Design of a Surrogate Model Assisted (1 + 1)-ES

Arash Kayhani and Dirk V. Arnold(✉)

Faculty of Computer Science, Dalhousie University,
Halifax, Nova Scotia B3H 4R2, Canada
Arash.Kayhani@dal.ca, dirk@cs.dal.ca

Abstract. Surrogate models are employed in evolutionary algorithms to replace expensive objective function evaluations with cheaper though usually inaccurate estimates based on information gained in past iterations. Implications of the trade-off between computational savings on the one hand and potentially poor steps due to the inaccurate assessment of candidate solutions on the other are generally not well understood. We study the trade-off in the context of a surrogate model assisted (1 + 1)-ES by considering a simple model for single steps. Based on the insights gained, we propose a step size adaptation mechanism for the strategy and experimentally evaluate it using several test functions.

1 Introduction

Surrogate models have been proposed as an approach for evolutionary algorithms (EAs) to deal with optimization problems where each evaluation of the objective function requires a considerable amount of time or incurs a significant cost. Surrogate models are built using information on candidate solutions that have been evaluated previously using the true objective function. Evaluating a new candidate solution using a surrogate model yields a potentially inaccurate estimate of its true objective function value at a much lower cost than would be incurred in the exact evaluation. Surrogate modelling is useful if the benefit of reduced cost outweighs the potentially poorer steps made due to the inexact evaluation of candidate solutions.

Numerous approaches for incorporating surrogate models in EAs exist and have been comprehensively surveyed by Jin [8] and Loshchilov [11]. Algorithms usually are heuristic in nature, and potential consequences of design decisions are not always well understood. Most recent work on surrogate model assisted EAs considers relatively sophisticated algorithms. Strategies usually are evaluated by comparing the approach that uses surrogate modelling techniques with a corresponding algorithm that does not. A potential pitfall in such comparisons arises in connection with the use of large populations: if an algorithm for a given optimization problem uses a larger than optimal population size, then efficiency can be gained simply by using a trivial surrogate modelling approach that classifies a fraction of candidate solutions as poor, at no computational cost.

A. Auger et al. (Eds.): PPSN 2018, LNCS 11101, pp. 16–28, 2018.
https://doi.org/10.1007/978-3-319-99253-2_2

Clearly, the computational savings in this case are due to the effective reduction of the population size rather than to surrogate modelling.

We contend that it is desirable to develop an improved understanding of the potential implications of the use of surrogate modelling techniques, and that such an understanding can be gained by analyzing the behaviour of surrogate model assisted EAs using simple test functions that allow comparing the performance of the algorithms against a well established baseline. The contributions of this paper are as follows: after briefly reviewing related work in Sect. 2, in Sect. 3 we propose a simple model for surrogate model assisted EAs and use it to study the single-step behaviour of a surrogate model assisted (1 + 1)-ES[1] on quadratic sphere functions. We then use the insights gained to propose a step size adaptation mechanism for that algorithm in Sect. 4, and we evaluate its performance using several test functions. Section 5 concludes with a brief discussion and future work.

2 Related Work

The use of surrogate models in EAs can be traced back to the 1980s. Both Jin [8] and Loshchilov [11] present comprehensive surveys of the development of the field. Notable strategies include, though are not limited to, the Gaussian Process Optimization Procedure (GPOP) by Büche et al. [4] and the Local Meta-Model Covariance Matrix Adaptation Evolution Strategy (lmm-CMA-ES) by Kern et al. [9]. GPOP iterates the optimization of a Gaussian process based model of the objective using CMA-ES [7] and the subsequent evaluation and addition of the solution obtained to the training set. With computational cost determined by the number of (exact) objective function evaluations required to reach the optimal solution to within some target accuracy, Büche et al. [4] report a speed-up by a factor between four and five compared to CMA-ES on quadratic sphere functions and on Schwefel's function, and smaller speed-ups on Rosenbrock's function. lmm-CMA-ES use locally weighted regression models in connection with an approximate ranking procedure within the CMA-ES. With full quadratic models, Kern et al. [9] report a speed-up by a factor between two and eight compared to CMA-ES on unimodal functions, including the quadratic sphere, Schwefel's function, and Rosenbrock's function. More recent surrogate model assisted CMA-ES variants include the Surrogate-Assisted Covariance Matrix Adaptation Evolution Strategy (^{s*}ACM-ES) by Loshchilov et al. [13] as well as several further algorithms surveyed and compared by Pitra et al. [14].

It is interesting to note that when considering unimodal test functions and comparing with relatively sophisticated black box optimization algorithms such as CMA-ES, the speed-ups reported as a result of using surrogate models appear to be a small factor (usually less than eight, frequently no larger than four), irrespective of the dimension of the problem. While larger speed-ups can be achieved when using surrogate models that perfectly fit the functions being optimized (e.g., quadratic models for optimizing quadratic functions), this observation is

[1] See Hansen et al. [6] for evolution strategy terminology.

not altogether unexpected in light of the performance bounds for black box optimization algorithms derived by Teytaud and Gelly [17].

A further interesting observation is that surrogate model assisted EAs tend to be relatively complicated and combine multiple heuristics for good performance. Notably, no surrogate model assisted version of the (1+1)-ES can be found in the literature. A seeming exception proposed by Chen and Zou [5] is not invariant to translations of the coordinate system — a property considered crucial for solving general unconstrained optimization problems — and does not include a mechanism for the adaptation of its step size. The Model Assisted Steady-State Evolution Strategy (MASS-ES) by Ulmer et al. [18] is a $(\mu + \lambda)$-ES that can in principle be run with $\mu = \lambda = 1$, but was not designed with those settings in mind and it is unclear whether its step size adaptation approach is effective under those conditions. Given the relative efficiency of the $(1+1)$-ES for unimodal black box problems and the relatively large body of knowledge regarding its convergence properties on convex functions, we argue that it is natural to ask to what degree the algorithm can be accelerated through the use of surrogate models, and how its step size can be adapted successfully.

3 Analysis

In order to gain a better understanding of potential implications of the use of surrogate models in EAs, in this section, we employ a simple model for the use of surrogate models. Specifically, we propose that an EA have the options of either evaluating a candidate solution accurately, at the cost of one objective function call, or of obtaining an inaccurate estimate of the solution's objective function value at vanishing cost. For simplicity, we assume that the inaccurate objective function value is a Gaussian random variable with a mean that coincides with the candidate solution's exact objective function value and some variance that models the accuracy of the surrogate model. As a result, techniques previously employed for the analysis of the behaviour of evolution strategies in the presence of Gaussian noise become applicable (see [1] and references therein). It would be straightforward to extend the analysis to biased surrogate models (i.e., models where the distribution mean differs from the exact objective function value). Models with a skew distribution of estimation errors could likely be considered based on analyses of the effects of non-Gaussian noise on the performance of evolution strategies (see [2]). Also not directly addressed in the present work are comparison based surrogate models. Loshchilov et al. [12] persuasively argue for such models in order to preserve invariance properties of comparison based optimization algorithms. We expect that an analysis analogous to what follows can be performed for such models.

We consider minimization of the quadratic sphere function $f : \mathbb{R}^n \to \mathbb{R}$ with $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$ using a surrogate model assisted $(1 + 1)$-ES, where throughout this section the simple model described above substitutes for a "true" surrogate model. We initially consider a single iteration of the strategy and defer the discussion of step size adaptation until Sect. 4. The algorithm in each iteration generates single offspring candidate solution $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$, where $\mathbf{x} \in \mathbb{R}^n$ is the best candidate solution obtained so far and is referred to as the parent, $\mathbf{z} \in \mathbb{R}^n$ is a standard normally distributed random vector, and $\sigma > 0$ is a step size parameter the adaptation of which is to be discussed below. The strategy uses the surrogate model to obtain an estimate $f_\epsilon(\mathbf{y})$ of the objective function value that according to the above assumptions is a random variable with mean $f(\mathbf{y})$ and some standard deviation $\sigma_\epsilon > 0$. Better surrogate models result in smaller values of σ_ϵ. If $f_\epsilon(\mathbf{y}) > f(\mathbf{x})$ (i.e., if the surrogate model suggests that the offspring candidate solution is inferior to the parent), then $\mathbf{y}$ is discarded and the strategy proceeds to the next iteration; otherwise it computes $f(\mathbf{y})$ at the cost of one objective function call and replaces $\mathbf{x}$ with $\mathbf{y}$ if and only if $f(\mathbf{y}) < f(\mathbf{x})$ (i.e., if the offspring candidate solution truly is superior to the parent). In the terminology of Loshchilov [11] this procedure can be considered a natural implementation of preselection in the $(1+1)$-ES.

The expected step of the strategy can be studied by using a decomposition of $\mathbf{z}$ first proposed by Rechenberg [15]. Vector $\mathbf{z}$ is written as the sum of two components: one in direction of the negative gradient direction $-\nabla f(\mathbf{x})$ and the other orthogonal to that. Due to symmetry, the length of the former component is standard normally distributed; the squared length of the latter is governed by a χ^2-distribution with $n-1$ degrees of freedom. The mean of that distribution is $n-1$ and its coefficient of variation tends to zero as n increases. Referring to $\delta = n(f(\mathbf{x}) - f(\mathbf{y}))/(2R^2)$, where $R = \|\mathbf{x}\|$, as the normalized fitness advantage of $\mathbf{y}$ over its parent, and introducing normalized step size $\sigma^* = n\sigma/R$, it follows

$$\begin{aligned}\delta &= \frac{n}{2R^2}\left(\mathbf{x}^{\mathrm{T}}\mathbf{x} - (\mathbf{x}+\sigma\mathbf{z})^{\mathrm{T}}(\mathbf{x}+\sigma\mathbf{z})\right) = \frac{n}{2R^2}\left(-2\sigma\mathbf{x}^T\mathbf{z} - \sigma^2\|\mathbf{z}\|^2\right) \\ &\overset{n\to\infty}{=} \sigma^* z_1 - \frac{\sigma^{*2}}{2}, \end{aligned} \tag{1}$$

where $z_1 = -\mathbf{x}^{\mathrm{T}}\mathbf{z}/R$ is a standard normally distributed random variable representing the length of the component of $\mathbf{z}$ in the direction of $-\nabla f(\mathbf{x})$ and $\overset{n\to\infty}{=}$ denotes convergence in distribution. Moreover, introducing $\sigma^*_\epsilon = n\sigma_\epsilon/(2R^2)$, the estimated normalized fitness advantage (i.e., the normalized fitness advantage estimated by using the surrogate model to evaluate $\mathbf{y}$) is $\delta_\epsilon = \delta + \sigma^*_\epsilon z_\epsilon$, where z_ϵ is standard normally distributed.

From the above, the estimated normalized fitness advantage is normally distributed with mean $-\sigma^{*2}/2$ and variance $\sigma^{*2} + \sigma^{*2}_\epsilon$ and thus has probability density

$$p_{\delta_\epsilon}(y) = \frac{1}{\sqrt{2\pi(\sigma^{*2}+\sigma^{*2}_\epsilon)}} \exp\left(-\frac{1}{2}\frac{\left(y+\sigma^{*2}/2\right)^2}{\sigma^{*2}+\sigma^{*2}_\epsilon}\right). \tag{2}$$

Moreover, the probability density of z_1 conditional on the estimated normalized fitness advantage δ_ϵ can be obtained as[2]

$$p_{z_1|\delta_\epsilon}(z \mid y) = \frac{\sqrt{\sigma^{*2}+\sigma^{*2}_\epsilon}}{\sqrt{2\pi}\sigma^*_\epsilon} \exp\left(-\frac{1}{2}\frac{\left((\sigma^{*2}+\sigma^{*2}_\epsilon)z - \sigma^*(y+\sigma^{*2}/2)\right)^2}{(\sigma^{*2}+\sigma^{*2}_\epsilon)\sigma^{*2}_\epsilon}\right). \tag{3}$$

[2] Detailed derivations of Eqs. (3), (4), (5), and (6) can be found in a separate document at `web.cs.dal.ca/~dirk/PPSN2018addendum.pdf`.

As $\mathbf{y}$ is evaluated using the objective function if and only if it appears superior to the parent based on the surrogate model, we write $p_{\text{eval}} = \text{Prob}[\delta_\epsilon > 0]$ for the probability of making a call to the objective function. From Eq. (2),

$$p_{\text{eval}} = \text{Prob}\left[\delta_\epsilon > 0\right] = \int_0^\infty p_{\delta_\epsilon}(y)\,\mathrm{d}y = \Phi\left(\frac{-\sigma^{*2}/2}{\sqrt{\sigma^{*2}+\sigma_\epsilon^{*2}}}\right), \tag{4}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Due to the accounting for computational costs, p_{eval} represents the expected cost per iteration of the algorithm. Similarly, as $\mathbf{y}$ replaces $\mathbf{x}$ if and only if $\delta_\epsilon > 0$ and $\delta > 0$, we write $p_{\text{step}} = \text{Prob}[\delta_\epsilon > 0 \wedge \delta > 0]$ for the probability of the offspring replacing the parent. From Eqs. (2) and (3),

$$p_{\text{step}} = \text{Prob}\left[\delta_\epsilon > 0 \wedge \delta > 0\right] = \int_0^\infty p_{\delta_\epsilon}(y) \int_{\sigma^*/2}^\infty p_{z_1|\delta_\epsilon}(z \,|\, y)\,\mathrm{d}z\,\mathrm{d}y = \frac{1}{\sqrt{2\pi}} \int_{\sigma^*/2}^\infty \mathrm{e}^{-z^2/2}\,\Phi\left(\frac{\sigma^* z - \sigma^{*2}/2}{\sigma_\epsilon^*}\right)\mathrm{d}z \tag{5}$$

as $\delta > 0$ is equivalent to $z_1 > \sigma^*/2$. Finally, the expected value of the normalized change in objective function value

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

from one iteration to the next can be computed as

$$\mathrm{E}\left[\Delta\right] = \int_0^\infty p_{\delta_\epsilon}(y) \int_{\sigma^*/2}^\infty \left(\sigma^* z - \frac{\sigma^{*2}}{2}\right) p_{z_1|\delta_\epsilon}(z \,|\, y)\,\mathrm{d}z\,\mathrm{d}y = \frac{1}{\sqrt{2\pi}} \int_{\sigma^*/2}^\infty \left(\sigma^* z - \frac{\sigma^{*2}}{2}\right) \mathrm{e}^{-z^2/2}\,\Phi\left(\frac{\sigma^* z - \sigma^{*2}/2}{\sigma_\epsilon^*}\right)\mathrm{d}z\,. \tag{6}$$

Equations (4), (5), and (6) describe the behaviour of the algorithm for $n \to \infty$ and can serve as approximations for finite but not too small n.

If a step size adaptation mechanism and surrogate modelling approach are in place such that the distributions of σ^* and σ_ϵ^* are independent of the iteration number, then the algorithm converges in expectation linearly with dimension-normalized rate of convergence

$$c = -\frac{n}{2}\,\mathrm{E}\left[\log\left(\frac{f(\mathbf{x}_{t+1})}{f(\mathbf{x}_t)}\right)\right] = -\frac{n}{2}\,\mathrm{E}\left[\log\left(1 - \frac{2\Delta}{n}\right)\right], \tag{7}$$

where subscripts denote iteration number. However, the rate of convergence does not account for computational cost as costs are incurred only in those iterations

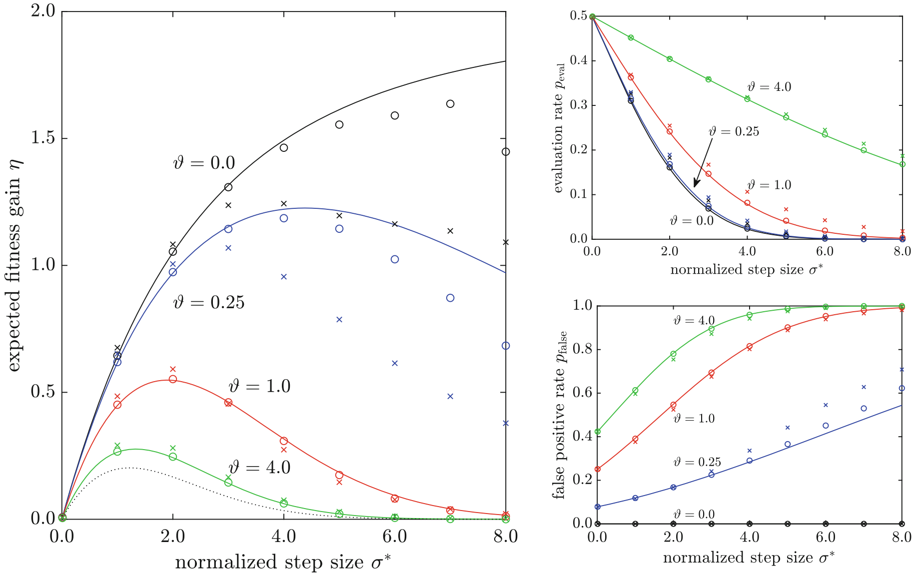

Fig. 1. Expected single step behaviour of the surrogate model assisted (1 + 1)-ES with unbiased Gaussian surrogate error. The solid lines represent results obtained analytically in the limit $n \to \infty$. The dots show values observed experimentally for $n = 10$ (crosses) and $n = 100$ (circles). The dotted line in the left hand plot illustrates the corresponding relationship for the (1 + 1)-ES without surrogate model assistance.

where a call to the objective function is made. We thus 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. (7) 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 quality of the surrogate model relative to the step size of the algorithm and in Fig. 1 plot the evaluation rate p_{eval}, the false positive rate $p_{\text{false}} = 1 - p_{\text{step}}/p_{\text{eval}}$ (i.e., the probability of a candidate solution that is deemed superior by the surrogate model to be inferior to the parent according to the true objective function), and the expected fitness gain against the normalized step size. The lines show results obtained from Eqs. (4), (5), and (6). The dots show corresponding values observed in experiments with unbiased Gaussian surrogate error for $n \in \{10, 100\}$ that have been obtained by averaging over 10^7 iterations. Deviations of the experimental measurements from values obtained in the limit $n \to \infty$ are considerable primarily for large normalized step size and small noise-to-signal ratio.

It can be seen from Fig. 1 that for given noise-to-signal ratio, the evaluation rate of the algorithm decreases with increasing step size. For very small steps, one out of every two steps is deemed successful by the surrogate model; with larger steps, the algorithm becomes more "selective" when deciding whether to obtain an exact objective function value for a candidate solution. At the same time,

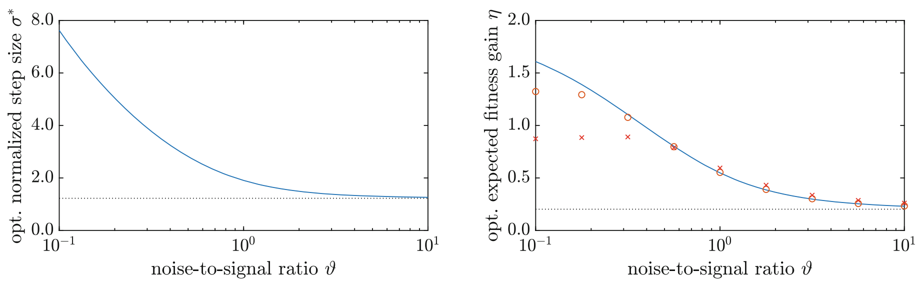

Fig. 2. Optimal normalized step size and resulting expected fitness gain of the surrogate model assisted (1 + 1)-ES plotted against the noise-to-signal ratio. The solid lines represent results obtained analytically in the limit $n \to \infty$. The dots show values observed experimentally for $n = 10$ (crosses) and $n = 100$ (circles). The dotted lines represent the optimal values for the (1 + 1)-ES without surrogate model assistance.

except for the case of zero noise-to-signal ratio, the false positive rate increases with increasing step size. The effect on the expected fitness gain (that accounts for computational costs) is such that for $\vartheta > 0$ the gain peaks at a finite value of σ^*. With increasing noise-to-signal ratio, the expected fitness gain decreases. For $\vartheta \to \infty$ the surrogate model becomes useless and the corresponding relationship for the (1 + 1)-ES without surrogate model assistance first derived by Rechenberg [15] is recovered (dotted line in the left hand plot in Fig. 1). That strategy achieves a maximal expected fitness gain of 0.202 at a normalized step size of $\sigma^* = 1.224$. For moderate values of ϑ, the surrogate model assisted algorithm is capable of achieving much larger expected fitness gain values at larger step sizes (e.g., for $\vartheta = 1.0$, the maximal achievable expected fitness gain is 0.548 and is achieved at a normalized step size of $\sigma^* = 1.905$). For $\vartheta = 0$ (i.e., a perfect surrogate model), both the optimal normalized step size and the expected fitness gain with increasing step size tend to infinity. However, it is important to keep in mind that the analytical results have been derived in the limit of $n \to \infty$ and merely are approximations in the finite-dimensional case. Figure 2 illustrates the dependence of the optimal normalized step size on the noise-to-signal ratio derived in the limit $n \to \infty$ and shows values of the expected fitness gain achieved with that step size, both derived analytically for $n \to \infty$ and measured experimentally for $n \in \{10, 100\}$. In the finite-dimensional cases the speed-up achieved through surrogate model assistance for small noise-to-signal ratios appears to top out between four and five for $n = 10$ and between six and seven for $n = 100$. Notice that these values are roughly in line with speed-ups reported for surrogate model assisted CMA-ES variants mentioned in Sect. 2.

4 Step Size Adaptation and Experiments

In this section we propose a step size adaptation mechanism for the surrogate model assisted (1 + 1)-ES. We then evaluate the algorithm by using a

Surrogate Model Assisted (1+1)-ES

Generate standard normally distributed $\mathbf{z} \in \mathbb{R}^n$ 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

Fig. 3. Single iteration of the surrogate model assisted (1 + 1)-ES.

Gaussian Process surrogate model in place of the simple model for surrogate models employed in Sect. 3 and applying it to several test functions.

The step size of the (1+1)-ES is commonly adapted using the 1/5th rule proposed by Rechenberg [15]. That rule stipulates that the step size of the strategy can be adapted based on the "success rate" (i.e., the probability of the parent being replaced by the offspring candidate solution). If this rate exceeds one fifth then the step size is increased; if it is below one fifth then the step size is decreased. An ingenious implementation of that rule has been proposed by Kern et al. [10]: rather than approximating the success rate by counting successes over a number of iterations, increase the step size by multiplication with $\mathrm{e}^{0.8/D}$ in each iteration where the offspring is successful; decrease it by multiplication with $\mathrm{e}^{-0.2/D}$ whenever the parent prevails. Constant D controls the magnitude of the step size updates and according to Hansen et al. [6] can be set to $\sqrt{1+n}$. If one out of every five offspring generated is successful, then the step size updates cancel each other out on average and the logarithm of the step size remains unchanged. If the success rate exceeds one fifth, then increasing updates occur more frequently and the step size will systematically increase and vice versa.

The one-fifth rule is not suitable for the adaptation of the step size of the surrogate model assisted (1+1)-ES. From Fig. 1, there is no single value of either the evaluation rate or the false positive rate (both of which are observable) such that optimal values of the expected fitness gain are obtained near those rates, for all values of the noise-to-signal ratio that the strategy may operate under. However, we suggest that the step size can be adapted by considering a *combination* of those rates and propose the algorithm shown in Fig. 3. Nonnegative constants c_1, c_2, and c_3 remain to be determined below. The algorithm decreases the step size (potentially by differing rates) if the offspring candidate solution is rejected either based on the objective function value estimate provided by the

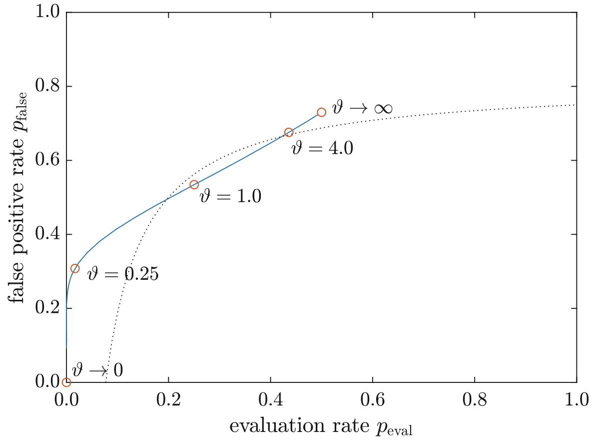

Fig. 4. False positive rate of the surrogate model assisted (1+1)-ES plotted against the evaluation rate. The solid line represents the optimally performing strategy under the conditions from Sect. 3, the dotted line the solution of Eq. (8) for $c_1 = 0.05$, $c_2 = 0.2$, $c_3 = 0.6$.

surrogate model or on the exact value returned by the objective function; it is increased if the offspring candidate solution is successful.

To choose values for the constants in the algorithm in Fig. 3, consider Fig. 4. The solid line in that plot has been obtained by using Eq. (6) to numerically determine the optimal normalized step size for values of the noise-to-signal ratio that vary from the very small to the very large. Corresponding values of the evaluation rate and the false positive rate were then obtained from Eqs. (4) and (5) and plotted against each other to obtain the solid curve in the plot. Considering the algorithm in Fig. 3, the step size will be unchanged in expectation if

$$-(1 - p_{\text{eval}})c_1 - p_{\text{eval}}p_{\text{false}}c_2 + p_{\text{eval}}(1 - p_{\text{false}})c_3 = 0. \tag{8}$$

The solution of Eq. (8) defines a branch of a hyperbola that is shown with a dotted line in Fig. 4 for the case that $c_1 = 0.05$, $c_2 = 0.2$, and $c_3 = 0.6$. If the combination of evaluation rate and false positive rate falls above the dotted line, then the logarithm of the step size will decrease in expectation; if it falls below, then the step size will increase. One could attempt to tune parameters c_1, c_2, and c_3 to better match the solid curve in the figure. However, the likely inaccuracy of the simple model for surrogate models employed in Sect. 3 may render such efforts futile. For example, biased surrogate models would result in a shift of the solid curve either to the left or to the right.

In order to test the step size adaptation mechanism thus proposed, we use a set of five ten-dimensional test problems: sphere functions $f(\mathbf{x}) = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{\alpha/2}$ for $\alpha \in \{1, 2, 3\}$ that we refer to as linear, quadratic, and cubic spheres, Schwefel's Problem 1.2 with $f(\mathbf{x}) = \sum_{i=1}^{n}(\sum_{j=1}^{i} x_j)^2$ (a convex quadratic function with condition number of the Hessian approximately equal to 175.1; see [16]), and

Table 1. Median test results.

	Median number of objective function calls		Speed-up
	Without model assistance	With model assistance	
Linear sphere	1270	503	2.5
Quadratic sphere	673	214	3.1
Cubic sphere	472	198	2.4
Schwefel's function	2367	1503	1.6
Quartic function	4335	1236	3.5

$f(\mathbf{x}) = \sum_{i=1}^{n-1}[\beta(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$ (see [3]). For $\beta = 100$ the latter function is the Rosenbrock function, the condition number of the Hessian of which at the optimizer exceeds 3,500, making it tedious to solve without adaptation of the shape of the mutation distribution. We use $\beta = 1$ instead, resulting in the condition number of the Hessian at the optimizer being 49.0, and we refer to it as the quartic function. The optimal function value for all problems is zero. We conduct 101 runs for each problem, both for the surrogate model assisted (1+1)-ES and for the strategy that does not use model assistance. For surrogate models, as Büche et al. [4], we employ Gaussian processes. We use a squared exponential kernel and for simplicity set the length scale parameter of that kernel to $8\sigma\sqrt{n}$, where σ is the step size parameter of the evolution strategy. The training set consists of the 40 most recently evaluated candidate solutions. The surrogate model assisted algorithm does not start to use surrogate models until after iteration 40. All runs are initialized by sampling the starting point from a Gaussian distribution with zero mean and unit covariance matrix and setting the initial step size to $\sigma = 1$. Runs are terminated when a solution with objective function value below 10^{-8} has been found.

Histograms showing the numbers of objective function calls used to solve the test problems to within the required accuracy are shown in the top row of Fig. 5, with median values represented in Table 1. The speed-up reported in the table is the median number of function evaluations used by the algorithm without surrogate model assistance divided by the corresponding number used by the surrogate model assisted (1 + 1)-ES. Speed-ups observed are between 1.6 for Schwefel's function and 3.5 for the quartic function. Despite the simplicity of the surrogate models, the speed-up of 3.1 observed for the quadratic sphere function is not far below the maximal speed-up between four and five expected from Fig. 2. Speed-ups observed for the linear and cubic sphere functions are below that observed for the quadratic sphere, suggesting that the Gaussian process based models are more accurate for the latter than for the former. Encouragingly, the simple step size adaptation mechanism proved successful in all runs.

Convergence graphs for the median runs are shown in the middle row of Fig. 5. Eventually linear convergence appears to be achieved in all runs. The bottom row of the figure shows values of the relative model error $|f(\mathbf{y}) - f_\epsilon(\mathbf{y})|/|f(\mathbf{y}) - f(\mathbf{x})|$, where $\mathbf{x}$ and $\mathbf{y}$ are parent and offspring candidate solutions, respectively,

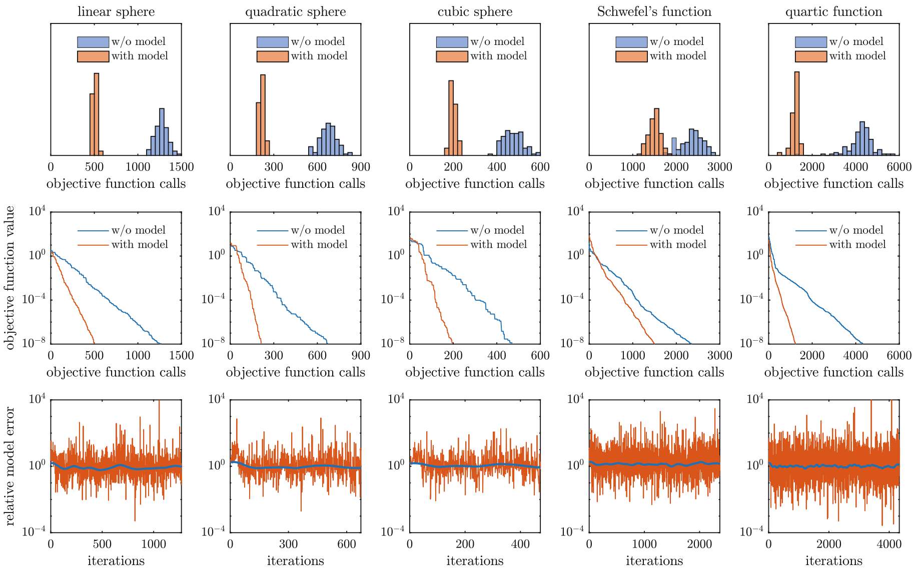

Fig. 5. Top row: Histograms showing the numbers of objective function calls used to solve the five test problems. Middle row: Convergence graphs for the median runs. Bottom row: Relative model error measured in the median runs.

observed in the median runs. The bold line in the centre of the plots represents the relative model error smoothed logarithmically by computing its convolution with a Gaussian kernel with a width of 40. We interpret the constancy of the smoothed curves as evidence that the algorithm operates under a relatively constant noise-to-signal ratio. Logarithmically averaging the relative model error across the median runs yields values between 0.786 and 0.989 for four of the five test problems, and a value of 1.292 for Schwefel's function.

5 Conclusions

To conclude, we have proposed unbiased Gaussian distributed noise as a model for surrogate modelling approaches. Using the model, we have presented an analysis of the behaviour of a surrogate model assisted $(1+1)$-ES on quadratic sphere functions. Based on that model we have proposed a step size adaptation mechanism for the surrogate model assisted $(1+1)$-ES and numerically evaluated it using a set of test functions. The mechanism successfully adapted the step size in all runs generated.

In future work, we will employ more sophisticated and possibly comparison based surrogate modelling approaches. Further goals include the development of adaptive approaches for setting the parameters c_1, c_2, and c_3 of the step size adaptation mechanism and the evaluation of the approach in the context of a $(1+1)$-ES with covariance matrix adaptation.

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

References

1. Arnold, D.V.: Noisy Optimization with Evolution Strategies. Kluwer, Dordrecht (2002)
2. Arnold, D.V., Beyer, H.-G.: A general noise model and its effects on evolution strategy performance. IEEE Trans. Evol. Comput. **10**(4), 380–391 (2006)
3. Auger, A., Hansen, N., Perez Zerpa, J.M., Ros, R., Schoenauer, M.: Experimental comparisons of derivative free optimization algorithms. In: Vahrenhold, J. (ed.) SEA 2009. LNCS, vol. 5526, pp. 3–15. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02011-7_3
4. Büche, D., Schraudolph, N.N., Koumoutsakos, P.: Accelerating evolutionary algorithms with Gaussian process fitness function models. IEEE Trans. Syst. Man Cybern. B Cybern. Part C **35**(2), 183–194 (2005)
5. Chen, Y., Zou, X.: Performance analysis of a (1+1) surrogate-assisted evolutionary algorithm. In: Huang, D.-S., Bevilacqua, V., Premaratne, P. (eds.) ICIC 2014. LNCS, vol. 8588, pp. 32–40. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09333-8_4
6. Hansen, N., Arnold, D.V., Auger, A.: Evolution strategies. In: Kacprzyk, J., Pedrycz, W. (eds.) Springer Handbook of Computational Intelligence, pp. 871–898. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-43505-2_44
7. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evol. Comput. **9**(2), 159–195 (2001)
8. Jin, Y.: Surrogate-assisted evolutionary computation: recent advances and future challenges. Swarm Evol. Comput. **1**(2), 61–70 (2011)
9. Kern, S., Hansen, N., Koumoutsakos, P.: Local meta-models for optimization using evolution strategies. In: Runarsson, T.P., Beyer, H.-G., Burke, E., Merelo-Guervós, J.J., Whitley, L.D., Yao, X. (eds.) PPSN 2006. LNCS, vol. 4193, pp. 939–948. Springer, Heidelberg (2006). https://doi.org/10.1007/11844297_95
10. Kern, S., Müller, S.D., Hansen, N., Büche, D., Ocenasek, J., Koumoutsakos, P.: Learning probability distributions in continuous evolutionary algorithms – a comparative review. Nat. Comput. **3**(1), 77–112 (2004)
11. Loshchilov, I.: Surrogate-Assisted Evolutionary Algorithms. PhD thesis, Université Paris Sud - Paris XI (2013)
12. Loshchilov, I., Schoenauer, M., Sebag, M.: Comparison-based optimizers need comparison-based surrogates. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN 2010. LNCS, vol. 6238, pp. 364–373. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15844-5_37
13. Loshchilov, I., Schoenauer, M., Sebag, M.: Intensive surrogate model exploitation in self-adaptive surrogate-assisted CMA-ES. In: Genetic and Evolutionary Computation Conference – GECCO 2013, pp. 439–446. ACM Press (2013)
14. Pitra, Z., Bajer, L., Repický, J., Holena, M.: Overview of surrogate-model versions of covariance matrix adaptation evolution strategy. In: Genetic and Evolutionary Computation Conference Companion, pp. 1622–1629. ACM Press (2017)
15. Rechenberg, I.: Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag, Stuttgart (1973)

16. Schwefel, H.-P.: Numerical Optimization of Computer Models. Wiley, Hoboken (1981)
17. Teytaud, O., Gelly, S.: General lower bounds for evolutionary algorithms. In: Runarsson, T.P., Beyer, H.-G., Burke, E., Merelo-Guervós, J.J., Whitley, L.D., Yao, X. (eds.) PPSN 2006. LNCS, vol. 4193, pp. 21–31. Springer, Heidelberg (2006). https://doi.org/10.1007/11844297_3
18. Ulmer, H., Streichert, F., Zell, A.: Model-assisted steady-state evolution strategies. In: Cantú-Paz, E. (ed.) GECCO 2003. LNCS, vol. 2723, pp. 610–621. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45105-6_72