# Local and global oder 3/2 convergence of a surrogate evolutionary algorithm

Anne Auger, Marc Schoenauer, Olivier Teytaud

# Local and Global Order 3/2 Convergence of a Surrogate Evolutionary Algorithm

Anne Auger
Computational Laboratory
Institute of Technology ETH
Hirshengraben 84, 8092
Zürich, Switzerland
anne.auger@inf.ethz.ch

Marc Schoenauer
Équipe TAO - INRIA Futurs
LRI, Bât. 490,
Université Paris-Sud
91405 Orsay Cedex. France
Marc.Schoenauer@inria.fr

Olivier Teytaud
Équipe TAO - INRIA Futurs
LRI, Bât. 490,
Université Paris-Sud
91405 Orsay Cedex. France
Olivier.Teytaud@lri.fr

## ABSTRACT

A Quasi-Monte-Carlo method based on the computation of a surrogate model of the fitness function is proposed, and its convergence at super-linear rate 3/2 is proved under rather mild assumptions on the fitness function – but assuming that the starting point lies within a small neighborhood of a global maximum. A memetic algorithm is then constructed, that performs both a random exploration of the search space and the exploitation of the best-so-far points using the previous surrogate local algorithm, coupled through selection. Under the same mild hypotheses, the global convergence of the memetic algorithm, at the same 3/2 rate, is proved.

## Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*Global optimization,Unconstrained optimization*; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: [Numerical Algorithms and Problems]

## General Terms

Algorithms, Theory

## 1. INTRODUCTION

Evolutionary Algorithms (EAs) have to find a trade-off between exploitation (of the best-so-far individuals) and exploration (of the yet-unknown regions of the search space). And though in practice the many successes of EAs witness the ability of researchers and engineers to actually find efficient compromises, theoretical results only address one of the two issues. This is the case in the discrete domain of Genetic Algorithms for instance: whether the global convergence is proved either asymptotically [16, 17] or in finite time [3], no estimation whatsoever is available for the convergence rate. On the other hand, the few results giving hints about the convergence rate only use particular (unimodal) functions [18, 4, 8]. And this is true also in the continuous framework, a domain in which the notion of convergence rate is of utter importance to practitioners using classical deterministic methods (e.g. gradient-based Newton-like methods). Most works address the issue of local convergence, assuming the unimodality of the fitness function, either from the point of view of the progress rate in Evolutions Strategies [21, 2], or with global results with linear convergence rates for quasi-convex functions [19, 1].

In this paper, two algorithms are proposed for continuous optimization. The first one is a quasi-random optimization algorithm that makes use of a surrogate model which approximates the function to optimize and introduces the minimum of this surrogate model within the population. This algorithm with a calibration obtained by a least square minimization has been developed in [14]. We will thoroughly investigate this case but will introduce two differences in the algorithm. First we make use of quasi-random numbers to calibrate the model. Quasi-random numbers have already proved to be successful in many areas one of which is the field of Monte Carlo methods allowing to speed up the convergence of those methods [5, 13]. Second the step-size update we use is also different (see Section 3). The first result of this paper is that this first algorithm exhibits a super-linear convergence rate 3/2 with mild hypotheses on the fitness function assuming that we start close enough to the global optimum of the objective function. When the calibration is made with a least-square minimization, $\Theta(dim^2)$[1] evaluations are needed to calibrate the model. Therefore to gain one order and a half of magnitude $\Theta(dim^2)$ evaluations are needed.

Many authors have proposed to couple a global stochastic algorithm with some local search. Such algorithms are often called *memetic algorithms* [12], and have been recognized as powerful techniques for combinatorial optimization in particular [7, 11] (see also the Special Issue of *Evolutionary Computation* on memetic algorithms [20]). In the continuous framework, however, very few work ni that area have been published, we refer to [10] for a recent survey on that domain.

But most work in the area of memetic algorithms use the standard "memetization" procedure: apply the local search algorithm to every offspring that is generated by the global


*GECCO'05,* June 25–29, 2005, Washington, DC, USA.


[1]Where dim is the dimension of the search space

exploration algorithm. Most of the work cited above use experimental validations of their algorithms. However, Hart [9] in the framework of Evolutionary Pattern Search proposes a convergence analysis of the algorithm designed. But the convergence rate of the algorithm remains unknown.

A slightly different coupling between exploitation and exploration has been explored by O. Franois, in which the two procedures run somehow in parallel, and the coupling is ensured by the selection algorithm alone [6]. A detailed analysis of the speed of convergence of the local algorithm and some exploration capabilities of the random global search allow to draw some conclusions about the global convergence of the resulting hybrid algorithm.

The second algorithm proposed here borrows to the later approach: global search capabilities are added to the quasi-random local algorithm; but because it has a high convergence speed, the resulting hybrid algorithm exhibits the same convergence rate than the local algorithm alone: in that sense, it can be considered as the first evolutionary algorithm for which there exists a theoretical proof of global convergence with a proved (super-linear) convergence rate.

Of course, those results are theoretical results only, and the practitioner will immediately find out that the proposed algorithm is far too costly in large dimensionality. Indeed, the exploration of the search space is a simple Monte Carlo technique using the same constant distribution allowing to cover the whole search space. Moreover, even though the order of the convergence is super-linear, the values of the constant appearing in the demonstration are far too large, and the number of required function evaluations remains unaffordable when the dimension is large. However we believe that this is a first step toward a better understanding of how to design algorithms that realize an optimal compromise between a competitive local convergence rate and an efficient global exploration capability.

The paper is organized as follows. We first consider in Section 2 a rather general global algorithm coupling local search and exploration. It constitutes the outline of the global algorithm considered next. Under certain assumption we prove that if the local search allows convergence with order $p$ then almost sure convergence occurs with order $p$ as well. In Section 3 we specify the local algorithm, building a quadratic approximation of the objective function and prove its convergence with order 3/2. Section 4 defines the global algorithm we consider and proves its convergence making use of the result of Section 2. Section 5 gives the details for the proofs of convergence with order 3/2 of the local algorithm omitted in Section 3 for the sake of clarity. Section 6 discusses those results and concludes.

## Notations and definitions

**Notations** The notation $\mathbb{R}_*^+$ will be used to denote strictly positive real numbers or $\mathbb{R}_*^+ = \mathbb{R}^+ \backslash \{0\}$.
We will denote $\|x\|_2$ (or $\|x\|$) and $\|x\|_\infty$ respectively the Euclidean norm ($\|x\|_2 = (\sum_i x_i^2)^{1/2}$) and the *sup* norm on $\mathbb{R}^{dim}$ ($\|x\|_\infty = sup_i |x_i|$).
For $(X, \sigma) \in \mathbb{R}^{dim} \times \mathbb{R}^+$, $B(X, \sigma)$ is the ball of center $X$ and radius $\sigma$, and $S(X, \sigma)$ is the sphere of center $X$ and radius $\sigma$.
Consider $\mathcal{Q}$ the vector space of quadratic forms on $\mathbb{R}^{dim}$. $A$ is a set of $\lambda - 1$ quasi-random points in $B(0, 1)$ such that the restriction of $\mathcal{Q}$ to $A$ has the same dimension as $\mathcal{Q}$. In particular this implies that $\lambda - 1 \geq 1 + dim + \frac{dim(dim+1)}{2}$.
The cardinal of a set $G$ will be denoted $|G|$.

**On the objective function $f$.** Consider a real-valued objective function $f$, to be minimized, defined on $\mathbb{R}^{dim}$, for some positive integer $dim$. We assume without loss of generality that its minimum value is 0 and is attained at least once. We will denote $X^*$ a global minimum of $f$ and $f(X^*) = 0$.
We will denote $\|\!|f|\!\|_{B(X,\sigma)}$ the following regularity measure on $f$

$$\|\!|f|\!\|_{B(X,\sigma)} = \inf_{\tilde{f} \text{ quadratic}} \sup_{x \in B(X,\sigma)} \frac{|f(x) - \tilde{f}(x)|}{\sigma^3}.$$

And define $\|\!|f|\!\|$ as $\|\!|f|\!\| = \sup_{X,\sigma} \|\!|f|\!\|_{B(X,\sigma)}$ Note that if $f$ has a bounded third derivatives on $\mathbb{R}^{dim}$, then taking for $\tilde{f}$ the second order Taylor developement of $f$ proves that $\|\!|f|\!\| \leq \frac{1}{6}\|f'''\|_\infty$

## 2. FROM FAST LOCAL CONVERGENCE TO FAST GLOBAL CONVERGENCE

Consider a general operator $op(.)$ from $\mathbb{R}^{\dim}$ to $\mathbb{R}^{\dim}$ and consider the local algorithm induced by consecutive applications of this operator, i.e. a sequence generated by this algorithm is $(op^k(x))_{k\geq 0}$. Let $p$ be a real number $\geq 1$. We assume that for this local algorithm two scenarios can happen, either convergence with order $p$ or the sequence $(op^k(x))_{k\geq 0}$ is uniformly bounded. This assumption amounts to say that the search space $\mathbb{R}^{dim}$ is split in two, we formalize this assumption as follows.

**Assumption A**.1. **Dichotomy property** *We assume that for $x \in \mathbb{R}^{dim}$ only two scenarios concerning the sequence $(op^k(x))_{k\geq 0}$ can happen, either convergence with order $p$ or the sequence is lower bounded. More formally, $\mathbb{R}^{dim} = [\mathbb{R}^{dim}]_1 \cup [\mathbb{R}^{dim}]_2$ where*

$$[\mathbb{R}^{dim}]_1 := \{x \text{ s.t. } (op^k(x))_{k\geq 0} \text{ converges with order } p\}$$
$$[\mathbb{R}^{dim}]_2 := \{x \text{ s.t. } \exists K_x > f(X^*) \text{ s.t. } f(op^k(x))_{k\geq 0} \geq K_x\}$$

For instance such a local algorithm is the one defined at section 3 and in that case $p = \frac{3}{2}$.

Now we consider the global algorithm, where at each generation we apply one iteration of $op()$ to each point of the current population, and we moreover sample new points (to ensure the exploration part) according to a distribution $\mu_x$ which allows us to see with a positive probability a neighborhood of the global optimum. The resulting algorithm is:

ALGORITHM 1 .

1. *Randomly choose an initial population $P_0$ ;*
2. *Let $n$ go from 1 to $\infty$*
   (a) $P'_n = \{op(x) | x \in P_{n-1}\} \cup G_n$ ;
   (b) $P_n$ *is the selection of the $|P_{n-1}|$ best elements of $P'_n$.*

where $G_n$ is a set of points randomly generated by sampling a distribution $\mu_x$. In the sequel, when the sequence

$(op^k(x))_{k\geq 0}$ generated by successive application of the local algorithm will be such that the selection step 2.b will select all the terms, we will say that the sequence is **non-interrupted** (otherwise the sequence is interrupted by the selection). Note that if at iteration $n$, points generated at step 2.a are selected it means that $G_n \cap P_n \neq \emptyset$. We then make the following assumptions on the global algorithm:

**Assumption A**.2. *1.* $0 < \inf_n |G_n| \leq \sup_n |G_n| < \infty$

*2. Assume that* $c\epsilon^d \leq \mathbb{P}(f(X) \leq \epsilon) \leq a\epsilon^b$*, where* $X$ *is a random variable of distribution* $\mu_x$*. In other words we assume that the random generation with the distribution* $\mu$ *allows us to see an* $\epsilon$ *neighborhood of the global optimum with a probability upper and lower bounded with respectively* $a\epsilon^b$ *and* $c\epsilon^d$*.*

*3. There exists some* $C > 0$ *such that*

$$\mathbb{P}\Big(\exists x \in P_n \cap G_n \text{ s.t. } (op^k(x))_{k\geq 0} \text{ converges to one optimum with order } p \text{ without interruption}\Big) \geq C$$

*Unformally, this assumption means that the probability of generating and selecting fast-convergence sequences is lower bounded by some* $C > 0$*.*

**Lemma** 1 . *Consider* $p$ *a real number* $\geq 1$*. Let assume that Assumptions* **A**.*1 and* **A**.*2 are satisfied, then almost surely, global convergence occurs with order* $p$*.*

Proof. In a first step we prove that with probability one, infinitely many randomly generated points at step 2.a are selected or more formally

$$\mathbb{P}(\forall n_0, \exists n > n_0 \text{ such that } P_n \cap G_n \neq \emptyset) = 1$$

Assume that the reverse is true:

$$\exists n_0 \text{ such that } \forall n \geq n_0 \; P_n \cap G_n = \emptyset$$

We can moreover assume the scenario that there is no non-interrupted sequence of points converging to the optimum with order $p$, i.e.

$$\forall n, \forall x \in P_n \; f(op^k(x)) \nrightarrow f(X^*) \text{ when } k \to \infty.$$

Indeed if there exists such a sequence, then we directly conclude to the convergence with order $p$ of the algorithm. Then, by the dichotomy property, all non-interrupted sequences of points are lower bounded in fitness, above the optimal value. In particular $\forall x \in P_{n_0}$, $f(op^{k-n_0}(x)) \geq \epsilon \; \forall k \geq 0$. Applying now Assumption **A**.2.2, we have that the probability to select elements of $G_n$ is lower bounded by $c\epsilon^d$, i.e. for all $k \geq 0$, $\mathbb{P}(P_{k-n_0} \cap G_{k-n_0} \neq \emptyset) > c\epsilon^d$. Therefore

$$\mathbb{P}(\forall n_0, \exists n > n_0 \text{ s.t. } P_n \cap G_n \neq \emptyset) = 1,$$

*i.e.* with probability one, infinitely many randomly generated points are selected. This is the end of the first step.

Let us now index those infinitely many randomly generated and selected points as $(x_{\tilde{n}}^{g,s})_{\tilde{n}\in\mathbb{N}}$. Note that by Step 1 with probability one such a sequence exists. From Assumption **A**.2.3, for all $\tilde{n} \in \mathbb{N}$, the probability that $((op^k(x_{\tilde{n}}^{g,s}))_{k\in\mathbb{N}}$ converges without interruption with order p to the optimum is lower bounded by $C$. For the following equations, we will use the abbreviation *cv. w.int. order p* for "*converges without interruption to the optimum with order p*".

$$\mathbb{P}(\text{ cv. to the optimum with order p }) = \mathbb{P}(\exists \tilde{n_0} \text{ s.t. } (op^k(x_{\tilde{n_0}}^{g,s}))_{k\in\mathbb{N}} \text{ cv. w.int. order p })$$

Besides, the right hand side of the previous equation is equal to

$$\begin{aligned}
&\mathbb{P}(\cup_{\tilde{n_0}}\{(op^k(x_{\tilde{n_0}}^{g,s}))_{k\in\mathbb{N}} \text{ cv. w.int. order p }\}) \\
=&1 - \mathbb{P}(\cap_{\tilde{n_0}}\{(op^k(x_{\tilde{n_0}}^{g,s}))_{k\in\mathbb{N}} \text{ cv. w.int. order p }\}^c) \\
\geq&1 - \Pi_{\tilde{n_0}}(1 - C) \\
=&1
\end{aligned}$$

Therefore, with probability one, this leads to a convergence with order $p$.

□

# 3. A LOCAL ALGORITHM WITH CONVERGENCE 3/2

We now consider the following algorithm, building a surrogate model of the objective function to minimize and prove its convergence rate. Let $X_0 \in \mathbb{R}^{dim}$, $\sigma_0 \in \mathbb{R}_*^+$ and $L \in \mathbb{R}^+$. The main loop of the local algorithm is the following:

Algorithm 2 .

1. *Let* $(X_n, \sigma_n)$ *be the current individual,* $B_n = B(X_n, \sigma_n)$*.*
2. *Compute* $A_n$*, a quasi-random set of* $\lambda - 1$ *points in* $B_n$*, from the (fixed) set* $A$ [2]
3. *Evaluate all points in* $A_n$ *(compute* $\{f(P), P \in A_n\}$*)*
4. *Build* $H_n$*, the least-square approximation of* $f$ *using the examples* $(P, f(P))_{P\in A_n}$*.*
5. *Compute*
$$X_n^* = argmin_{x\in B_n} H_n(x) \tag{1}$$
6.
$$X_{n+1} = argmin\{f(x), x \in A_n\} \bigcup \{X_n, X_n^*\} \tag{2}$$
7. $\sigma_{n+1} = L\sigma_n^{\frac{3}{2}}$

**Local convergence results**

The following assumptions will be assumed for the objective function $f$:

**Assumption A**.3.

1. *$f$ is lower-bounded, with at least one global minimum*
2. *There exists a finite number of global minima, each of them with a positive definite hessian*
3. $\|\!|f|\!\| < +\infty$

[2]It means that $A_n$ is computed by translating by $X_n$ and dilating with a factor $\sigma_n$ the set $A$ : $A_n = X_n + \sigma_n \times A$

Then we shall prove the following convergence result:

**Theorem** 1 . *There exists $L'$ and $\sigma'_0(L)$ such that if $L \geq L'$, if $\sigma_n < \sigma'_0(L)$, and if $X^* \in B(X_n, \sigma_n)$, then $X_{n+k}$ converges to $X^*$ with order 3/2, in the following senses:*

$$||X_{n+k} - X^*|| = O(\sigma_{n+k}) \text{ as } k \to \infty \quad (3)$$

$$|f(X_{n+k})| = O(\sigma^3_{n+k}) \text{ as } k \to \infty \quad (4)$$

*where both $\sigma_{n+k}$ and $\sigma^3_{n+k}$ are of order 3/2, i.e.*

$$\sigma_{n+k} = O(\sigma^{3/2}_{n+k-1}) \text{ as } k \to \infty \quad (5)$$

$$(\sigma_{n+k})^3 = O((\sigma^3_{n+k-1})^{3/2}) \text{ as } k \to \infty \quad (6)$$

*$L'$ only depends upon $f$ and $\sigma'_0(.)$ is a decreasing function of $L$, depending only upon $f$ and $L$.*

PROOF. We choose $L'$ such that $1/(L')^2$ is small enough in the sense of Lemma 3 (*i.e.* if $\sigma_n < 1/(L')^2$ then the assumption "small enough" for $\sigma_n$ in Lemma 3 is satisfied). Let $L \geq L'$, and let $\sigma'_0(L) = 1/L^2$.
Let $\sigma_n < \sigma'_0(L) = 1/L^2$, then $\sigma_{n+1} = L\sqrt{(\sigma_n)}\sigma_n < L\frac{1}{L}\sigma_n$, therefore

$$\sigma_{n+1} < \sigma_n. \quad (7)$$

Since $\sigma_{n+1} = L\sqrt{\sigma_n}\sigma_n$ we deduce that $L\sqrt{\sigma_n} < 1$. As

$$\sigma_{n+k} \leq (L\sqrt{\sigma_n})^k \sigma_{n+1}$$

we have that $\sigma_{n+k} \to 0$ when $k \to \infty$, and for any $k \geq 0$, $\sigma_{n+k-1} > \sigma_{n+k}$. Besides by definition of $\sigma_{n+k}$ ($\sigma_{n+k} = L(\sigma_{n+k-1})^{3/2}$ ) $\sigma_{n+k} \to 0$ when $k \to \infty$ with order 3/2. Therefore Eq. 5 and Eq. 6 are satisfied. From Lemma 3 (Eq. 7 and the choice of $L'$ give us that the assumptions of the Lemma are satisfied), we have that for any $k \geq 0$, $X^* \in B(X_{n+k}, \sigma_{n+k}/2)$. This implies first that the Assumptions of Lemma 2 are satisfied and second that if $X_n$ converges, it converges to $X^*$.
We prove now the order 3/2 for the convergence of $X_n$. We have that $\|X_{n+k+l} - X_{n+k+l-1}\|$ is upper bounded by $\sigma_{n+k+l-1}$ which is of order 3/2. Besides

$$\begin{aligned}\|X_{n+k} - \lim_{l\to\infty} X_{n+l}\| &= \|\sum_{l=1}^{\infty} X_{n+k+l} - X_{n+k+l-1}\| \\ &\leq \sum_{l=1}^{\infty} \|X_{n+k+l} - X_{n+k+l-1}\| \\ &\leq \sum_{l=1}^{\infty} \sigma_{n+k+l-1} = \sigma_{n+k} \sum_{l=1}^{\infty} (\underbrace{L\sqrt{\sigma_{n+k}}}_{<1})^l\end{aligned}$$

As $\lim_{l\to\infty} X_{n+l} = X^*$, we have shown that $||X_{n+k} - X^*|| = O(\sigma_{n+k})$ (Eq. 3). Thanks to Assumptions **A**.3.2, this convergence of order 3/2 is transposed to $f(X_{n+k})$ (Eq. 4).

□

This results relies on the two following Lemmas.

**Lemma** 2 . *There exists $K_\infty$ such that if $X^* \in B(X_n, \sigma_n/2)$, then the $L_\infty$ approximation error $\epsilon_n$ defined as $sup_{x\in B_n}|H_n(x) - f(x)|$ can be bounded as follows:*

$$\epsilon_n \leq K_\infty \sigma^3_n. \quad (8)$$

PROOF. See section 5.

**Lemma** 3 . *Assume that $\sigma_n$ is small enough and that $\sigma_{n+1} < \sigma_n$. Assume that $X^* \in B(X_n, \frac{\sigma_n}{2})$ then*

$$X^* \in B(X_{n+1}, \frac{\sigma_{n+1}}{2}).$$

PROOF. See section 5.

# 4. THE GLOBAL ALGORITHM

From the previous local algorithm, we build a global algorithm. The basic idea of the global algorithm is to make use of the previous algorithm for the local search and to ensure the exploration stage by sampling some new points of the search space independently at each generation. The reasonable assumption made for this independent sampling is that it allows to see with a lower bounded probability an epsilon-neighborhood of the global optimum. Without loss of generality, we assume that the value of the fitness function at the global optimum is zero.

To formalize this algorithm, let us now denote loc(.) the local operator which associates to a triplet $(x, \sigma, L) \in \mathbb{R}^{dim} \times \mathbb{R}^+_* \times \mathbb{R}^+$ one iteration from the local algorithm 2 (Section 3).

$$\text{loc}() : \mathbb{R}^{dim} \times \mathbb{R}^+_* \times \mathbb{R}^+ \to \mathbb{R}^{dim} \times \mathbb{R}^+_* \times \mathbb{R}^+.$$

In other words $\text{loc}(X_n, \sigma_n, L) = (X_{n+1}, L\sigma_n^{3/2}, L)$, where $X_{n+1}$ is given by Equation 2. The $k^{th}$ iterates of the operator loc() will be denoted $\text{loc}^k()$.

The reason why we include $L$ is the definition of the operator loc whereas it is not changed by Algorithm 2 is because when building the global algorithm we will generate randomly different values for $L$ (see the Random generation description of Algorithm 3 for more details).

We will now call elements of the population some triplets $(x, \sigma, L) \in \mathbb{R}^{dim} \times \mathbb{R}^+_* \times \mathbb{R}^+$. Moreover we will sometimes use the notation $f(x, \sigma, L)$ to design the fitness of $x$ first element of the triplet $(x, \sigma, L)$. In the same way we will use the notation $f(\text{loc}^k(x, \sigma, L))$.

ALGORITHM 3 (GLOBAL ALGORITHM).

1. *Randomly choose an initial population $P_0$ ;*
2. *Let $n$ go from 1 to $\infty$*
   (a) $P'_n = \{\text{loc}(x, \sigma, L) | (x, \sigma, L) \in P_{n-1}\} \cup G_n$
   (b) *$P_n$ is the selection of the $|P_{n-1}|$ best elements of $P'_n$.*

Where $G_n$ is a set of randomly generated points that we detail now:

**Random generation** For the random generation at step 2.a, we of course need to sample points of the search space $\mathbb{R}^{\text{dim}}$ but we also need to sample step-sizes $\sigma$ as well as constants $L$ that will be needed once one will apply the local search operator defined by Algorithm 2 to one such point (if it is selected at step 2.b). Therefore a finite number of triplets $(x, \sigma, L)$ are generated. We assume that the points $x$ of $\mathbb{R}^{\text{dim}}$ are generated using a same distribution that we will denote $\mu_x$. The step-sizes $\sigma$ are sampled using a distribution $\mu_\sigma$ with support in $\mathbb{R}^+_*$ and the constants $L$ are sampled using a distribution $\mu_L$ with support in $\mathbb{R}^+_*$.

The required assumptions on $\mu_x$, $\mu_\sigma$ and $\mu_L$ will be stated in Lemma 1 and Theorem 2 .

Note that at the selection step 2.b, the step-size and constant L are selected with an individual. This technique is inspired from the mutative step-size technique introduced by Rechenberg [15] and Schwefel [21].

All $P_n$ have the same cardinality. In case of equality between fitnesses at step 2.b, elements generated by application of loc(.) are preferred to randomly drawn elements.

The global convergence of this algorithm relies in fact on a more general result proved in Lemma 1 , stating that the previous construction ensures that if the local operator ensures a fast local convergence, then the global algorithm will converge globally fast as well. Section 2 states and proves this result.

**Global convergence result**

To prove the global convergence, we will use Lemma 1 .

**Assumption A**.4.

1. *There is a finite set of optima ;*
2. *The fitness function $f$ is twice differentiable and at each global optimum $X^*$, the Hessian matrix is positive definite.*
3. *$\|f\| < \infty$ (this assumption could be slightly weakened by a restriction to a neighborhood of $X^*$)*
4. *$c\epsilon^d \leq P(f(X) \leq \epsilon) \leq a\epsilon^b$, for $X$ with distribution $\mu_x$ (see Lemma 1 for the interpretation of this assumption)*
5. *For some $\alpha$ and $\beta$, for all $\delta > 0$, $\forall x \in \mathbb{R}^{dim}$ with $f(x) \leq \delta$, there exists an optimal $X^*$ such that $d(x, X^*) \leq \alpha\delta^\beta$,*
6. *In the random generation part of step 2.a, $\sigma$ is drawn according to the distribution $\mu_\sigma$ where we assume that $\mu_\sigma([\sigma_1, \sigma_2]) > 0$ for all segments $[\sigma_1, \sigma_2]$ included in $\mathbb{R}_*^+$.*
7. *In the random generation part of step 2.a, $L$ is drawn according to the distribution $\mu_L$ where we assume that $\mu_L([c, +\infty[) > 0$ for all positive constant $c$.*
8. *The selection step 2b will discard all points such that $L \geq 1/\sqrt{\sigma}$.*

**Theorem** 2 . *We assume that Assumptions **A**.3 and **A**.4 are satisfied. Then Algorithm 3 converges almost surely with order* 3/2.

Proof. The proof consists in showing that the different assumptions of Lemma 1 are satisfied and then applying Lemma 1 . We start by proving assumption **A**.1 (Dichotomy property).

Hypothesis A.4-8 (removal of points such that $L \geq \frac{1}{\sqrt{\sigma}}$) ensures the dichotomy property of lemma 1 : the sequence $f(loc^k(x, \sigma, L))$ is non-increasing and lower bounded by $f(X^*) = 0$, and so

- either it is lower bounded by a quantity $> 0$ and $x \in [\mathbb{R}^{dim}]_2$;
- or it converges to 0, and then
  - $||loc^k(x, \sigma, L) - \lim_k loc^k(x, \sigma, L)|| \to 0$ as $k \to \infty$ with order $\frac{3}{2}$ thanks to hypothesis **A**.4-8 which ensures $L < \frac{1}{\sqrt{\sigma}}$ ;
  - thanks to **A**.4-5, $\lim_k loc^k(x, \sigma, L) = (X^*, 0, L)$ ;
  - **A**.4-2 transposes the convergence of $loc^k(x, \sigma, L)$ with order 3/2 to the convergence of $f(loc^k(x, \sigma, L))$ with order 3/2.

  and in that case $x \in [\mathbb{R}^{dim}]_1$.

Assumption **A**.2.2 is satisfied with Assumption **A**.4.4.

We verify now Assumption **A**.2.3.

We have to find a lower bound on the probability of convergence with order 3/2 without interruption.

Consider some real number $M > 0$ to be specified later. Let $\Pi$ be the probability that $(loc^k(x))_{k\geq 0}$ converges with order 3/2 and $L \leq M$, for $(x, \sigma, L)$ randomly generated at step 2a and selected at step 2b. This probability depends on $n$ and $M$. With Theorem 1 we have

$$\Pi = \mathbb{P}(L \leq M \wedge (loc^k(x, \sigma, L))_{k\geq 0} \text{ converges with order} 3/2))$$

$$\Pi \geq \mathbb{P}(L' \leq L \leq M \wedge \sigma \leq \sigma_0'(L) \wedge d(x, x^*) \leq \sigma)$$

$$\Pi \geq \mathbb{P}(L' \leq L \leq M \wedge \sigma \leq \sigma_0'(M) \wedge d(x, x^*) \leq \sigma)$$

as $\sigma_0'(.)$ decreases. Consider $\sigma_1 < \sigma_0'$ to be defined later, then we can split the above probability by independence and we get

$$\Pi \geq \mathbb{P}(L' \leq L \leq M)\mathbb{P}(\sigma_1 \leq \sigma \leq \sigma_0'(M))\mathbb{P}(d(x, x^*) \leq \sigma_1)$$

Thanks to **A**.4-5,

$$\Pi \geq \underbrace{\mathbb{P}(M \geq L \geq L')}_{P_1} \underbrace{\mathbb{P}(\sigma_1 \leq \sigma \leq \sigma_0')}_{P_2} \underbrace{\mathbb{P}(f(x) \leq \sqrt[\beta]{\sigma_1/\alpha})}_{P_3}$$

Let $M$ be sufficiently large to ensure that $\mathbb{P}(L' \leq L \leq M) > 0$, then We have proved that $\Pi \geq P_1P_2P_3$. $P_1$ and $P_2$ only depend upon $f$ and are $> 0$. $P_3$ is $> 0$ and non-decreasing as $n$ increase (points are more strongly selected, which increases $P_3$). So, $\Pi$ can be lower bounded by $P_1P_2P_3^{min}$ for some $P_3^{min}$. Therefore, $\Pi$ can be lower bounded by some $\Pi_0 > 0$. We have lower bounded $\Pi$.

We now need to lower bound $\Pi'$, the probability of convergence to the optimum with order 3/2 *and* of no interruption. The probability of interruption of $k \mapsto loc^k(x)$ by one given random generation $(x', \sigma', L)$ at step $k$ is less than $\mathbb{P}(f(x') \leq f(loc^k(x)))$ (we omit $\sigma$ and $L$ for short)

Consider points $(x, \sigma, L)$ satisfying $E$.

Note that for $\sigma$ small enough, as $L$ is upper bounded by $M$, the rule $\sigma' = L\sigma^{3/2}$ ensures that $\sigma' \leq \sigma \times (\sigma^{1/2} \times M) \leq \sigma/2$. In particular, $d(x, loc^k(x)) \leq \sigma \times (1 + \frac{1}{2} + \frac{1}{4} + \dots) \leq 2\sigma$. On the other hand, for $\sigma$ small enough, $f(loc^k(x))$ is upper bounded by some $g \times d(loc^k(x), X^*)^2$ (because of the assumption on the hessian of the fitness). For $\sigma$ small enough, we have therefore established that :

$$d(loc^k(x), X^*) \leq 2\sigma/2^k$$

and

$$\mathbb{P}(interruption) \leq \sum_k a(g \times d(loc^k(x), X^*)^2)^b$$

Therefore, there exists some $\sigma_2$ such that for $\sigma \leq \sigma_2$

$$\mathbb{P}(interruption) \leq \sum_k a(g(\sigma/2^k)^2))^b$$

This is upper bounded by $1/2$ if $\sigma$ is small enough, say for $\sigma \leq \sigma_3$. We now consider $\Pi_0$ associated to $\sigma_1 \leq min(\sigma_2, \sigma_3)$, $M$ sufficiently large, and we define $E$ the event $M \geq L \geq L' \wedge \sigma_1 \leq \sigma \leq \sigma_0' \wedge f(x) \leq \sqrt[\beta]{\sigma_1/\alpha}$. Hence,

$$\begin{aligned}\Pi' &= \mathbb{P}(\textit{no interruption and convergence with order } 3/2\ ) \\ &\geq \mathbb{P}(\textit{no interruption and convergence} \\ &\textit{...with order } 3/2\ |E)\mathbb{P}(E) \\ &\geq \mathbb{P}(\textit{no interruption}|E)\mathbb{P}(E) \\ &\geq \mathbb{P}(\textit{no interruption}|E)\Pi_0 \geq \frac{1}{2}\Pi_0\end{aligned}$$

Therefore, the probability of generating a never interrupted sequence of points with convergence order $3/2$ is lower bounded by $\frac{1}{2}\Pi_0$ at each new generated point which is selected (for $n$ large enough).

All assumptions in Lemma 1 are satisfied. Hence the expected result.

□

# 5. THE DETAILED PROOFS OF LOCAL CONVERGENCE

## 5.1 Proof of Lemma 2

We want to derive bounds on $\epsilon_n = sup_{x \in B_n} |H_n(x) - f(x)|$.

Let us introduce $\tilde{f}_n$, best quadratic approximation of $f$ on $B_n$ for the $L^\infty$-norm (this is different from $H_n$, best quadratic approximation of $f$ on the sample points in $A_n$ for the $L^2$-norm). We can then write $\epsilon_n \leq \zeta_1 + \zeta_2$ with

$$\begin{aligned}\zeta_1 &= \sup_{x \in B_n} |f(x) - \tilde{f}_n(x)| \\ \zeta_2 &= \sup_{x \in B_n} |\tilde{f}_n(x) - H_n(x)|\end{aligned}$$

By definition of $\|\!|.|\!\|$ and of $\tilde{f}_n$, we have

$$\zeta_1 \leq \|\!|f|\!\|_{B_n} \sigma_n^3. \tag{9}$$

For $\zeta_2$, note first that both $\tilde{f}_n$ and $H_n$ are quadratic functions defined on $B_n$: $\tilde{f}_n$ is the best possible quadratic approximation of $f$ while $H_n$ is the best possible quadratic approximation of $f$ built from the values of $f$ associated to the points of $A_n$.

Let $D$ be $1 + dim + dim(dim+1)/2$, for $x \in B_n$, let define $h$ as follows:

$$h := \begin{cases} \mathbb{R}^{dim} \to \mathbb{R}^D \\ \quad x \to (1, \tilde{x}_1^2, \tilde{x}_2^2, \ldots, \tilde{x}_{dim}^2, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{dim}, \\ \quad \tilde{x}_1\tilde{x}_2, \tilde{x}_1\tilde{x}_3, \ldots, \tilde{x}_{dim-1}\tilde{x}_{dim}) \end{cases}$$

where $\tilde{x} = (x - X_n)/\sigma_n$.

Because $\tilde{f}_n - H_n$ is quadratic, there exists a **linear** application $w$ from $\mathbb{R}^D$ to $\mathbb{R}$ such that

$$w(h(x)) = (\tilde{f}_n - H_n)(x)\ \forall x \in \mathbb{R}^{dim} \tag{10}$$

We can now rewrite $\zeta_2$ as $\sup_{x \in B_n} |w(h(x))|$ and by introducing for all $x \in B_n$ the point $\hat{x}$ which is the closest point in $A_n$ from $x$ we get

$$\zeta_2 \leq \sup_{x \in A_n} |w(h(x))| + \sup_{x \in B_n} |w(h(x)) - w(h(\hat{x}))| \tag{11}$$

and we can bound the second term of equation 11:

$$\begin{aligned}|w(h(x)) - w(h(\hat{x}))| &= |w(h(x) - h(\hat{x}))| \\ &= \Delta |w(\frac{h(x) - h(\hat{x})}{\Delta})| \\ &\leq \Delta |w(z)| \text{ for } z = ((h(x) - h(\hat{x}))/\Delta) \in B(0,1)\end{aligned}$$

where $\Delta = \sup_{x \in B_n} \inf_{y \in A_n} d(h(x), h(y))$ is a constant only depending upon $A$ (by dilatation/translation of $B(0,1)$ and $A$, see definition of $A_n$ in algorithm 2 ).

Let us now consider the matrix $P$ whose lines are $h(P)$ for $P \in A$. Because $\lambda \geq D$ (by definition of $A$), we can extract from this matrix a $D \times D$ square matrix of rank $D$, that we will denote $P'$. The $D$ lines of $P'$ that we denote $(p_i')_{1 \leq i \leq D}$ form a basis of $\mathbb{R}^D$. Therefore each element $z \in h(B(0,1))$ is a linear combination of those lines, and can be written $\Gamma P'$ where $\Gamma$ is a $D \times 1$ vector. Hence the bound above becomes

$$\begin{aligned}|w(h(x)) - w(h(\hat{x}))| &\leq \Delta |w(\Gamma P')| \\ &\leq \Delta |w(\sum_{i=1}^D \Gamma_i p_i')| \\ &\leq \Delta |\sum_{i=1}^D \Gamma_i w(p_i')|\end{aligned}$$

For each $i$, $|w(p_i')| \leq \sup_{x \in A_n} |w(h(x))|$ and $|\sum_{i=1}^D \Gamma_i| \leq D\|\Gamma\|_2$. Therefore

$$|w(h(x)) - w(h(\hat{x}))| \leq \Delta D \|\Gamma\|_2 \sup_{x \in A_n} |w(h(x))| \tag{12}$$

But $\|\Gamma\|_2 = \|z(P')^{-1}\|_2 \leq \frac{1}{v}\|z\|_2 \leq \frac{1}{v}$, where $v$ is the smallest eigenvalue of the matrix $P'$. Hence

$$|w(h(x)) - w(h(\hat{x}))| \leq \frac{\Delta D}{v} \sup_{x \in A_n} |w(h(x))| \tag{13}$$

From Eq. 10, Eq. 11, Eq. 13 we have

$$\zeta_2 \leq (1 + \frac{\Delta D}{v}) \sup_{x \in A_n} |(\tilde{f}_n - H_n)(x)|. \tag{14}$$

We now investigate an upper bound for $\sup_{x \in A_n} |(\tilde{f}_n - H_n)(x)|$ and start by introducing $f$:

$$\sup_{A_n} |(\tilde{f}_n - H_n)| \leq \sup_{A_n} |(\tilde{f}_n - f)| + \sup_{A_n} |(f - H_n)| \tag{15}$$

Besides

$$\sup_{x \in A_n} |(f - H_n)(x)| \leq \sqrt{\sum_{x \in A_n} |(f - H_n)(x)|^2}$$

From the definition of $H_n$ (see algorithm 2 ) :

$$\sqrt{\sum_{x \in A_n} |(f - H_n)(x)|^2} \leq \sqrt{\sum_{x \in A_n} |(f - \tilde{f}_n)(x)|^2}$$

Moreover it holds

$$\sqrt{\sum_{x\in A_n} |(f-\tilde{f}_n)(x)|^2} \leq \sqrt{\lambda-1}\sup_{x\in A_n}|(f-\tilde{f}_n)(x)|$$

From the previous equation and from Eq. 15 we get

$$\begin{aligned}\sup_{A_n}|(\tilde{f}_n - H_n)| &\leq (1+\sqrt{\lambda-1})\sup_{A_n}|(\tilde{f}_n - f)| \\ &\leq (1+\sqrt{\lambda-1})\sigma_n^3 \|\!| f |\!\|_{B(X_n,\sigma_n)}\end{aligned}$$

Introducing this inequality in Eq. 14, we have

$$\zeta_2 \leq (1+\frac{\Delta D}{v})(1+\sqrt{\lambda-1})\sigma_n^3 \|\!| f |\!\|_{B(X_n,\sigma_n)} \qquad (16)$$

Therefore with Eq. 9 we get

$$\zeta_1+\zeta_2 \leq (1+(1+\frac{\Delta D}{v})(1+\sqrt{\lambda-1}))\sigma_n^3 \|\!| f |\!\|_{B(X_n,\sigma_n)}$$

Hence the expected result.

□

## 5.2 Proof of Lemma 3

Let $X^* \in B(X_n, \frac{\sigma_n}{2})$, let $H$ be the Hessian matrix of $f$ at point $X^*$ and $\lambda_{min,H}$ be its smallest eigenvalue. Let $\lambda_{min,H_n}$ be the smallest eigenvalue of the hessian of $H_n$. We want to prove that $X^* \in B(X_{n+1}, \frac{\sigma_{n+1}}{2})$.

Let $X_n^*$ defined at Eq. 1. From its definition we have that $X_n^* \in B(X_n, \sigma_n)$. A straightforward consequence (see Eq. 2) is that

$$X_{n+1} \in B(X_n, \sigma_n) \qquad (17)$$

Let $x \in (B(X_n,\sigma_n) \setminus B(X_n^*, \frac{1}{4}\sigma_{n+1}))$. Then

$$f(x) = f(X_n^*) + \underbrace{f(x) - H_n(x)}_{\geq -\epsilon_n} + \underbrace{H_n(x) - H_n(X_n^*)}_{\geq \frac{1}{2}\lambda_{min,H_n}\sigma_{n+1}^2/16} + \underbrace{H_n(X_n^*) - f(X_n^*)}_{\geq -\epsilon_n}$$

where $\epsilon_n = \sup_{x\in B_n}|H_n(x) - f(x)|$. Taking the min of the previous equation, we derive

$$\inf_{x\in B(X_n,\sigma_n)\setminus B(X_n^*,\frac{\sigma_{n+1}}{4})} f(x) \geq f(X_n^*) + \frac{1}{2}\lambda_{min,H_n}\frac{\sigma_{n+1}^2}{16} - 2\epsilon_n$$

From Lemma 2 , $\epsilon_n < K_\infty \sigma_n^3$ and from Lemma 4 , $\lambda_{min,H_n}$ is lower bounded, therefore for $\sigma_n$ small enough, we have $f(x) > f(X_n^*)$ for every $x \in B(X_n,\sigma_n)\setminus B(X_n^*, \frac{\sigma_{n+1}}{4})$. Moreover $f(X^*) \leq f(X_n^*)$ and $f(X_{n+1}) \leq f(X_n^*)$ (see Eq. 2), we deduce that

$$X^* \notin B(X_n,\sigma_n) \setminus B(X_n^*, \frac{\sigma_{n+1}}{4})$$

and $X_{n+1} \notin B(X_n,\sigma_n) \setminus B(X_n^*, \frac{\sigma_{n+1}}{4})$. As $X^* \in B(X_n,\sigma_n)$ (initial assumption) and $X_{n+1} \in B(X_n,\sigma_n)$ (see Eq. 17), we have that $X^* \in B(X_n^*, \frac{\sigma_{n+1}}{4})$ and $X_{n+1} \in B(X_n^*, \frac{\sigma_{n+1}}{4})$. Therefore the distance between $X_{n+1}$ and $X^*$ is smaller than $\frac{\sigma_{n+1}}{2}$, i.e. $X^* \in B(X_{n+1}, \frac{\sigma_{n+1}}{2})$.

□

**Lemma** 4 (A LOWER BOUND FOR $\lambda_{min,H_n}$).

$$\lambda_{min,H_n} \geq \frac{1}{4} min(Sp(H)) + O(\sigma_n)$$

*where $H$ denotes the Hessian matrix of $f$ and $min(Sp(H))$ its smallest eigenvalue.*

It remains to find a lower bound for $\lambda_{min,H_n}$, the smallest eigenvalue of the hessian of $H_n$. This will come from the fact that $f$ has a positive definite Hessian at $X^*$ and that $H_n$ is a quadratic approximation of $f$.

Let us show first that $X_n^* \in B(X_n, 3\sigma_n/4)$ and that this minimum is unique. Remember that $X_n^*$ is defined as the minimum of the quadratic approximation $H_n$ of $f$ built on points from $A_n$.
For all $x$ in $S(X_n, \frac{3}{4}\sigma_n)$ we have from Lemma 2

$$H_n(x) \geq f(x) - \epsilon_n$$

Besides from the definition of $\|\!| f |\!\|$

$$\begin{aligned} f(x) - \epsilon_n &\geq H(x-X^*) - \epsilon_n - \|\!| f |\!\|\sigma_n^3 \\ &\geq \frac{1}{2}\min\ Sp(H)\sigma_n^2/16 - \epsilon_n - \|\!| f |\!\|\sigma_n^3 \end{aligned}$$

But from Lemma 2 we also know that $|H_n(x^*)| \leq \epsilon_n$
Hence if we have

$$2\|\!| f |\!\|[1+(1+\sqrt{\lambda-1})(\Delta\sqrt{dim}/v+1)]\sigma_n^3 + \|\!| f |\!\|\sigma_n^3 < \frac{1}{2}w\sigma_n^2/16$$

*i.e.*

$$\left\{2\|\!| f |\!\|[1+(1+\sqrt{\lambda-1})(\Delta\sqrt{dim}/v+1)] + \|\!| f |\!\|\right\}\sigma_n < \frac{1}{2}w/16$$

then the minimum of $H_n$ will be inside $B(X_n, \frac{3}{4}\sigma_n)$.

Denote $H_n(x) = Q(x - X_n^*)$ sor some quadratic positive definite $Q$, and let $g = f - H_n$.
Let $x \in B(X_n,\sigma_n)$. We know that $g = O(\sigma_n^3)$ and that $f(x) = H(x-X^*) + O(\sigma_n^3)$, hence $g(x) = H(x-X^*) - Q(x-X_n^*) + O(\sigma_n^3)$.
From which we derive that $Q(x-X_n^*) \geq H(x-X^*) + O(\sigma_n^3)$.
Consider now the line that has direction $V_n$, the eigenvector corresponding to the smallest eigenvalue $\lambda_{min,H_n}$ of $H_n$ going through $X_n^*$. It meets the sphere $S(X_n,\sigma_n)$ in 2 points. Because $X^*$ lies in $B(X_n, \frac{\sigma_n}{2})$, those 2 points are at distance at least $\sigma_n/2$ from $X^*$. Moreover, at least one of them, say $x$, satisfies $||x - X_n^*|| \leq \sigma_n$. For this $x$ we can write $Q(x-X_n^*) \geq H(x-X^*) + O(\sigma_n^3)$, that is $\frac{1}{2}\lambda_{min,H_n}||x-X_n^*||^2 \geq \frac{1}{2}\min(Sp(H))||x-X^*||^2 + O(\sigma_n^3)$, and hence it shows that $\lambda_{min,H_n} \geq \frac{1}{4} min(Sp(H)) + O(\sigma_n)$.

□

# 6. DISCUSSION AND CONCLUSIONS

In this paper we built up a memetic algorithm and analyzed its convergence properties. The valuable property of the algorithm is that the order of convergence of the local algorithm is conserved when constructing the global algorithm.

The order 3/2 (with $\Theta(dim^2)$ evaluations at each generation) for the convergence of the local algorithm relies on the use of pseudo-random numbers instead of random numbers. Though losing a stochastic part for the algorithm we gain in terms of convergence properties. Therefore, the result suggests that this technique which is now very popular in other communities should also be used for evolutionary computation.

The technique used for this analysis is rather general: Section 2 establishes a general Lemma to prove that global convergence occurs almost surely with order $p$ when local

convergence occur with order $p$. We believe that it could be applied for analyzing other algorithm. The next step would be to consider more realistic scenario for the exploration part, in particular to have a better scaling in the dimension.

## 7. REFERENCES


[1] A. Auger. Convergence results for (1,$\lambda$)-SA-ES using the theory of $\varphi$-irreducible markov chains. *Theoretical Computer Science*, 334(1-3):35–69, 2005.

[2] H.-G. Beyer. *The Theory of Evolutions Strategies*. Springer, Heidelberg, 2001.

[3] R. Cerf. An asymptotic theory of genetic algorithms. In J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution*, volume 1063 of *LNCS*, pages 37–53. Springer Verlag, 1996.

[4] S. Droste, T. Jansen, , and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276:51–81, 2002.

[5] K. Fang and Y. Wang. *Number-Theoretic Methods in Statistics*. London: Chapman and Hall, 1994.

[6] O. François. An evolutionary strategy for global minimization and its markov chain analysis. *IEEE Transactions on Evolutionary Computation*, 2(3):77–91, 1999.

[7] B. Freisleben and P. Merz. New genetic local search operators for the TSP. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *PPSN96*, LNCS 1141, pages 890–899. Springer Verlag, 1996.

[8] J. Garnier, L. Kallel, and M. Schoenauer. Rigorous hitting times for binary mutations. *Evolutionary Computation*, 7(2):167–203, 1999.

[9] W. H. Hart. A Convergence Analysis of Unconstrained and Bound Constrained Evolutionary Pattern Search. *Evolutionary Computation Journal*, 9(1):1–23, 2001.

[10] M. Lozano, F. Herrera, N. Krasgonor, and D. Molina. Real-coded memetic algorithms with crossover hill-climbing. *Evolutionary Computation Journal*, 2004.

[11] P. Merz and B. Freisleben. A genetic local search approach for the QAP. In T. Bäck, editor, *Proceedings of the $7^{th}$ International Conference on Genetic Algorithms*, pages 465–470. Morgan Kaufmann, 1997.

[12] P. Moscato. On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms. Technical Report Caltech Concurrent Computation Program, Report. 826, California Institute of Technology, Pasadena, California, USA, 1989.

[13] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM, 1992.

[14] J. Poland. Explicit local models: Towards "optimal" optimization algorithms. Technical Report 09-04, IDSIA / USI-SUPSI, 2004.

[15] I. Rechenberg. *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien des Biologischen Evolution*. Fromman-Holzboog Verlag, Stuttgart, 1973.

[16] G. Rudolph. Convergence analysis of canonical genetic algorithm. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.

[17] G. Rudolph. Convergence of non-elitist strategies. In Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel, and H. Kitano, editors, *Proceedings of the First IEEE International Conference on Evolutionary Computation*, pages 63–66. IEEE Press, 1994.

[18] G. Rudolph. How mutation and selection solve long path problems in polynomial expected time. *Evolutionary Computation*, 4(2):195–205, Summer 1996.

[19] G. Rudolph. Convergence rates of evolutionary algorithms for a class of convex objective functions. *Control and Cybernetics*, 26(3):375–390, 1997.

[20] M. Schoenauer, editor. *Special Issue on Memetic Algorithms*. MIT press, 2004.

[21] H.-P. Schwefel. *Numerical Optimization of Computer Models*. John Wiley & Sons, New-York, 1981. 1995 – $2^{nd}$ edition.