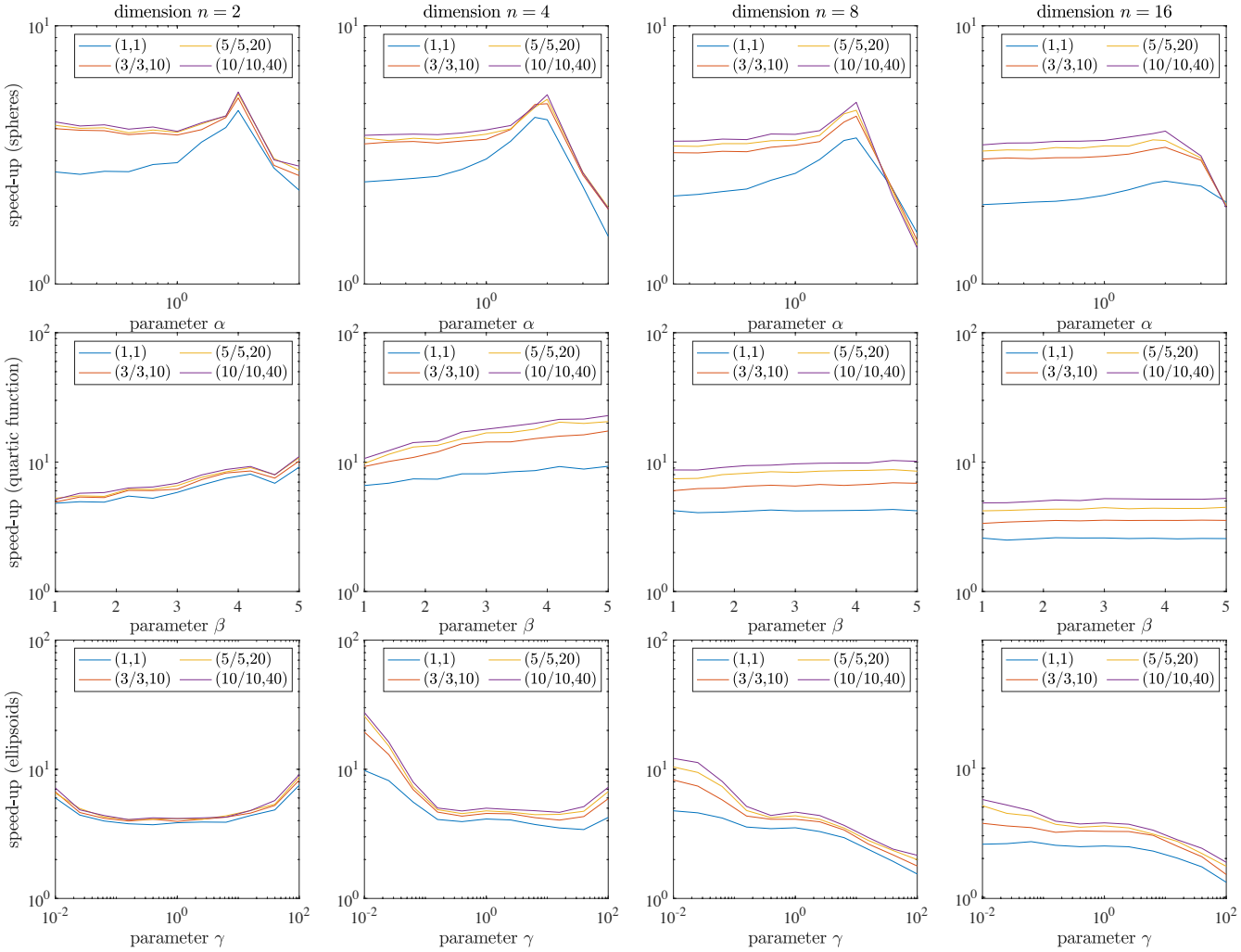

Figure 4: Speed-up

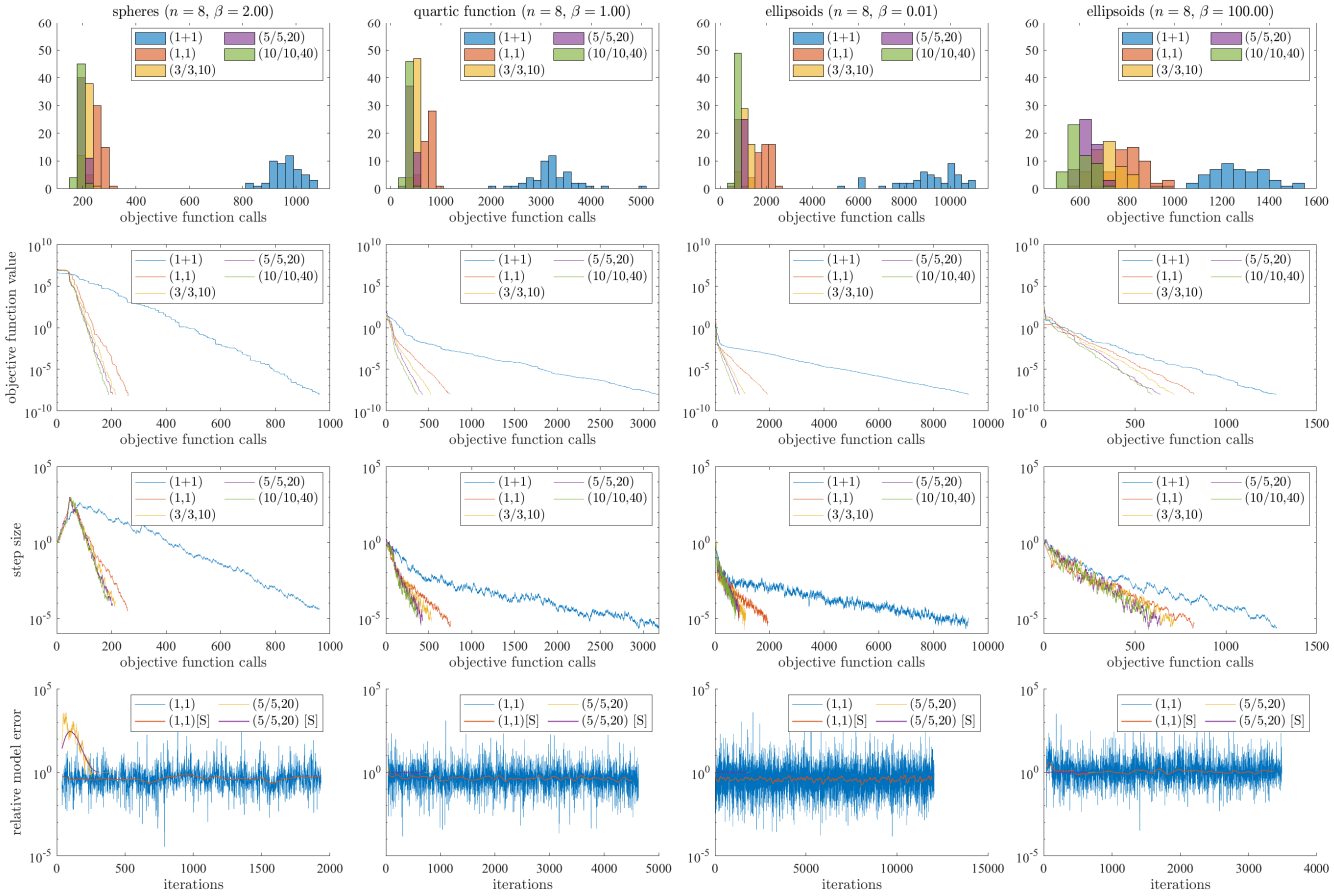

Figure 5: Step behaviour