Gaussian Adaptation Revisited – An Entropic View on Covariance Matrix Adaptation

Christian L. Müller and Ivo F. Sbalzarini

Institute of Theoretical Computer Science and Swiss Institute of Bioinformatics,
ETH Zurich, CH-8092 Zurich, Switzerland
christian.mueller@inf.ethz.ch, ivos@ethz.ch
http://www.mosaic.ethz.ch

Abstract. We revisit Gaussian Adaptation (GaA), a black-box optimizer for discrete and continuous problems that has been developed in the late 1960's. This largely neglected search heuristic shares several interesting features with the well-known Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and with Simulated Annealing (SA). GaA samples single candidate solutions from a multivariate normal distribution and continuously adapts its first and second moments (mean and covariance) such as to maximize the entropy of the search distribution. Sample-point selection is controlled by a monotonically decreasing acceptance threshold, reminiscent of the cooling schedule in SA. We describe the theoretical foundations of GaA and analyze some key features of this algorithm. We empirically show that GaA converges log-linearly on the sphere function and analyze its behavior on selected non-convex test functions.

Keywords: Gaussian Adaptation, Entropy, Covariance Matrix Adaptation, Evolution Strategy, Black-Box Optimization.

1 Introduction

High-dimensional, non-convex, and noisy optimization problems are commonplace in many areas of science and engineering. In many cases, the applied search algorithms have to operate in a black-box scenario, where only zeroth-order information about the objective is available. Such problems can usually only be tackled by stochastic search heuristics, such as Simulated Annealing (SA) [1] or Evolutionary Algorithms (EA). For non-convex, real-valued objective functions, Evolution Strategies (ES), a subclass of EA's, are nowadays the preferred optimization paradigm. A particularly successful example is the Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) [2].

In the present paper we revisit Gaussian Adaptation (GaA), a stochastic design-centering and optimization method for discrete and continuous problems that has been introduced and developed since the late 1960s by Gregor Kjellström [3,4]. Although the method shares several interesting features with CMA-ES and with Simulated Annealing, it has been largely neglected in the optimization literature. It is the scope of this work to reintegrate GaA into the field of optimization, as it builds on theoretical concepts that might prove valuable also for other search heuristics. We hereby focus on GaA for continuous sampling and optimization.

Gaussian Adaptation is a stochastic black-box optimizer that works on discontinuous and noisy functions, where gradients or higher-order derivatives may not exist or are not available. During exploration of the search space, GaA samples *single* candidate solutions from a multivariate normal distribution and iteratively updates the first and second moments of the sampling distribution. While the selection mechanism and moment adaptation of CMA-ES are intended to increase the likelihood of sampling better candidate solutions, GaA adapts the moments such as to maximize the entropy of the search distribution under the constraint that acceptable search points are found with a predefined, fixed hitting (success) probability. If minimization of an objective function is considered, sample-point selection (acceptance) is controlled by a monotonically decreasing, fitness-dependent threshold, reminiscent of the cooling schedule in SA. This ensures that the algorithm focuses on regions with better fitness values.

In order to facilitate understanding of the GaA algorithm and to highlight the key differences in the mean and covariance matrix adaptation strategies, we briefly review several variants of CMA-ES in the following section. We then describe the theoretical foundations of GaA, the algorithmic flow, and the strategy parameter settings. Section 3 illustrates the convergence of GaA on the sphere function and its dependence on search space dimensionality. In order to demonstrate the efficiency of the covariance matrix update, we also report convergence results on Rosenbrock's function. We conclude Section 3 by sketching the maximum entropy behavior of GaA using a test function introduced by Kjellström. Section 4 discusses the obtained results and concludes this work by formulating some theoretical challenges around GaA.

2   Covariance Matrix Adaptation and Gaussian Adaptation

This section summarizes the key concepts of ES with Covariance Matrix Adaptation. Equipped with these preliminaries we then outline the canonical Gaussian Adaption algorithm as developed by Kjellström and co-workers and propose a general parametrization, constraint handling, and initialization protocol.

2.1   Evolution Strategies with Covariance Matrix Adaptation

Standard CMA-ES [2], [5] is a $(\mu/\mu_{\mathrm{w}},\lambda)$-ES that uses *weighted intermediate recombination*, cumulative step size adaptation, and a combination of rank-μ update and rank-one update for the covariance adaptation [6]. At each iteration of the algorithm, the λ members of the candidate population are sampled from a multivariate normal distribution $\mathcal{N}$ with mean $\mathbf{m} \in \mathbb{R}^n$ and covariance $\mathbf{C} \in \mathbb{R}^{n \times n}$. The sampling radius is controlled by the overall standard deviation (step size) σ. Let $\mathbf{x}_k^{(g)}$ the kth individual at generation g. The new individuals at generation $g + 1$ are sampled as:

$$\mathbf{x}_k^{(g+1)} \sim \mathbf{m}^{(g)} + \sigma^{(g)}\mathcal{N}\Big(\mathbf{0}, \mathbf{C}^{(g)}\Big) \quad k = 1, \ldots, \lambda. \tag{1}$$

Selection is done by ranking the λ sampled points in order of ascending fitness and retaining the μ best. This procedure renders the algorithm invariant to strictly monotonic transformation of the objective function. The mean of the sampling distribution

given in Eq. 1 is updated using weighted intermediate recombination of the selected points. The covariance matrix for the next generation is adapted using a combination of rank-μ and rank-one update (see [5] for details). The fundamental objective behind the covariance adaptation scheme is to increase the likelihood of finding good samples in the next generation. In addition, self-adaptation of the step size σ enables the algorithm to explore the search space at different scales. Standard settings for the strategy parameters of CMA-ES have been derived from theoretical and empirical studies (see [6] for a comprehensive summary). The restart variant of CMA-ES with iteratively increasing population size (IPOP-CMA-ES) [7] can be considered a parameter-free CMA-ES.

The (1+1)-variant of CMA-ES establishes a direct link to GaA by combining the classical (1+1)-ES with a rank-one update of the covariance matrix [8]. In the next subsection, we show that Gaussian Adaptation has been designed in a similar spirit, yet grounding its theoretical justification on a different foundation.

2.2 Gaussian Adaptation

Gaussian Adaptation has been developed in the context of electrical network design. There, the key goal is to find an optimal setting of design parameters $\mathbf{x} \in \mathbb{R}^n$, e.g., nominal values of resistances and capacities in an analog network, that fulfill two requirements. First, the parameter settings satisfy the specifications imposed by the engineer, i.e. some (real-valued) objective (or criterion) function $f(\mathbf{x})$ applied to the network output, and second, the nominal values should be robust with respect to intrinsic random variations of the components during operation of the electrical device. Kjellström realized that with increasing network complexity classical optimizers such as conjugate gradients perform poorly, especially when analytical gradients are not readily available or when the objective function is multimodal. He suggested to search the space of valid parameter settings with stochastic methods that only rely on evaluations of the objective function. Starting from an exploration method that can be considered an adaptive random walk through design space [3], he refined his algorithm to what he called Gaussian Adaptation [4].

Before turning to the problem of optimization, Kjellström considered the following simpler situation: Assume that the engineer of an electrical circuit can vary the set of design parameters and can decide whether these settings fulfill a specified criterion or not. How can one describe the set $\mathcal{A} \subset \mathbb{R}^n$ of acceptable solutions in a general and compact manner? Based on Shannon's information theory, Kjellström derived that under the assumption of finite mean $\mathbf{m}$ and covariance $\mathbf{C}$ of the samples, a Gaussian distribution may be used to characterize $\mathcal{A}$ [4]. Although not specifically stated in the original publication, Kjellström applied the maximum entropy principle, developed by Jaynes in 1957 [9]. This principle is a type of statistical inference that gives the least biased estimate possible on the given information. In the case of given mean and covariance information, the Gaussian distribution maximizes the entropy $\mathcal{H}$, and hence is the preferred choice to describe the region of acceptable points. The entropy of a multivariate Gaussian distribution is:

$$\mathcal{H}(\mathcal{N}) = \log\left(\sqrt{(2\pi e)^n \det(\mathbf{C})}\right) , \tag{2}$$
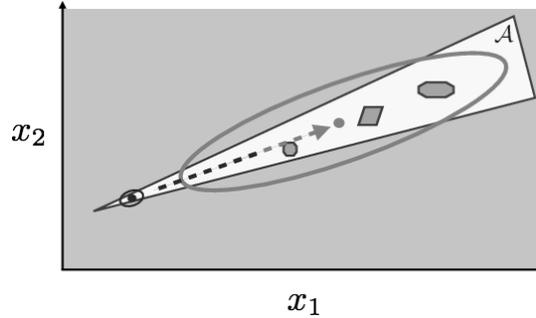
Fig. 1. Illustration of Gaussian Adaptation. The white, non-convex region defines the acceptable region $\mathcal{A}$ in a 2D design-parameter space $\mathbf{x}$. Both the left (dark) and right (light gray) dots and ellipsoids represent the means and covariances of two Gaussian distributions with the same hitting probability P. GaA moves away from the boundary toward the center and adapts the distribution to the shape of $\mathcal{A}$.

where $\mathbf{C}$ is the covariance matrix. In order to get the most informative characterization of the region $\mathcal{A}$, Kjellström envisioned an iterative sampling strategy with a Gaussian distribution that satisfies the following criteria: (i) The probability of finding a feasible design parameter set should be fixed to a predefined value $P < 1$, and (ii) the spread of the samples quantified by their entropy should be maximized. As Eq. 2 shows, this can be achieved by maximizing the determinant of the covariance matrix. In the situation where the parameters have to fulfill a predefined static criterion, the iterative sampler should push the mean of the distribution toward the *center* of the feasible design space. Simultaneously, it should adapt the orientation and scale of the covariance matrix to the shape of $\mathcal{A}$ under the constraint of the fixed hitting probability. The final mean can, e.g., be used as the nominal design parameter set. Fig. 1 illustrates this process, which is called "design centering" or "design tolerancing" in electrical engineering.

When the criterion function $f(\mathbf{x})$ yields real values, the sampler can be turned into a minimizer by introducing a fitness acceptance threshold c_T that is monotonically lowered until some convergence criteria are met. A similar idea has later also been employed in the popular Simulated Annealing algorithm [1].

The GaA Algorithm. In order to realize an iterative procedure that works both on design-tolerancing and optimization problems, Kjellström proposed the Gaussian Adaptation method. The process starts by setting the mean $\mathbf{m}^{(0)}$ of a multivariate Gaussian to an initial point $\mathbf{x}^{(0)} \in \mathcal{A}$. The covariance $\mathbf{C}^{(g)}$ is decomposed as follows:

$$\mathbf{C}^{(g)} = \left(r \cdot \mathbf{Q}^{(g)} \right) \left(r \cdot \mathbf{Q}^{(g)} \right)^T = r^2 \left(\mathbf{Q}^{(g)} \right) \left(\mathbf{Q}^{(g)} \right)^T , \tag{3}$$

where r is the scalar step size and $\mathbf{Q}^{(g)}$ is the normalized square root of $\mathbf{C}^{(g)}$. Like in CMA-ES, $\mathbf{Q}^{(g)}$ is found by eigendecomposition of the covariance matrix $\mathbf{C}^{(g)}$. The initial $\mathbf{Q}^{(0)}$ is set to the identity matrix $\mathbf{I}$. A point in iteration $g + 1$ is sampled from a Gaussian distribution according to:

$$\mathbf{x}^{(g+1)} = \mathbf{m}^{(g)} + r^{(g)} \mathbf{Q}^{(g)} \eta^{(g)} , \tag{4}$$

where $\eta^{(g)} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$. The new sample is evaluated by the criterion function $f(\mathbf{x}^{(\mathbf{g+1})})$. *Only if* the sample fulfills the specification, i.e. $\mathbf{x}^{(g+1)} \in \mathcal{A}$ in the design-tolerancing scenario or $f(\mathbf{x}^{(g+1)}) < c_{\mathrm{T}}^{(g)}$ in the optimization scenario, the following adaptation rules are applied: The step size r is increased according to $r^{(g+1)} = ss \cdot r^{(g)}$, where $ss > 1$ is called the expansion factor. The mean is updated via

$$\mathbf{m}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{m}}}\right)\mathbf{m}^{(g)} + \frac{1}{N_{\mathrm{m}}}\mathbf{x}^{(g+1)}. \tag{5}$$

N_{m} is a weighting factor that controls how fast the mean is shifted. The covariance matrix is updated through:

$$\mathbf{C}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{C}}}\right)\mathbf{C}^{(g)} + \frac{1}{N_{\mathrm{C}}}\left(\mathbf{x}^{(g+1)} - \mathbf{x}^{(g)}\right)\left(\mathbf{x}^{(g+1)} - \mathbf{x}^{(g)}\right)^{T}. \tag{6}$$

N_{C} weights the influence of the accepted sample point on the covariance adaptation. Kjellström introduced an alternative update rule that is mathematically equivalent to Eq. 6, but numerically more robust. It acts directly on the square root $\mathbf{Q}^{(g)}$ of the covariance matrix:

$$\mathbf{\Delta C}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{C}}}\right)\mathbf{I}^{(g)} + \frac{1}{N_{\mathrm{C}}}(\eta^{(g)})(\eta^{(g)})^{T}, \quad \mathbf{\Delta Q}^{(g+1)} = (\mathbf{\Delta C}^{(g+1)})^{\frac{1}{2}}. \tag{7}$$

$\mathbf{Q}^{(g+1)}$ is then updated as $\mathbf{Q}^{(g+1)} = \mathbf{Q}^{(g)}\mathbf{\Delta Q}^{(g+1)}$. In order to decouple the volume of the covariance (controlled by $r^{(g+1)}$) and its orientation, $\mathbf{Q}^{(g+1)}$ is normalized such that $\det(\mathbf{Q}^{(g+1)}) = 1$. As in CMA-ES, the full adaptation of the covariance matrix gives GaA the appealing property of being invariant to arbitrary rotations of the problem.

In case $\mathbf{x}^{(g+1)}$ is not accepted at the current iteration, only the step size is adapted by $r^{(g+1)} = sf \cdot r^{(g)}$, where $sf < 1$ is the contraction factor.

A crucial ingredient for optimization using GaA is the adaptation of the acceptance threshold c_{T}. Kjellström suggested the following rule:

$$\mathbf{c_{\mathrm{T}}}^{(g+1)} = \left(1 - \frac{1}{N_{\mathrm{T}}}\right)\mathbf{c_{\mathrm{T}}}^{(g)} + \frac{1}{N_{\mathrm{T}}}f(\mathbf{x}^{(g+1)}), \tag{8}$$

where N_{T} controls the weighting between the old threshold and the objective value of the *accepted* sample. It can readily be seen that this fitness-dependent threshold update leaves the algorithm invariant to linear transformations of the objective function.

Strategy Parameters in GaA. Gaussian Adaptation comprises strategy parameters that influence its exploration behavior. We outline a standard parameter setting that is expected to work for a large class of design centering and optimization problems. We first consider the hitting (acceptance) probability P. Kjellström investigated the information-theoretic optimality of P for a random walk in a simplex region [3] and for Gaussian Adaptation in general regions [4]. In both cases, he concluded that the efficiency of the process and P are related as $E \propto -P \log P$, leading to $P = \frac{1}{e} \approx 0.3679$, where e is Euler's number. A proof is provided in [10]. Maintaining the desired hitting probability

corresponds to leaving the volume of the distribution, $\det(\mathbf{C})$, constant under stationary conditions. As $\det(\mathbf{C}) = r^{2n} \det(\mathbf{Q}\mathbf{Q}^T)$, the expansion and contraction factors ss and sf increase or decrease the volume by a factor of ss^{2n} and sf^{2n}, respectively. After S successful and F failed samples, a necessary condition for constant volume thus is:

$$\prod_{i=1}^{S}(ss)^{2n} \prod_{i=1}^{F}(sf)^{2n} = 1 \,. \tag{9}$$

Using $P = \frac{S}{S+F}$ and introducing a small $\beta > 0$, one can verify that $ss = 1 + \beta(1 - P)$ and $sf = 1 - \beta P$ satisfies Eq. 9 to first order. The scalar rate β is coupled to the strategy parameters N_C and N_T, but not to N_m. As N_m influences the update of $\mathbf{m} \in \mathbb{R}^n$, it is reasonable to set $N_m \propto n$. In this study we propose $N_m = en$. A similar reasoning is employed for N_C and N_T. N_C influences the update of $\mathbf{C} \in \mathbb{R}^{n \times n}$ that contains n^2 entries. Hence, N_C should be proportional to n^2. Kjellström suggests using $N_C = \frac{(n+1)^2}{\log(n+1)}$ as a standard value, and coupling $N_T = \frac{N_C}{2}$ and $\beta = \frac{1}{N_C}$ [11].

It is noteworthy that the simple (1+1)-ES is a limit case of GaA. Setting $N_m = N_T = 1$ moves GaA's mean directly to the accepted sample and c_T to the fitness of the accepted sample. For $N_C \to \infty$, the covariance stays completely isotropic and GaA becomes equivalent to the (1+1)-ES with a Pth-success rule. Keeping N_C finite results in an algorithm that is almost equivalent to the (1+1)-CMA-ES [8]. Slight differences, however, remain in deciding when to update the covariance and how to adapt the step size. Moreover, (1+1)-CMA-ES does not normalize the volume of the covariance matrix. Replacing GaA's acceptance rule by a probability based on Boltzmann-weighted fitness differences, and setting $N_C \to \infty$, makes GaA equivalent to SA.

Constrained Handling and Initialization. In the context of box-constrained optimization, where the boundaries are explicitly given by $\mathbf{x} \in [\mathbf{A}, \mathbf{B}] \subset \mathbb{R}^n$, several boundary handling techniques can be employed. We suggest projecting the components of samples that violate the constraints onto the boundary along the coordinate axes, and evaluating the projected samples. In the case of box constraints, the initial mean $\mathbf{m}^{(0)}$ is drawn from a uniform distribution in the box. The initial step size is set to $r^{(0)} = 1/e\,(\max \mathbf{B} - \min \mathbf{A})$, similar to the global search setting of the initial σ in CMA-ES [7]. The initial threshold $c_T^{(0)}$ is set to $f(\mathbf{m}^{(0)})$.

3 Numerical Examples

We show the convergence of GaA on quadratic functions by considering the sphere function. The covariance matrix adaptation mechanism is then demonstrated on Rosenbrock's function. Finally, we sketch the entropic behavior of GaA on a multimodal function introduced by Kjellström [12].

3.1 Gaussian Adaptation on the Sphere Function

We consider the sphere function as a prototypical quadratic function in order to study convergence of the GaA algorithm. It is defined as:

$$f_{\text{Sphere}}(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \,. \tag{10}$$

The global minimum is at the origin $\mathbf{0}$ with $f_{\text{Sphere}}(\mathbf{0}) = 0$. For practical purposes, search is restricted to $\mathbf{x} \in [-5, 5]^n$. In order to study the dimension-dependence of GaA's convergence properties, we use the standard strategy parameter values, constraint handling, and initialization. 10 repetitions are conducted for dimensions $n = 2, 5, 10, 20, 30, 40, 50$. Fig. 2 summarizes the results. We observe the expected log-linear convergence of GaA on the sphere function. Nevertheless, we show the results in a log-log plot in order to better discriminate the trajectories for different dimensions (Fig. 2a). The mean number of function evaluations (FES) needed to achieve an accuracy of 10^{-9} grows slightly faster than quadratically with n (Fig. 2b). The measured (empirical) hitting probability $\hat{P}$ approaches the optimal $P = 1/e$ with increasing dimension. A least-squares fit of a power law yields $\hat{P}(n) = -0.2077n^{-0.1831} + 0.4265$ (Inset in Fig. 2b).
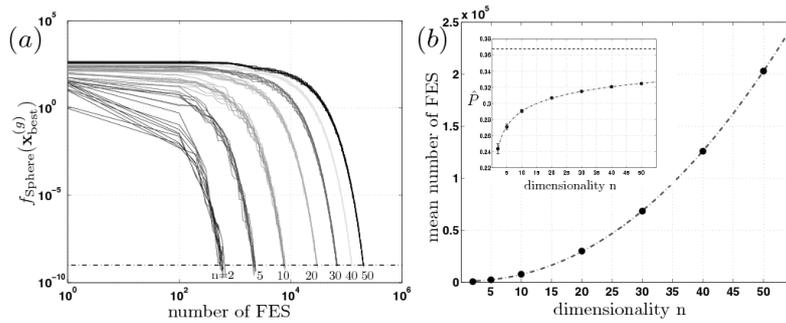

Fig. 2. (a) Log-log plot of the current best fitness value $f(\mathbf{x}_{\text{best}}^{(g)})$ vs. the number of function evaluations (FES) on the sphere function for $n = 2, 5, 10, 20, 30, 40, 50$ (from left to right). The dashed line shows the target fitness (stopping criterion). (b) Average number of FES needed to reach the target fitness vs. dimensioniality n. The dashed curve is a perfect fit of the power law $\text{FES}(n) = 47.11n^{2.138} + 857.8$. The inset shows the mean and standard deviation of the empirical hitting probability $\hat{P}$ vs. n. The dashed line represents the desired optimal $P = 1/e$.

3.2   Gaussian Adaptation on Rosenbrock's Function

We study the behavior of the covariance matrix adaptation scheme on Rosenbrock's valley function, defined as:

$$f_{\text{Rosen}}(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right) \,. \tag{11}$$

The global minimum is at $\mathbf{1}$ with $f_{\text{Rosen}}(\mathbf{1}) = 0$. Search is constrained to $\mathbf{x} \in [-2, 2]^n$. Rosenbrock's function is multimodal for $n > 3$, and it exhibits an interesting topology. On a global length scale ($\|x_i\| > 1$), the first summand dominates and attracts
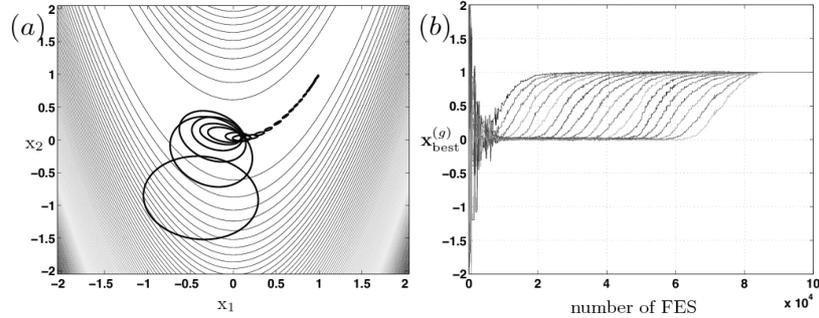
Fig. 3. Typical trajectory of GaA on Rosenbrock's function for $n = 20$. (a) 2D Contour plot with the sub-covariances along the first two dimensions (black ellipses) shown every 1000 iterations. (b) Evolution of the components of $\mathbf{x}_{\text{best}}^{(g)}$ vs. the number of FES (= iterations) for the same run.

most search heuristics toward the origin. On smaller length scales ($\|x_i\| \ll 1$), however, the second term dominates and forms a bent parabolic valley that leads from the origin to the global minimum **1**. Hence, it becomes favorable to constantly re-orient the covariance matrix along the valley. We perform 10 optimization runs with the same protocol for $n = 2, 5, 10, 20, 30, 40$. GaA finds the global minimum in all cases. Similar to CMA-ES, GaA's search on Rosenbrock can be divided into three phases: (1) log-linear convergence toward the origin; (2) a plateau region for covariance adjustment along the valley; (3) log-linear convergence near the global minimum. Fig. 3 shows a typical trajectory of GaA for $n = 20$. The same qualitative behavior is observed also in all other dimensions. After rapidly approaching the origin, GaA efficiently adapts its covariance to follow the valley. The objective variables migrate, in order of increasing dimension, toward the global minimum. The mean number of function evaluations needed to achieve an accuracy of 10^{-9} follows the power law $\text{FES}(n) = 60.13n^{2.462} + 2807$, with offset, prefactor, and exponent being larger than on the sphere function. $\hat{P}$, however, converges toward the optimal value faster than on the sphere function: $\hat{P}(n) = -0.1405n^{-0.917} + 0.3582$.

3.3   Gaussian Adaptation on Kjellström's Function

We consider the highly multimodal test function f_{Kjell} as introduced by Kjellström [12]. This function allows demonstrating under which circumstances a maximum-entropy method such as GaA is effective and efficient. It is defined as:

$$f_{\text{Kjell}}(\mathbf{x}) = \prod_{i=1}^{n}(1 + h(x_i)), \quad h(x_i) = 0.01 \sum_{j=1}^{5}[\cos(jx_i + b_j)], \tag{12}$$

with $\mathbf{b} = [b_1, \ldots, b_5] = [1.982, 5.720, 1.621, 0.823, 3.222]$ and $\mathbf{x} \in [0, 2\pi]^n$. The n-dimensional f_{Kjell} is the Cartesian product of n 1D functions. We thus first consider f_{Kjell} in 1D, i.e. $\mathbf{x} = x_1$, as depicted in Fig. 4a. In 1D, f_{Kjell} has 5 minima. The global minimum is at $\mathbf{x}_{\min} \approx 2.3486$ with $f_{\text{Kjell}}(\mathbf{x}_{\min}) \approx 0.9692$. The global maximum $\mathbf{x}_{\max}$
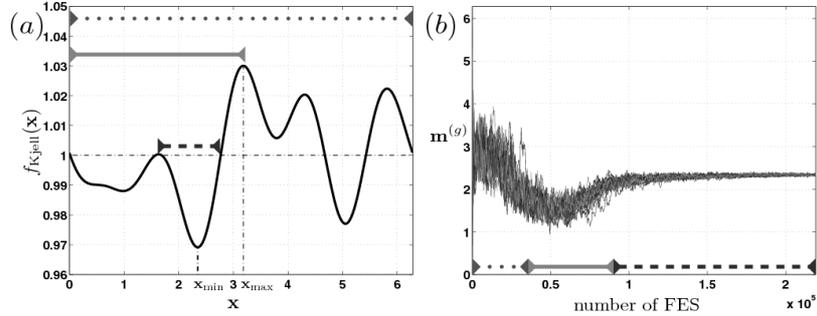
Fig. 4. (a) The multimodal function f_{Kjell} in 1D. The global minimum $\mathbf{x}_{\min}$ is contained in a locally convex region (dashed bar) that belongs to the larger (i.e., maximum entropy) sub-region of the space (solid gray bar). The global maximum $\mathbf{x}_{\max}$ separates this region from the right part of the space. (b) Typical evolution of GaA's mean $\mathbf{m}^{(g)}$ on f_{Kjell} in $n = 25$. After a global search phase (dotted bar), GaA first adapts a high-entropy distribtion to the broad region (solid gray bar) before it converges to the locally convex global-minimum region (dashed bar).

(located at a value slightly larger than $\mathbf{x} = \pi$) divides the search space into two parts. The region $\mathbf{x} < \mathbf{x}_{\max}$ covers a bit more than half of the space and therefore yields higher entropy for an adapted Gaussian distribution (solid gray bar in Fig. 4a). More-over, it contains, on average, lower function values than the other region, including the global minimum. In n dimensions, f_{Kjell} is a separable multimodal function with 5^n minima. The global minimum is within a region that allows for hierarchical adaptation of Gaussians with high entropy and includes, on average, low objective function values. Despite the staggering number of minima, the function can be solved efficiently. The standard setting for N_{C} (and the coupled parameters N_{T} and β), however, results in premature convergence of the search in one of the $5^{25} \approx 3 \cdot 10^{17}$ minima (success rate $<10\%$). A simple parameter search on N_{C} shows that the setting $N_{\mathrm{C}} = 10n^2$ is better on f_{Kjell} with respect to success rate. It leads to a 100% success rate up to $n = 50$. Fig. 4b depicts the trace of the mean $\mathbf{m}^{(g)}$ for a typical run in $n = 25$ dimensions with optimized N_{C}. In the first phase (dotted interval), GaA explores the entire search space (dotted interval in Fig. 4a). In the second phase (solid interval), it adjusts a high-entropy distribution to the center of the broad region that contains low fitness values (solid in-terval in Fig. 4a). Finally, GaA proceeds to the region that contains the global minimum with function value $f_{\mathrm{Kjell}}(\mathbf{x}_{\min}) \approx 0.9692^{25} \approx 0.4570$ (dashed interval).

4   Conclusions and Discussion

We have revisited Gaussian Adaptation (GaA), a stochastic design-centering and opti-mization method that dates back to the 1960's. We have summarized GaA's key con-cepts, including the maximum entropy principle and the adaptation of the covariance matrix. Furthermore, we have proposed a standard parameter setting, as well as con-straint handling and initialization procedures for GaA. We have empirically shown the convergence of GaA on the sphere function and the covariance adaptation mechanism

on Rosenbrock's function. Furthermore, we have re-introduced a highly multimodal function, referred to as Kjellström's function. This function can efficiently be solved by GaA despite the exponentially growing number of minima. This is due to the function's global topology, where the global minimum is located in a region that is suited to maximum-entropy adaptation. Future work will involve a more comprehensive study of GaA's convergence behavior, including a full evaluation on the CEC 2005 benchmark test suite and a convergence proof for quadratic functions. Due to GaA's theoretical foundation on the maximum entropy principle, a framework relating it to modern sampling strategies, such as Adaptive Metropolis or Slice Samplers, is also conceivable.

References

1. Kirkpatrick, S., Gelatt, C., Vecchi, M.P.: Optimization by simulated annealing. Science 220(4598), 671–680 (1983)
2. Hansen, N., Ostermeier, A.: Completely Derandomized Self-Adaption in Evolution Strategies. Evolutionary Computation 9(2), 159–195 (2001)
3. Kjellström, G.: Network Optimization by Random Variation of Component Values. Ericsson Technics 25(3), 133–151 (1969)
4. Kjellström, G., Taxen, L.: Stochastic Optimization in System Design. IEEE Trans. Circ. and Syst. 28(7) (July 1981)
5. Hansen, N., Kern, S.: Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In: Yao, X., Burke, E.K., Lozano, J.A., Smith, J., Merelo-Guervós, J.J., Bullinaria, J.A., Rowe, J.E., Tiňo, P., Kabán, A., Schwefel, H.-P. (eds.) PPSN 2004. LNCS, vol. 3242, pp. 282–291. Springer, Heidelberg (2004)
6. Hansen, N.: The CMA Evolution Strategy: A Tutorial (2007)
7. Auger, A., Hansen, N.: A restart CMA evolution strategy with increasing population size. In: Proc. of IEEE Congress on Evolutionary Computation (CEC 2005), vol. 2, pp. 1769–1776 (2005)
8. Igel, C., Suttorp, T., Hansen, N.: A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In: GECCO 2006: Proceedings of the 8th annual conference on Genetic and evolutionary computation, pp. 453–460. ACM, New York (2006)
9. Jaynes, E.T.: Information Theory and Statistical Mechanics. Phys. Rev. 106(4), 620–630 (1957)
10. Kjellström, G.: On the Efficiency of Gaussian Adaptation. J. Optim. Theory Appl. 71(3) (December 1991)
11. Kjellström, G.: Personal communication
12. Kjellström, G., Taxen, L.: Gaussian Adaptation, an evolution-based efficient global optimizer. In: Comp. Appl. Math., pp. 267–276. Elsevier Science, Amsterdam (1992)