Directed Evolutionary Algorithms

Dissertation

zur Erlangung des Grades eines

D o k t o r s   d e r   N a t u r w i s s e n s c h a f t e n

der Universität Dortmund
am Fachbereich Informatik
von

Stefan Berlik

Dortmund
2006

Tag der mündlichen Prüfung:
22.12.2006

Dekan:
Prof. Dr. Peter Buchholz

Gutachter:
Prof. Dr. Bernd Reusch
Prof. Dr. Günter Rudolph

Contents

Notation

Lists of Symbols

Spaces

$\mathbb{B}$	Space of binary numbers
$\mathcal{F}$	Fitness function space
$\mathcal{I}$	State space
$\mathbb{R}$	Space of real numbers
$\mathbb{R}^+$	Space of positive real numbers, $(0, \infty)$
$\mathbb{R}_0^+$	Space of positive real numbers including zero, $[0, \infty)$
$\mathcal{S}$	Strategy parameter space
$\mathcal{X}$	Object parameter space
$\mathbb{Z}$	Space of integer-valued numbers

Scalars, Vectors, and Matrices

$\mathbf{0}$	Vector or matrix containing zeros
$\mathbf{1}$	Vector or matrix containing ones
β	Learning rate of angle adaptation
κ	Life span of an individual
λ	Number of parents
μ	Number of offspring
ρ	Number of parents procreating a descendant
τ	Individual learning rate of n step size adaptation
τ'	Generational learning rate of n step size adaptation
τ_0	Learning rate of single step size adaptation
$\mathbf{C}$	Covariance matrix

$\mathbf{e}_i$	Unit vector in the ith dimension
g	Generation
i	Imaginary number
$\mathbf{I}$	Identity vector
$\mathbf{I}$	Individual
$\mathbf{P}^{(g)}$	Population at generation g
$\mathbf{s}$	Strategy parameter vector
$\mathbf{x}$	Object parameter vector

Relations, Operators, and Functions

$\sim$	Distributed according to
$\preceq$	Domination
$\nabla(\cdot)$	Vector differential operator del
$\nabla^2(\cdot)$	Hessian matrix
$\|\cdot\|$	L2 norm
$\partial(\cdot)$	Partial derivative
$\overset{+}{,}$	Selection operators
$\gamma_1(\cdot)$	Skewness
$\gamma_2(\cdot)$	Kurtosis
$\pi(\cdot)$	Skewing function
$\mathbf{A^{-1}}$	Matrix inverse
$\mathbf{A^T}$	Transposition
$\dim(\cdot)$	Dimension of a vector
$\mathrm{erf}(\cdot)$	Error function, also called Gauss error function
$\mathrm{erf}^{-1}(\cdot)$	Inverse error function
$\mathrm{E}(\cdot)$	Expectation
$f(\cdot)$	Skalar-valued fitness function
$\mathbf{f}(\cdot)$	Vector-valued fitness function
$\mathrm{H}(\cdot)$	Heaviside step function

$\inf(\cdot)$	Infimum
$\min(\cdot)$	Minimum
$\max(\cdot)$	Maximum
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with expectation μ and standard deviation σ
$\mathrm{OP}(\cdot)$	Outer product of a vector with itself
$\mathrm{sign}(\cdot)$	Signum function
$\mathcal{U}(a, b)$	Uniform distribution with support $[a, b]$
$\mathrm{Var}(\cdot)$	Variance
$\mathcal{SN}(\lambda)$	Skew-normal distribution with shape parameter λ
$\mathcal{SN}_\mathrm{n}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$	n-dimensional multivariate skew-normal distribution with mean $\boldsymbol{\mu}$, correlation matrix $\boldsymbol{\Omega}$, and shape vector $\boldsymbol{\alpha}$

List of Acronyms

BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm
cdf	Cumulative distribution function
CMA	Covariance Matrix Adaptation
CSA	Cumulative step size adaptation
DCMA	Directed Covariance Matrix Adaptation
DES	Directed Evolution Strategy
EA	Evolutionary Algorithm
EDA	Estimation of Distribution Algorithm
EP	Evolutionary Programming
ES	Evolution Strategy
EVOP	Evolutionary Operation
GA	Genetic Algorithm
MCDM	Multiple criteria decision making
MOEA	Multiobjective Evolutionary Algorithm
MOP	Multiobjective optimization problem

MSN	Multivariate skew-normal distribution
NSGA-II	Non-Dominated Sorting Genetic Algorithm II
NSN	Naïve skew-normal distribution
pdf	Probability density function
RPMI	Related polar moment of inertia
rv	Random variable
SA	Simulated Annealing
SN	Skew-normal distribution
$\mathrm{SN_n}$	n-dimensional multivariate skew-normal distribution
SNSN	Standardized naïve skew-normal distribution
SSN	Standardized skew-normal distribution

1 Introduction

In recent years, more and more frequently optimization problems are appearing defined by functions for which derivatives are unavailable or available only at a prohibitive cost. These problems originate from various disciplines, like for example science, engineering, or finance and cover a broad spectrum of problem classes. As two main reasons why derivative-free optimization is currently an area of impact can be seen the growing complexity in mathematical modeling and increasing sophistication of scientific computing.

A significant group of algorithms within the derivative-free optimization methods are Evolutionary Algorithms. Evolutionary Algorithms are inspired by biology and especially by those processes that allow populations of organisms to adapt to their surrounding environment, namely the principles of variation and selection. These concepts were established in the 19th century by Charles Darwin and introduced into the engineering context in the 1960's by Ingo Rechenberg, Hans-Paul Schwefel, John H. Holland, and others.

Natural evolution is driven by the principles of recombination and mutation of genetic information and fitness-based selection. In nature, the genetic information of the descendants is either a copy of the genes of a single parent or a mixture of gene sequences from the parents as the result of a mating process. Due to reproduction errors or other random perturbations, genetic information of the descendants also includes minor modifications. Based on their genetic habitude, individuals in a population differ in their fitness with respect to their environment. Those who are well adapted to their environment are likely to survive the natural selection process. They are more probable to become parents thus spreading their genetic information into the following generations.

In an engineering context, the genetic information is represented by the decision variables which specify the properties of a solution to the optimization problem. The fitness of the solution is determined by the objective function. Variation is introduced by recombination and mutation operators. The recombination operator exchanges information of different solutions while the mutation operator adds random perturbations to the variables.

It was the time when Dr. Hildebrand was working at his *Asymmetrische Mutation* (German, asymmetric mutation) that resulted 2001 in his dissertation *Asymmetrische Evolutionsstrategien* (German, asymmetric evolution strategies) [Hildebrand, 2001]. There he reported his mutation method to be somewhat superior compared to conventional symmetric mutation operators. Prof. Dr. Reusch[1] and me were working at that time together with

[1]Chair Automata- and Switching Theory and Computational Intelligence, Department of Computer Science, Universität Dortmund

Prof. Dr.-Ing. Kauder[2] and Dr. Helpertz of the mechanical engineering faculty at a DFG[3] founded project. Since this project aimed at finding optimal profiles of screw-type machines, there was the hope to benefit from the new mutation operator. However, as no gain could be recognized, further research into the area was necessary. The results of this work will be presented in what follows.

Since the early work of Rechenberg [1973] and Schwefel [1974] the design of mutation operators turned out to be one of the most critical points in Evolution Strategies. Even though till this day various mutation techniques have been proposed, almost all of them rely on symmetrically distributed mutations, usually normally distributed ones. As a consequence, no true directionality of the search can be modeled. The intention of directed mutation on the other hand is to impart exactly this. Put simply, for every problem dimension a tendency towards the positive or negative domain can be established by the mutation distribution. Thus, hopefully the mutation distribution will adapt favorable directions over the generations and sustain further advance into it. Hence, directed mutation introduces a different *mutation principle*.

The basis of any directed mutation operator is an appropriate customizable skew distribution. Usually introducing skewness into any distribution concurs with expectations unequal to zero. There are several methods to reach this goal whereof the two essential variants are the constructive approach and the skewing function approach.

The guiding principle of the constructive approach is to use a piecewise defined function whereof one half equals the normal distribution and the other half is scaled to some extent with respect to the abscissa. This construction principle traces back to Hildebrand [1996]. He used it to build several skew distributions for his asymmetric mutation. As these distributions have some serious drawbacks, with the naïve skew-normal distribution an alternative using the same construction principle has been developed.

However, there remain some construction principle immanent problems. The mentioned distributions are all of limited mathematical tractability and random variate generation by the proposed inversion method is expensive due to the relative complex functions.

All this will be mastered by a novel construction principle. The fundamental idea of the skewing function approach is to multiply an arbitrary symmetric probability density function with a skewing function. This operation amplifies one side of the density and attenuates the other at the same amount such that on average the density remains the same up to a constant factor. Loosely speaking, the two operations annihilate each other in total. Using this method skew distributions occur that are of striking simplicity and beauty.

This thesis is structured in four parts. Part I gives a survey on optimization. Optimization problems and algorithms are introduced and basic mathematical definitions are provided. Further, Evolution Strategies are recapitulated. In Part II Directed Evolutionary Algorithms are presented. Two approaches to

[2] Chair Fluid Energy Machines, Faculty of Mechanical Engineering, Universität Dortmund

[3] *Deutsche Forschungsgemeinschaft*, German Research Foundation

directed mutation are discussed and directed variants of the conventional Evolution Strategy and the Covariance Matrix Adaptation Evolution Strategy are presented. In Part III empirical results are subsumed. Part IV closes the thesis with conclusions and gives an outlook and suggestions for future work.

In the first chapter of Part I, i.e. Chapter 2, we specify what we will understand under the term *optimization*. We will see that different classes of optimization problems exist and various techniques to solve them. Several of the most widely used optimization algorithms are sketched and classified. Some theoretical fundamentals needed for the design of directed mutation operators are provided in Chapter 3. Besides others, these include random variables, cumulative distribution functions, probability density functions, and the moments of random variables. Chapter 4 deals with Evolution Strategies in detail. After the representation of the parameters within the strategies has been discussed, the main ingredients of general Evolution Strategies and different mechanisms of adaptation are outlined. Last, a state-of-the-art strategy with covariance matrix adaptation is presented.

Since Part II is about Directed Evolutionary Algorithms, in Chapter 5 the concept is motivated. The principle of directed mutation is sketched and some alternative approaches are rendered. The next chapter, Chapter 6, is dedicated to the biological foundations and questions that arise from these. In Chapter 7 the first and older constructive approach is presented in detail, which relies on piecewise defined distributions. With the Ξ distribution and the naïve skew-normal distribution two variants are discussed that differ only slightly in construction but vastly in operation. However, both suffer from some immanent drawbacks of this technique. Chapter 8 introduces the novel and more substantive skewing function approach that realizes the idea to multiply an arbitrary symmetric probability density function with a skewing function. With this the mentioned problems of the previous construction principle will be fixed. The construction principle, some skewing functions – whereof some are also distributions theirselves – and the class' most prominent member, the skew-normal distribution are presented thoroughly. In Chapter 9 the theoretical fundamentals of directed mutation are utilized. First the realization of directed mutation operators on basis of the previously presented distributions is discussed. With directed mutation operators being available, then Directed Evolution Strategies as a whole are introduced. The last chapter of this part is devoted to a very powerful directed variant of the Covariance Matrix Adaptation Evolution Strategy. Chapter 11, which is the first of the two chapters of Part III, presents simulation results of Directed Evolution Strategies applied to several single-objective test functions. In contrast to that, Chapter 12 is dedicated to simulation results achieved for a multiobjective real world application. To be more precise, the optimization of a screw-type machine is treated. Finally, Chapters 13 and 14 conclude this thesis and point to future work, respectively.

Part I

Survey of Optimization

2 Optimization Problems and Algorithms

Optimization is the activity that aims at finding the best, i.e. optimal solution to a given problem. To tackle the problem comprises the definition of the set of decision variables that are to be optimized plus one single or a set of objective functions that measures the quality of these decision variables. Furthermore, an appropriate optimization algorithm has to be selected.

This chapter introduces to the optimization problem in its single- and multi-objective variant. A classification of the main optimization approaches is given with some typical, most widely used optimization algorithms.

2.1 Optimization Problems

2.1.1 Definitions

The aim of every optimization process is to optimize a vector of some given objective functions $\mathbf{f}$ with respect to a set of decision variables $\mathbf{x}$. The optimization problem then reads:

$$\mathbf{f}(\mathbf{x}) \rightarrow \text{opt.}, \quad \mathbf{f} \in \mathcal{F}, \ \mathbf{x} \in \mathcal{X}. \tag{2.1}$$

In principle, the decision space $\mathcal{X}$ (also referred to as search space) can be any set of data structures of finite, not necessarily fixed length. The type of the components x_i of $\mathbf{x}$, and therefore the space $\mathcal{X}$ spanned by them, depends on the optimization problem. Examples for $\mathcal{X}$ are the n-dimensional real-valued search space $\mathbb{R}^n$, the integer search space $\mathbb{Z}^n$, and the binary search space $\mathbb{B}^n$; moreover mixtures of different spaces, as well as more complex data structures are possible. Due to constraints, $\mathcal{X}$ also can contain subspaces of these. In the sequel, we restrict ourselves to the real-valued decision space.

Without loss of generality, the optimization process can be restricted to the minimization of all objectives, since every maximization of a function f can be transformed into a corresponding minimization problem by

$$\max(f(\mathbf{x})) = -\min(-f(\mathbf{x})). \tag{2.2}$$

Depending on the number of objective functions given, we distinguish single-objective and multiobjective optimization. In the first case, the objective is scalar-valued; in the second, the objective is vector-valued.

2.1.2 Single-Objective Optimization

Single-objective optimization is the classical optimization task and simply called *optimization* if there is no likelihood of confusion. It is more intuitive in the sense that the objective function is scalar-valued and thus a total order on the search space $\mathcal{F}$ is induced by the well known "less than"-relation.

Definition 2.1. A single-objective optimization problem is defined as the search for a solution to the problem given by:

$$\min\ f(\mathbf{x}) \in \mathcal{F}, \quad \text{with } \mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathcal{X}, \tag{2.3}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is the n-dimensional decision space and $\mathcal{F} \subseteq \mathbb{R}$ is the one-dimensional objective space with a total order induced by the $<$-relation.

Definition 2.2. A point $\mathbf{x}^* \in \mathcal{X}$ is a *global optimum* of a function f, if

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall\, \mathbf{x} \in \mathcal{X}. \tag{2.4}$$

The corresponding objective value $f_{\min} := f(\mathbf{x}^*)$ is called *optimal value*.

Definition 2.3. A point $\mathbf{x}^*$ is a *local minimum* of a function f, if there exists some $\epsilon > 0$ such that

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall\, \mathbf{x} \text{ with } \|\mathbf{x} - \mathbf{x}^*\| < \epsilon. \tag{2.5}$$

Note that a norm is involved in this definition and hence also in the definition of a local minimum. Any global optimum is also a local optimum, but not vice versa.

2.1.3 Multiobjective Optimization

Most real-world problems require simultaneous optimization of more than one objective function. If these are conflicting, obviously a distinct configuration of the decision variables will not result in optimal values for all of them. Hence, some trade-off between the objectives is needed to ensure good compromise solutions.

Given a set of solutions to the problem, these compromise solutions can be identified by the principle of *dominance*. It states that a solution is definitely superior to another solution, i.e. dominating this solution, if it is superior or equal in all objectives and strictly superior for at least one objective. Applying this principle to the set of all solutions in order to remove solutions being dominated by at least one other, the subset of best compromise solutions results. This subset is termed *Pareto optimal set*, named after the work of the engineer and economist Vilfredo Pareto [1906].

All solutions within the Pareto optimal set are of equal quality and share the common feature that further improvement of any objective is possible only at the expense of at least one other objective. Further selection can be done on basis of preference information, known as Multiple Criteria Decision Making (MCDM).

Definition 2.4. A multiobjective optimization problem (MOP, hereafter) is defined as the search for solutions to the problem given by:

$$\min\ \mathbf{f}(\mathbf{x}) = \min\left(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\right)^T \in \mathcal{F}, \tag{2.6}$$

with

$$\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathcal{X},$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is the n-dimensional decision space and $\mathcal{F} \subseteq \mathbb{R}^m$ is the m-dimensional objective space.

To solve the optimization problem described by (2.6), a quality measure for comparing different solutions is needed. Such a partial ordering is given by the dominance criterion.

Definition 2.5 (Pareto dominance)**.** A vector $\mathbf{u} = (u_1, \ldots, u_k)^T$ is said to dominate a vector $\mathbf{v} = (v_1, \ldots, v_k)^T$, denoted by $\mathbf{u} \preceq \mathbf{v}$, if and only if $\mathbf{u}$ is partially less then $\mathbf{v}$, i.e.

$$\begin{aligned} \mathbf{u} \preceq \mathbf{v}, \quad \text{iff}\ \ \forall\, i \in \{1, \ldots, k\} : &\quad u_i \leq v_i \\ \wedge\, \exists\, i \in \{1, \ldots, k\} : &\quad u_i < v_i. \end{aligned} \tag{2.7}$$

If neither of two vectors is dominating the other one, the two are said to be indifferent.

Definition 2.6 (Pareto indifference)**.** Vector $\mathbf{u}$ is indifferent to vector $\mathbf{v}$, if and only if neither is dominating the other one.

With the domination principle the concepts of non-domination and Pareto optimality can be defined. Among a set of solutions $S \subset \mathcal{X}$, a non-dominated solution x is a solution that is not dominated by any member of the set S. When the set S is the entire search space $\mathcal{X}$, a non-dominated solution x is denoted as *Pareto optimal*.

Definition 2.7 (Non-dominance principle)**.** A solution $\mathbf{x} \in S$ with $S \subset \mathcal{X}$ is said to be *non-dominated with respect to the set S* if and only if there is no $\mathbf{y} \in S$ for which $\mathbf{v} = \mathbf{f}(\mathbf{y})$ dominates $\mathbf{u} = \mathbf{f}(\mathbf{x})$.

Definition 2.8 (Pareto optimality)**.** A solution $\mathbf{x} \in \mathcal{X}$ is said to be *Pareto optimal with respect to $\mathcal{X}$* if and only if there is no $\mathbf{y} \in \mathcal{X}$ for which $\mathbf{v} = \mathbf{f}(\mathbf{y})$ dominates $\mathbf{u} = \mathbf{f}(\mathbf{x})$.

The best solution to a multiobjective optimization problem is the subset of non-dominated solutions among all feasible solutions of the entire search space $\mathcal{X}$, called *Pareto optimal set*.

Definition 2.9 (Pareto optimal set)**.** For a given MOP $\mathbf{f}(\mathbf{x})$, the *Pareto optimal set* $\mathcal{P}^*$ is defined as:

$$\mathcal{P}^* := \{\mathbf{x} \in \mathcal{X} \mid \nexists\, \mathbf{y} \in \mathcal{X} :\ \mathbf{f}(\mathbf{y}) \preceq \mathbf{f}(\mathbf{x})\}\,. \tag{2.8}$$

The front spanned in the objective space $\mathcal{F}$ by the Pareto optimal set $\mathcal{P}^*$ is referred to as *Pareto front*.

Definition 2.10 (Pareto front). For a given MOP $\mathbf{f}(\mathbf{x})$ with the Pareto optimal set $\mathcal{P}^*$, the *Pareto front* $\mathcal{PF}^*$ is defined as:

$$\mathcal{PF}^* := \left\{ \mathbf{u} = \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^T \mid \mathbf{x} \in \mathcal{P}^* \right\}. \tag{2.9}$$

Depending on the relationship between the different objectives, the Pareto optimal set may contain one or several solutions. When no a priori preference is defined among the objectives, the dominance criterion is the only way to determine if one solution performs better than the other.

Example 2.1. The dominance principle for a two-objective minimization problem is illustrated in Figure 2.1. Solution a is dominating c, since a is superior in both objectives. a is indifferent to b, as b is to c; since in a mutual comparison, each solution is superior in one objective. The Pareto-optimal set is $\mathcal{P}^* = \{a, b\}$.

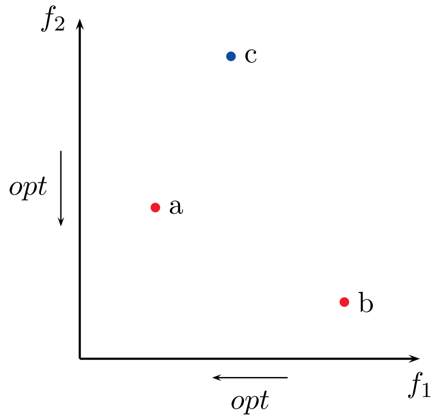


Figure 2.1: Illustration of the dominance principle for a two-objective minimization problem

2.2 Optimization Algorithms

2.2.1 Classification

The current literature (cf. e.g. [Goldberg, 1989]) identifies three main types of search methods, as illustrated in Figure 2.2: calculus based, enumerative, and guided random algorithms.

Calculus based algorithms rely on gradient information of the objective function. Direct and indirect methods are distinguished.

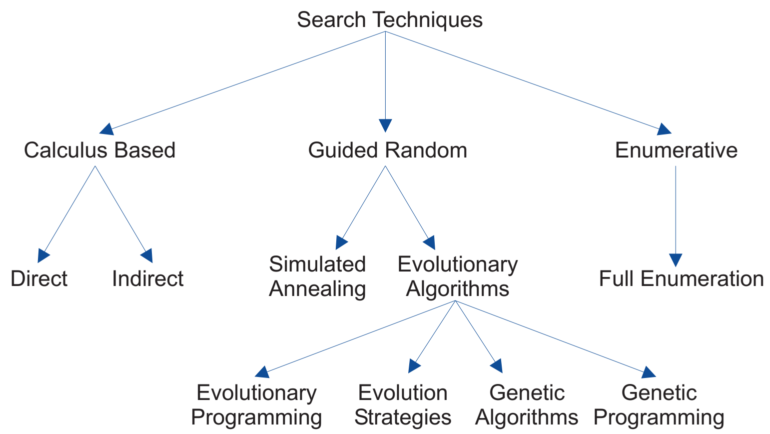


Figure 2.2: Classification of search and optimization algorithms

Guided random algorithms utilize random processes in trying to find the optimum. Many of them are inspired by nature where random processes occur e.g. in mutations, within the annealing process of metal, or in behaviors of bird flocks. Well-known representatives of guided random algorithms are Evolutionary Algorithms and Simulated Annealing.

Enumerative algorithms simply evaluate the objective function at every point in the search space. Full enumeration is exhaustive, but also most expensive in terms of the number of objective function evaluations. Other representatives of enumerative algorithms are branch and bound algorithms.

The three types are outlined in the following sections. Apart from the above classification, with hybrid methods another related approach is presented.

2.2.2 Calculus Based Algorithms

Calculus based algorithms utilize gradient or higher order derivative information of the objective function. Therefore they are sometimes also referred to as gradient methods. Depending on the way this information is exploited, direct and indirect approaches are distinguished. Direct methods work "directly" on the objective function and take advantage of the gradient information in an implicit manner whereas indirect methods apply techniques of classical analysis.

While there exists a broad consensus of which algorithms should be termed indirect there can be observed some inconsistency in the direct group. A rather strict classification is e.g. proposed by Wright [1995] in claiming that "A direct search method does not 'in its heart' develop an approximate gradient." In contrast to other authors, this excludes for instance finite difference Quasi-Newton schemes.

Direct Methods

Direct calculus based methods are iterative and most often deterministic algorithms. They uncover gradient information of the objective function only indirectly by reasoning on the search steps to execute. Kolda et al. [2003] give a recent overview about optimization by direct search methods. The probably most widely cited direct method is the Simplex Algorithm, to be presented next.

Simplex Algorithm The Simplex Algorithm proposed by Spendley, Hext, and Himsworth [1962] bases on the reiterated variation of the $n+1$ vertices of an object spanned in the n-dimensional search space. The start points are arranged equidistant. Hence, they are the vertices of a regular simplex. For $n = 2$, the simplex is a triangle, for $n = 3$, a tetrahedron and in general it is a polyhedron. Initially the objective values of the vertices have to be calculated. Then, at every iteration the worst vertex is replaced by its reflection on the midpoint of the other vertices, illustrated in Figure 2.3.

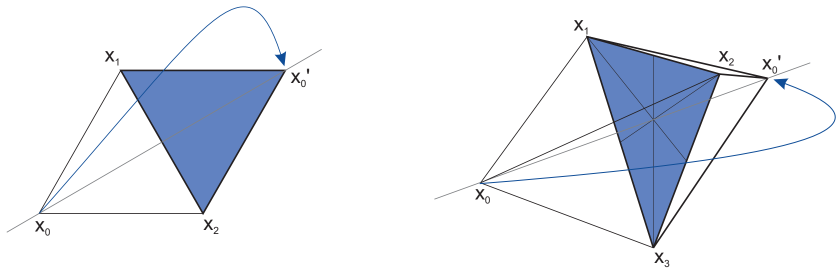


Figure 2.3: Reflection operator for 2-dimensional and 3-dimensional simplices

It may occur that the new vertex is again the one with the worst objective value which would in the next step anew be reflected to its origin. To avoid such oscillation, instead the second worse vertex is reflected, and so on. The consequence is that the entire simplex may begin to rotate around the vertex next to the optimum. Therefore, after some rotations the length of the edges is halved. It is proposed to do this shrinking whenever a vertex is included in more than $1.65n + 0.05n^2$ consecutive polyhedrons. The fact that the edge length can only be reduced limits the convergence velocity of the algorithm.

Nelder and Mead [1965] extended the method with a more flexible reflection operator and introduced an expansion and contraction operator. This enables the simplex algorithm to adapt to changing topologies and solves the collapsing size problem. Consequently, the algorithm operates with irregular simplices.

The simplex algorithm is widely used. However, Wright [1995] recently showed that it is slow for some problems. Even on convex functions the algorithm may fail.

Indirect Methods

Algorithms of the first group of indirect methods analytically compute positions of potential minima. This is done by differentiating the unconstrained objective function and solving the usually nonlinear set of equations resulting from setting the gradient of the objective function equal to zero. Thus the gradient and where required the Hessian matrix of the objective function are needed. The gradient ("first derivative") is the vector of the n partial derivatives of the objective function $f(\mathbf{x})$, denoted by $\nabla f(\mathbf{x})$ and defined as:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right). \tag{2.10}$$

The Hessian matrix ("second derivative") is defined by the n^2 partial derivatives of the n first derivatives with respect to the n decision variables. It is denoted by $\nabla^2 f(\mathbf{x})$ and defined as:

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix}. \tag{2.11}$$

A necessary condition for $\mathbf{x}^*$ being a local minimum is that the gradient $\nabla f(\mathbf{x}^*)$ vanishes. If additionally the Hessian matrix $\nabla^2 f(\mathbf{x}^*)$ becomes positive definite, the sufficient condition is fulfilled. This means, $\mathbf{x}^*$ is a local minimum if and only if

$$\begin{aligned} &\nabla f(\mathbf{x}^*) = \mathbf{0} \quad \text{and} \\ &\nabla^2 f(\mathbf{x}^*) \text{ is positive definite.} \end{aligned} \tag{2.12}$$

Hence this group of indirect gradient-based methods require the mathematical equations of the objective functions. While these are available for test functions, in real-world optimization problems this is usually not the case. Most often exact derivatives of the objective function cannot be calculated. Either the gradient simply does not exist, for example due to unpredictable discontinuities, or the function obviates the use of automatic differentiation techniques as it may be defined by a complex or convoluted computational structure.

This problem is circumvented by algorithms of the second group of indirect methods that utilize an iterative approach. The objective function is evaluated successively and search steps are preformed in the direction related to the local gradient. The simplest way to obtain the local gradient of the objective function f is finite differencing. The forward difference of the ith component of $g(\mathbf{x})$, i.e. the ith partial derivative of $f(\mathbf{x})$, is given by

$$g_i(\mathbf{x}) \approx \frac{f(\mathbf{x} + h\,\mathbf{e}_i) - f(\mathbf{x})}{h}, \tag{2.13}$$

where $\mathbf{e}_i$ is the unit vector of the ith space direction and h is the length of the finite step.

This gradient is the most obvious choice for the descent step direction, $\mathbf{p} = -g(\mathbf{x})$, as realized in the *steepest descent* method. That way the objective function value decreases at the fastest rate. However, the drawback with gradient descent is that moving into the descent direction is a *local* feature only.

A second order derivative approach is *Newton's method*. It is based on setting up a quadratic model of the objective function f in a neighborhood of its current position via the second order Taylor expansion of f around $\mathbf{x}_k$

$$f(\mathbf{x}_k + \mathbf{p}) \approx f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{p} + \frac{1}{2}\mathbf{p}^T \nabla^2 f(\mathbf{x}_k)\mathbf{p}, \tag{2.14}$$

where $\mathbf{p}$ is the step to the minimum. By differentiating (2.14) with respect to p, the minimum necessarily fulfills

$$\frac{\partial}{\partial \mathbf{p}} f(\mathbf{x}_k + \mathbf{p}) = \nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)\mathbf{p} = 0. \tag{2.15}$$

This can be reformulated to get the *Newton direction* p,

$$\mathbf{p} = -\frac{\nabla f(\mathbf{x}_k)}{\nabla^2 f(\mathbf{x}_k)}. \tag{2.16}$$

The Newton method is of second order convergence speed and considered to be fast. If $\nabla^2 f(\mathbf{x}_k)$ is a positive definite matrix, just one iteration step is required to find the minimum of the quadratic model from *any* starting point. However, good convergence can only be expected when the model (2.14) is accurate, otherwise the iterative Newton process might diverge.

A large number of varieties of Newton's method exist, all using the explicit computation of the gradient and the Hessian matrix. Since explicit computation of this matrix can be very expensive, if possible at all, this has to be seen as the main drawback with the Newton method.

By approximation of the Hessian, algorithms can circumvent that. The latter are also referred to as Quasi-Newton methods, often of same convergence speed as the original method. One example is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which was independently proposed by Broyden [1970], Fletcher [1970], Goldfarb [1970], and Shanno [1970]. It consists in calculating a matrix at every iteration by which the gradient vector is multiplied to take account of the additional knowledge gained during the iteration. With the Newton direction given by the vector p, a line search is executed that minimizes the objective function along the line $\mathbf{x}_k + \alpha\mathbf{p}, \alpha \in \mathbb{R}^+$, determined by the starting point and the search direction. An example for line search methods is the golden section method. Hence, from a more formal point of view, gradient descent methods can be characterized by successive minimizations in one-dimensional subspaces.

2.2.3 Guided Random Algorithms

Guided random algorithms, also referred to as semi-stochastic algorithms, represent one fraction of the group of random search algorithms, strictly random algorithms the other one. Examples for the latter are strictly random walks through the search space with saving the best known solution so far. A detailed comparison of guided random and related algorithms is e.g. given by

Berlik [2000]. Two prominent members of this group of algorithms are presented in the following: Evolutionary Algorithms and Simulated Annealing.

Evolutionary Algorithms

Evolutionary Algorithms (EAs) are a set of optimization algorithms inspired by biology and especially by those processes that allow populations of organisms to adapt to their surrounding environment, namely the *principles of variation and selection*. These concepts were established in the 19th century by Charles Darwin [1859] and are still today widely acknowledged as valid, even though complemented with further details [Futuyma, 1998, 2005], cf. also Chapter 6.

Natural evolution is driven by the principles of recombination and mutation of genetic information and fitness-based selection. In nature, the genetic information of the descendants is either a copy of the genes of a single parent or a mixture of gene sequences from the parents as the result of a mating process (*recombination*). Due to reproduction errors or other random perturbations, genetic information of the descendants also includes minor modifications (*mutations*). Based on their genetic habitude, individuals in a population differ in their fitness with respect to the environment. Those who are well adapted to the environment are likely to survive the natural selection process (*selection*). These individuals can become parents and spread their genetic information into the following generations.

In an engineering context, the genetic information is represented by the decision variables which specify the properties of a solution to the optimization problem. The fitness of the solution is determined by the objective function.

EAs are optimization algorithms of *zeroth order*, i.e. no gradient information is required. They belong to the class of stochastic optimization algorithms and are considered as robust [Bäck et al., 1997b] since they are applicable to a wide variety of problems. EAs can tackle problems with arbitrary combinations of the following features: static or dynamic problems; discrete or continuous decision variables; discontinuous, noisy, or multi-modal objective functions; one single or several objective functions [Schwefel, 1995]. However, EAs are also considered computationally expensive in terms of the number of objective function evaluations required for convergence.

EA is a generic term, coined in the year 1991. If only the optimization methods are considered, it comprises three main branches:

- Evolution Strategies,
- Genetic Algorithms,
- Evolutionary Programming.

All have independently been developed in the 1960's and will be outlined in the following paragraphs. For a more detailed presentation of each of the three groups of EAs see the monographs by Rechenberg [1994] or Schwefel [1995] for Evolution Strategies, Holland [1994] or Goldberg [1989] for Genetic Algorithms, and Koza [1992] for Evolutionary Programming. In addition a paper by Bäck et al. [1993], Bäck and Schwefel [1996], or a book by Fogel [1999] can be consulted for a unified presentation and comparison between them.

Evolution Strategies (ESs) were a joint development of a group of three students, Bienert, Rechenberg, and Schwefel, by the mid-1960's in Berlin [Bienert et al., 1966, Rechenberg, 1964, Schwefel, 1965]. Originally, they were used as experimental optimization techniques [Rechenberg, 1964, Klockgether and Schwefel, 1970], e.g. with hydrodynamical problems like driving a flexible pipe bending into a form with minimal loss of energy [Lichtfuss, 1965]. These early strategies relied on two individuals per generation only, one parent and one descendant. Like in Evolutionary Operation (EVOP) [Box, 1957, Box and Draper, 1969] the variables were altered in discrete steps, but stochastically instead of deterministically. With the appearance of the first computers the ES developed toward a multi-membered numerical optimization strategy based on real-valued vector representations.

In ESs, mutation is performed by adding normally distributed random numbers with zero mean and a certain standard deviation to the decision variables. In classical ESs, mutation was the solely variation operator and is even now considered the main operator. The principle of *self adaptation* is of core importance. It is used to steer the control parameters of the mutation operator, e.g. a global standard deviation, individual standard deviations or correlation information. In ESs, the parents of a descendant are chosen at random and selection is done deterministically; only the best individuals survive.

ESs for single-objective problems will be discussed in detail in Chapter 9 and for a multiobjective problem in Chapter 12. A more detailed review of the field's history is given e.g. by Bäck et al. [1991, 1997b].

Genetic Algorithms (GAs) were proposed by Holland [1962, 1969, 1973]. In its most canonical form, individuals are represented by bit strings of fixed length. Other than binary decision spaces have to be mapped to a binary representation and vice versa. Mutation and recombination is performed by flipping bits or exchanging substrings of different parents, respectively. The recombination operator is considered the key operator, justified by the *schema theorem*. Mutations are seen as less important and occur rather seldom. In GAs, selection is stochastic; proportional to an individual's fitness.

Evolutionary Programming (EP) was introduced by Fogel [1962, 1964] to simulate evolution as a learning process aiming at generating intelligent behavior. While its original form was proposed to operate on finite state machines and the corresponding discrete representations, currently individuals are represented in EP similar to ESs as real-valued vectors.

Simulated Annealing

Simulated Annealing (SA) is inspired by the annealing process in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. In the beginning of the annealing process, atoms are likely to change between different crystal structures and their internal energy might decrease, or due to the external energy of the heat even increase. At lower temperatures, merely small changes occur; especially no changes that would increase the internal energy. Hence, always the lowest energies are kept.

Kirkpatrick et al. [1983] proposed with SA a transformation of the annealing process in metallurgy to optimization. The algorithm works with two solutions $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. Starting with an arbitrary solution $\mathbf{x}$, another solution $\mathbf{x}'$ is created in the neighborhood of $\mathbf{x}$ by some random process. If the objective value of the new solution $f(\mathbf{x}')$ is equal or better than that of the old solution $f(\mathbf{x})$, the algorithm continues with $\mathbf{x}'$. But even if the objective value of the new solution is worse, it is not automatically rejected. While in the real annealing process atoms may change from lower to higher energy levels, in SA the worse of two solutions may be selected. The selection of worse solutions may help to overcome local minima and is done with probability $P(\Delta, T)$. Δ measures the resulting decrease in the objective value, $\Delta = f(\mathbf{x}) - f(\mathbf{x}')$, and T is a control parameter (analog of temperature). Similar to the real annealing, the parameter T decreases during the optimization. In the beginning, a stronger increase in the objective value is accepted, whereas at the end only better solutions are. One possible function for the calculation of the selection probability is given by

$$P(\Delta, T) = e^{-\Delta/T}. \tag{2.17}$$

2.2.4 Enumerative Algorithms

Enumerative methods use the simple technique of checking the objective function value of every point in the search space. Therefore, they are usually applied to finite search spaces, or at least discretized infinite search spaces. Since they are *exhaustive*, they are guaranteed to find the optimum. However, enumerative methods are applicable only to search spaces of limited size which makes them unsuitable for many real world problem domains.

Hybrid Methods

Hybrid methods fusion different optimization algorithms aiming at improving its overall performance. Utilizing the power of local search inside EAs has been conceived since the very beginning of EAs. Usually the EA adopts the part of global optimization and starts a local search algorithm from time to time to explore the neighborhood of the best solution found so far.

Hybrid methods are a lively field of research. An overview is given in [Bäck et al., 1997a, Part D3], for a recent work in the field of Covariance Matrix Adaptation-ES see [Auger et al., 2004].

3 Basic Definitions

This chapter provides the foundation of probability theory needed for the design of directed mutation operators. Therefore issues like random variables, cumulative distribution functions, probability density functions, and the moments of random variables are treated. Especially the higher moments, skewness and kurtosis are presented. Since the aim is to lay a solid foundation, also definitions of the σ-algebra, probability spaces and measures are given. However, this chapter does not claim to be complete. To get a profound introduction to probability theory, the reader is referred to the standard monographs. Good sources of information are for example the ones of Chow and Teicher [1997], Karr [1993], Rasch [1995], or Ross [2000], where the presented matter is compiled from.

3.1 Sets and Set Operations

A set is a collection of well-defined and well-distinguished objects considered as a single whole. The objects are also called elements and the set is also said to be the aggregate of these elements. In a set the order of its elements has no significance and multiplicity is generally also ignored. We think of the set Ω as a universal set if every set we consider is a subset of Ω. A set containing no elements is called an empty set, denoted by $\emptyset$, while a set whose elements are themselves sets is called a class.

To indicate that ω is an element of the set Ω we write $\omega \in \Omega$ and $\omega \notin \Omega$ otherwise. If A and B are sets and every element of A is likewise an element of B, then A is called a subset of B, denoted by $A \subset B$. An axiom of set theory called extensionality says that two sets are equal if and only if they contain exactly the same elements. Then both $A \subset B$ and $B \subset A$ and we write $A = B$.

Definition 3.1. Let Ω be a universal set.

1. The *union* of the sets A_i; $i = 1, \ldots, n$ is the set

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ for at least one } i\}.$$

Similarly, the union of the infinite sequence of sets A_i; $i = 1, \ldots$ is the set

$$A_1 \cup A_2 \cup \cdots = \bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ for at least one } i\}.$$

2. The *intersection* of the sets A_i; $i = 1, \ldots, n$ is the set

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ for every } i\}.$$

   Similarly, the intersection of the infinite sequence of sets A_i; $i = 1, \ldots$ is the set

$$A_1 \cap A_2 \cap \cdots = \bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ for every } i\}.$$

   Notice that two sets are disjoint if and only if $A \cap B = \emptyset$.

3. The *difference* $A \setminus B$ of two sets A and B is defined by

$$A \setminus B = \{\omega \in \Omega \mid \omega \in A,\ \omega \notin B\}.$$

   Notice the definition of the *complement*

$$A^c = \{\omega \in \Omega \mid \omega \notin A\}$$

   as special case of the difference.

4. The *symmetric difference* $A \mathbin{\Delta} B$ of two sets A and B is defined by

$$A \mathbin{\Delta} B = \{\omega \in \Omega \mid (\omega \in A) \text{ XOR } (\omega \in B)\}.$$

Union, intersection, difference, and symmetric difference are referred to as set operations.

Definition 3.2. The collection $A_1, A_2, \ldots$ of sets is *mutually* (or *pairwise*) *disjoint* if $A_i \cap A_j = \emptyset$ for every $i \neq j$. Then $\bigcup_{i=1}^{n} A_i$ sometimes is written as $\sum_{i=1}^{n} A_i$, and similarly for infinite sums.

Lemma 3.1. *For any sets $A, B, C \in \Omega$ the following identities hold:*

1. Associative laws

$$A \cup (B \cup C) = (A \cup B) \cup C, \qquad A \cap (B \cap C) = (A \cap B) \cap C$$

2. Commutative laws

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A$$

3. Distributive laws

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

4. Idempotent laws *and* complement laws

$$A \cup A = A, \qquad A \cap A = A$$
$$A \cup A^c = \Omega, \qquad A \cap A^c = \emptyset$$

5. Identity laws *and* domination laws

$$A \cup \emptyset = A, \qquad A \cap \Omega = A$$
$$A \cup \Omega = \Omega, \qquad A \cap \emptyset = \emptyset$$

6. De Morgan's laws

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c$$

Proof. We only proof the first part of 1 here. The other identities are derived similarly.

$$\begin{aligned} A \cup (B \cup C) &= \{\omega \in \Omega \mid \omega \in A \vee (\omega \in B \cup C)\} \\ &= \{\omega \in \Omega \mid \omega \in A \vee (\omega \in B \vee \omega \in C)\} \\ &= \{\omega \in \Omega \mid (\omega \in A \vee \omega \in B) \vee \omega \in C\} \\ &= \{\omega \in \Omega \mid (\omega \in A \cup B) \vee \omega \in C)\} \\ &= (A \cup B) \cup C \end{aligned}$$

□

3.2 σ-Algebras and Measurable Spaces

When we think of a *non-deterministic experiment* we intuitively think of it as an experiment whose outcome is uncertain. The set Ω of all conceivable outcomes associated with such an experiment is referred to as the *sample space*. Often it is not appropriate to consider the universal set Ω with all possible subsets on the whole. Instead, only the collection of sets of interest is to be treated.

3.2.1 Algebras

Definition 3.3. Let Ω be a universal set. An *algebra* $\mathcal{A}$ is a nonempty collection of subsets of Ω, satisfying the axioms:

(A1) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(A2) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$

Example 3.1. Let $\Omega = \{a, b, c, d\}$. Then $\mathcal{A}_1 = \{\emptyset, \{a, b, c\}, \{d\}, \Omega\}$ is an algebra while $\mathcal{A}_2 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ is not.

The sets of an algebra $\mathcal{A}$ are called *events*. Mutually disjoint sets in $\mathcal{A}$ are called *mutually exclusive events*. Thus an algebra $\mathcal{A}$ can be thought of as a collection of events associated with an experiment whose sample space is Ω. It follows that algebras are closed under unions, intersections, and differences.

Lemma 3.2. *Let $\mathcal{A}$ be an algebra and $A_1, A_2, \dots, A_n \in \mathcal{A}$ with $i, j \in \{1, \dots, n\}$. Then the following holds:*

1. $\emptyset, \Omega \in \mathcal{A}$
2. $A_i \cap A_j \in \mathcal{A}$
3. $A_i \setminus A_j \in \mathcal{A}$
4. $\bigcup_{k=1}^{n} A_k \in \mathcal{A}, \bigcap_{k=1}^{n} A_k \in \mathcal{A}$

Proof.

1. Because $\mathcal{A}$ is not empty, an $A \in \mathcal{A}$ exists, with $A \subset \Omega$. It then follows from (A1) that $A^c \in \mathcal{A}$, from (A2) that $A \cup A^c = \Omega \in \mathcal{A}$, and last from (A1) that $\Omega^c = \emptyset \in \mathcal{A}$.
2. $A_i \cap A_j = \left(A_i^c \cup A_j^c\right)^c \in \mathcal{A}$
3. $A_i \setminus A_j = A_i \cap A_j^c = (A_i^c \cup A_j)^c \in \mathcal{A}$
4. By induction.

□

3.2.2 σ-Algebras and Measurable Spaces

To investigate limits it is necessary that the system of sets $\mathcal{A}$ is not only closed under unions and intersections of *finite* many sets but also under unions and intersections of *countably infinite* many sets. This can be reached by adding the following axiom.

Definition 3.4. A *σ-algebra* is an algebra satisfying the additional axiom

(A3) $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definition 3.5. The pair $(\Omega, \mathcal{A})$ is called a *measurable space* if $\mathcal{A}$ is a σ-algebra relative to Ω.

For every Ω the power set $\mathcal{P}$ is always also a σ-algebra. If Ω is finite or countably infinite, $\mathcal{A} = \mathcal{P}$ can be chosen. If Ω is uncountable (e.g. $\Omega = \mathbb{R}$ or $\Omega = [0, 1]$), a smaller σ-algebra has to be chosen.

3.3 Probability Spaces and Measures

Definition 3.6 (Kolmogorov Axioms of Probability). Let $(\Omega, \mathcal{A})$ be a measurable space. A *probability function* (or *probability measure*) *on* Ω is a function $P : \mathcal{A} \to [0, 1]$ such that

(P1) P is non-negative: $P(A) \geq 0$ for each $A \in \mathcal{A}$

(P2) P is normed: $P(\Omega) = 1$

(P3) P is σ-additive: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for pairwise disjoint $A_1, A_2, \ldots \in \mathcal{A}$

$P(A)$ is termed *probability of the event* $A \in \mathcal{A}$. Let $(\Omega, \mathcal{A})$ be a measurable space and P be a measure on $\mathcal{A}$. Then the tripel $(\Omega, \mathcal{A}, P)$ is called a *measure space*. If $P(\Omega) = 1$, then it is called a *probability space*.

Theorem 3.1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and the event $A \in \mathcal{A}$, then*

1. $P(\emptyset) = 0$
2. $P(A) \leq 1$
3. $P(A^c) = 1 - P(A)$

Proof.

1. Since Ω and $\emptyset$ are pairwise disjoint, by axiom (P3) then $P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$. Also, $P(\Omega) = 1$ by (P1), hence $1 = 1 + P(\emptyset)$, so that $P(\emptyset) = 0$.
2. $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ by (P2) and (P3). Since $P(A)$ and $P(A^c)$ are non-negative by P(1), then $P(A) \leq 1$.
3. This follows immediately from $1 = P(\Omega) = P(A) + P(A^c)$.

□

Theorem 3.2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $A, B \in \mathcal{A}$ be two events, then*

1. $A \subset B \Rightarrow P(A) \leq P(B)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. $P(A \cup B) \leq P(A) + P(B)$

Proof.

1. $P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$. Since $P(B \setminus A) \geq 0$ by (P1), then $P(B) \geq P(A)$.
2. Since $P(A \setminus B) = P(A) - P(B)$ for $A, B \in \mathcal{A}$ with $A \supset B$ and

$$A \cup B = (A \setminus (A \cap B)) \cup (A \cap B) \cup (B \setminus (A \cap B))$$

it holds

$$\begin{aligned} P(A \cup B) &= P(A \setminus (A \cap B)) + P(A \cap B) + P(B \setminus (A \cap B)) \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

3. This follows immediately from 2.

□

With Theorem 3.2 immediately further properties of probability measures follow:

1. P is finitely additive:

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{3.1}$$

   for every finite sequence $\{A_i\}_{i=1}^{n}$ of mutually disjoint events in $\mathcal{A}$.

2. Subadditivity of P:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) \tag{3.2}$$

   for every collection $\{A_i\}$ of events in $\mathcal{A}$.

As a generalization of the second part of Theorem 3.2, also Poincaré-Sylvester's inclusion-exclusion formula arises.

Corollary 3.1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. For every $n = 1, 2, \dots$ and every sequence $A_1, \dots, A_n \in \mathcal{A}$ holds*

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \leq k_1 < \ldots < k_i \leq n} P(A_{k_1} \cap \ldots \cap A_{k_i})\,. \tag{3.3}$$

Proof. By induction.

Further we can show with Theorem 3.2 that probability measures are continuous with respect to the monotone convergence of sets.

Corollary 3.2. *Let $A_1, A_2, \dots \in \mathcal{A}$ be events as defined above. Then it holds*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} P(A_i), \quad \text{if } A_1 \subset A_2 \subset \dots, \tag{3.4}$$

and

$$P\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} P(A_i), \quad \text{if } A_1 \supset A_2 \supset \dots. \tag{3.5}$$

Proof. Omitted.

The subadditivity of probability measures is not only valid for two events as discussed in part 3 of Theorem 3.2 or for finite many events as shown in (3.2). It is also valid for infinite collections.

Theorem 3.3. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\{A_i\}_{i=1}^{\infty}$ be a sequence of events from $\mathcal{A}$. Then it holds*

$$P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i). \tag{3.6}$$

Proof. Instead of the sequence $\{A_i\}_{i=1}^{\infty}$, define the sequence $\{A_i'\}_{i=1}^{\infty}$ as

$$A_1' = A_1, \qquad A_i' = A_i \setminus \bigcup_{k=1}^{i-1} A_k.$$

Now $A_i' \cap A_j' = \emptyset$ for $i < j$ so $\{A_i'\}_{i=1}^{\infty}$ is a sequence of disjoint sets. Since $\bigcup_{i=1}^{\infty} A_i' = \bigcup_{i=1}^{\infty} A_i$, and since $A_i' \subset A_i$, it holds

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} A_i') = \sum_{i=1}^{\infty} P(A_i') \leq \sum_{i=1}^{\infty} P(A_i),$$

and the claim follows. □

3.4 Random Variables

In Section 3.3 we studied properties of a set function P defined on a measurable space $(\Omega, \mathcal{A})$. Since P is a set function, it is not very easy to handle. Moreover, in practice we are usually more interested in some function $X(\omega)$ of elementary events ω.

Definition 3.7. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A *random variable* (rv) is a mapping $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \mid \omega \in \Omega,\ X(\omega) \leq x\} \in \mathcal{A}, \quad \forall\, x \in \mathbb{R}\,. \tag{3.7}$$

The regularity condition (3.7) is called *measurability* of the mapping X with respect to the σ-algebra $\mathcal{A}$. Often we are not only interested in the probability that the values $X(\omega)$ of a rv X do not exceed a given threshold x, i.e. take values in the interval $B = (-\infty, x]$. Rather we are interested in the probability that X takes values in a more general subset $B \subset \mathbb{R}$, where B might be the union of disjoint intervals. Therefore not only in the sample space, but also in the event space a system of subsets is considered that is closed under the set operations $\cup, \cap$, and $\setminus$. Usually the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is considered, that is defined as the minimal σ-algebra of subsets of $\mathbb{R}$ containing all open sets (a, b), with $-\infty < a < b < \infty$. Hence, $\mathcal{B}(\mathbb{R}) = \sigma\left(\{(a,b),\ -\infty < a < b < \infty\}\right)$ is a generating system. In particular, $\mathcal{B}(\mathbb{R})$ also contains all half-open resp. closed intervals, since e.g. $(a, b] = \bigcap_{n=1}^{\infty}(a, b + n^{-1}) \in \mathcal{B}(\mathbb{R})$ holds.

Thus, we can give an equivalent to the regularity condition (3.7) as follows:

Theorem 3.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A mapping $X : \Omega \to \mathbb{R}$ is a rv if*

$$\{\omega \mid \omega \in \Omega,\ X(\omega) \in B\} \in \mathcal{A}, \quad \forall\, B \in \mathcal{B}(\mathbb{R}) \tag{3.8}$$

where B is the Borel σ-algebra over $\mathbb{R}$.

Proof. Omitted.

Example 3.2. When rolling two dice, the sample space is given by $\Omega = \{\omega = (\omega_1; \omega_2),\ \omega_i \in \{1, \ldots, 6\}\}$. If we are interested in the sum of the two dice, then

$\omega \to X(\omega) = \omega_1 + \omega_2$. The event that we get a sum of 4, for example, is given by $A = \{\omega \mid X(\omega) = 4\} = \{(1,3),(2,2),(3,1)\}$, or in general $A = \{\omega \mid X(\omega) = k\}$ with $k \in \{2,\ldots,12\}$. In question is the probability $P(A)$, thus it is necessary that $A \in \mathcal{A}$. It has to hold $\{\omega \mid \omega \in \Omega,\ X(\omega) = k\} \in \mathcal{A}$ for every $k = 2,\ldots,12$. In this example this is equivalent to $\{\omega \mid \omega \in \Omega, X(\omega) \leq x\} \in \mathcal{A}$ for every $x \in \mathbb{R}$.

3.5 Distributions, Cumulative Distribution Functions, and Probability Density Functions

Theorem 3.4 leads to the definition of the *distribution* and the *cumulative distribution function* of a rv X.

3.5.1 Distributions

Definition 3.8. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ an arbitrary rv. The *distribution* of X is the set function $P_X : \mathcal{B}(\mathbb{R}) \to [0,1]$ with

$$P_X(B) = P\left(\{\omega \mid \omega \in \Omega,\ X(\omega) \in B\}\right), \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{3.9}$$

The set function defined in (3.9) is a *probability measure* over the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, since P_X is σ-additive and normed due to $P_X(\mathbb{R}) = P(\Omega) = 1$. Usually, (3.9) is written in its abbreviated form,

$$P(X \in B) = P\left(\{\omega \mid \omega \in \Omega,\ X(\omega) \in B\}\right), \quad \forall B \in \mathcal{B}(\mathbb{R})$$

and especially

$$P(X \leq x) = P\left(\{\omega \mid \omega \in \Omega,\ X(\omega) \leq x\}\right), \quad \forall x \in \mathbb{R}.$$

3.5.2 Cumulative Distribution Functions

With the definition of the distribution (3.9), the *cumulative distribution function* of a rv X can be defined.

Definition 3.9. The function $F_X : \mathbb{R} \to [0,1]$ with $F_X(x) = P(X \leq x)$ is called *cumulative distribution function* (cdf) of the rv X.

Next, some properties of cdfs are discussed.

Theorem 3.5. *Let $X : \Omega \to \mathbb{R}$ be an arbitrary rv and $F_X : \mathbb{R} \to [0,1]$ its cdf. Then it holds*

1. *Asymptotic behavior at infinity:*

$$F_X(-\infty) := \lim_{x\to-\infty} F_X(x) = 0, \qquad F_X(\infty) := \lim_{x\to\infty} F_X(x) = 1, \tag{3.10}$$

2. *Monotony:*

$$F_X(x) \leq F_X(x+h), \quad \forall x \in \mathbb{R} \text{ and } h \geq 0, \tag{3.11}$$

3. *Continuity from the right: $F_X(x)$ is continuous from the right, i.e. for every sequence $\{h_n\}$ with $h_n \geq 0$ and $\lim_{n\to\infty} h_n = 0$ it holds*

$$\lim_{n\to\infty} F_X(x + h_n) = F_X(x), \quad \forall x \in \mathbb{R}. \tag{3.12}$$

Proof.

1. Only the first part of (3.10) is shown. Since F_X is monotone we can w.l.o.g. assume that x converges monotone against $-\infty$. With Corollary 3.2 then follows

$$\lim_{x\to-\infty} F_X(x) = \lim_{x\to-\infty} P_X((-\infty, x]) = P_X\Big(\bigcap_{x\leq 0}(-\infty, x]\Big) = P_X(\emptyset) = 0.$$

The proof of the second part of (3.10) is analog.

2. Since $(-\infty, x] \subset (-\infty, x+h]$, it follows from the first part of Theorem 3.2 that

$$F_X(x) = P_X((-\infty, x]) \leq P_X((-\infty, x+h]) = F_X(x+h).$$

3. Analog to the proof of the first part, from Corollary 3.2 follows that

$$\begin{aligned}
\lim_{n\to\infty} F_X(x+h_n) &= \lim_{n\to\infty} P_X((-\infty, x+h_n]) \\
&= P_X\Big(\bigcap_{n\geq 1}(-\infty, x+h_n]\Big) \\
&= P_X((-\infty, x]) \\
&= F_X(x).
\end{aligned}$$

□

With the distribution function F_X also the following probabilities can be described

$$P(a \leq X \leq b), \quad P(a < X \leq b), \quad P(a < X < b), \quad P(a \leq X < b),$$

because for example

$$\begin{aligned}
P(a \leq X \leq b) &= P(\{X \leq b\} \setminus \{X < a\}) \\
&= P(X \leq b) - P(X < a) \\
&= F_X(b) - \lim_{h\downarrow 0} F_X(a-h).
\end{aligned}$$

However, in general $F_X(a) = \lim_{h\downarrow 0} F_X(a-h)$ does not hold, but

$$F_X(a) = \lim_{h\downarrow 0} F_X(a-h) + P(X = a). \tag{3.13}$$

In Theorem 3.5 it is shown that distribution functions are monotone and bounded. Hence, they can have for every $\varepsilon > 0$ only countable many jump discontinuities with jumps higher then ε; and thus only countable many jump discontinuities can exist.

Theorem 3.6. *Let $X : \Omega \to \mathbb{R}$ be an arbitrarily rv. Then the distribution P_X of X is uniquely determined by the cdf F_X of X.*

Proof. Omitted.

3.5.3 Probability Density Functions

Definition 3.10. The rv $X : \Omega \to \mathbb{R}$ (resp. its distribution) is called *absolutely continuous*, if the cdf F_X of X has the following integral representation

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy, \quad \forall\, x \in \mathbb{R} \tag{3.14}$$

where $f_X : \mathbb{R} \to [0, \infty)$ is a non-negative (Lebesgue-integrable) function, referred to as *probability density function* (pdf) or *density* of X.

The integral in (3.14) is usually regarded as Lebesgue integral.

The cdf F_X (and thus also the distribution P_X) of a absolutely continuous rv X is in the following sense completely determined by a pdf f_X.

Theorem 3.7.

1. *The rv $X : \Omega \to \mathbb{R}$ is absolutely continuous, if the distribution P_X of X can be written in the following form:*

$$P_X(B) = \int_{B} f_X(y)\,dy, \quad \forall\, B \in \mathcal{B}(\mathbb{R}) \tag{3.15}$$

2. *Let $X, Y : \Omega \to \mathbb{R}$ be absolutely continuous rvs. Then $P_X = P_Y$, if and only if*

$$f_X(x) = f_Y(x) \tag{3.16}$$

for almost all $x \in \mathbb{R}$, i.e. (3.16) holds for all $x \in \mathbb{R} \setminus B$, where the set of exceptions $B \subset \mathbb{R}$ has Lebesgue measure zero.

Proof. Omitted.

Often the density f_X is an (at least piecewise) absolutely continuous function. If X is absolutely continuous, then the cdf F_X has no jumps. With (3.13) follows especially

$$P(X = x) = 0, \quad \forall\, x \in \mathbb{R}. \tag{3.17}$$

To describe an absolutely continuous rv X it is sufficient to regard its pdf f_X, since f_X uniquely determines the cdf F_X and with this also the distribution P_X of X. Some examples of pdfs are given in the next definition.

Definition 3.11. Let X be a continuous rv. It is said to possess

(a) *uniform distribution*, $\mathcal{U}(a,b)$, with support (a,b) if the pdf is

$$f_X(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases} \quad \forall\, x \in \mathbb{R}, \tag{3.18}$$

(b) *normal distribution*, $\mathcal{N}(\mu,\sigma^2)$, with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if the pdf is

$$f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \forall\, x \in \mathbb{R}, \tag{3.19}$$

(c) *lognormal* or *logarithmic normal distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if the pdf is

$$f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma\, x} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right), \quad \forall\, x \in \mathbb{R}^+, \tag{3.20}$$

(d) *exponential distribution* with parameter $\lambda > 0$ if the pdf is

$$f_X(x;\lambda) = \lambda e^{-\lambda x}, \quad \forall\, x \in \mathbb{R}_0^+, \tag{3.21}$$

(e) *Cauchy distribution* with parameters $\gamma > 0$ and $m \in \mathbb{R}$ if the pdf is

$$f_X(x;m,\gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-m}{\gamma}\right)^2\right)}, \quad \forall\, x \in \mathbb{R}, \tag{3.22}$$

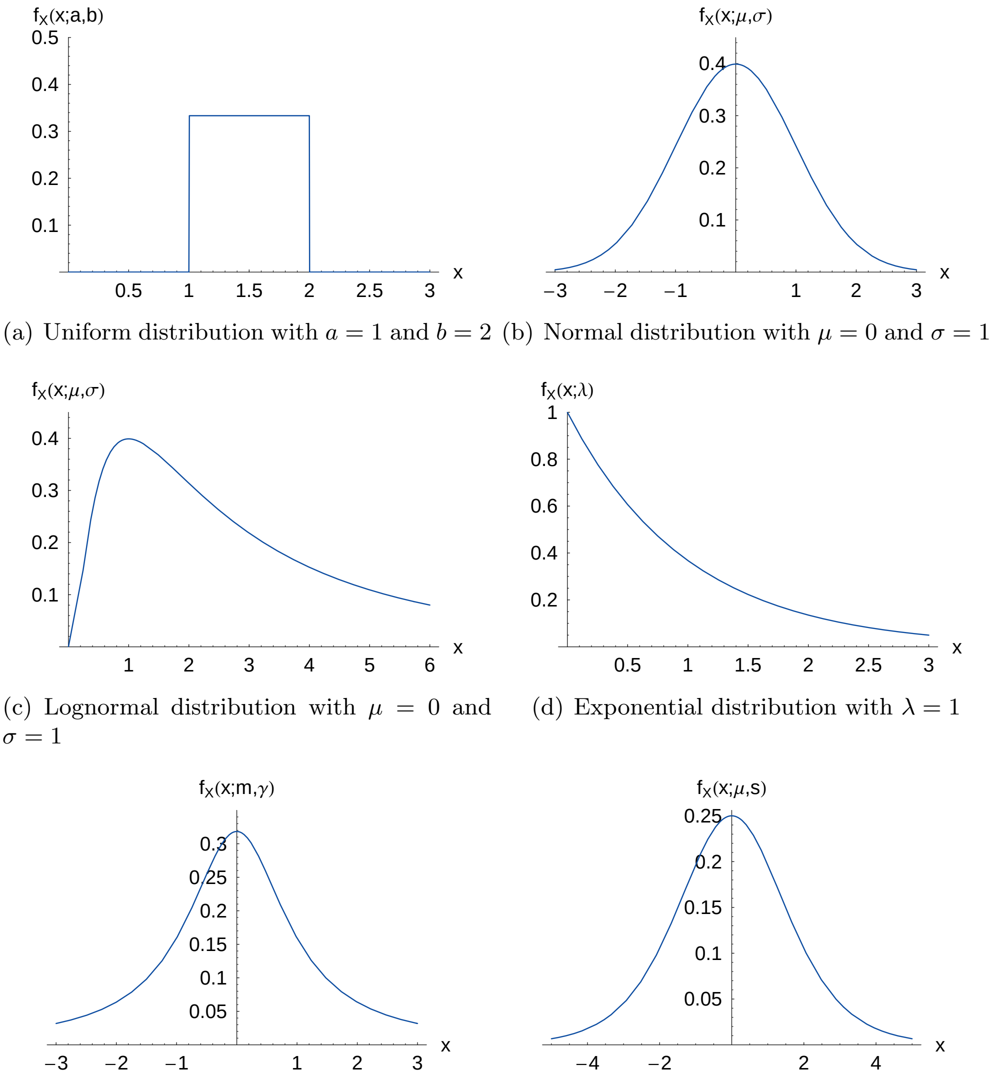
(f) *logistic distribution* with parameters $\mu \in \mathbb{R}$ and $s > 0$ if the pdf is

$$f_X(x;\mu,s) = \frac{e^{-(x-\mu)/s}}{s(1+e^{-(x-\mu)/s})^2}, \quad \forall\, x \in \mathbb{R}. \tag{3.23}$$

Plots of the distributions defined in the previous definition are shown in Figure 3.1.

3.6 Moments

In this section we will be concerned with calculating the expectation and the variance of a rv. These two statistics give an indication of the location and spread of a distribution. Further, we will also determine whether a distribution is skewed to the left or right of its expectation by calculating its skewness, and whether the distribution is strongly centered around its expectation by calculating its kurtosis.



(a) Uniform distribution with $a = 1$ and $b = 2$ (b) Normal distribution with $\mu = 0$ and $\sigma = 1$

(c) Lognormal distribution with $\mu = 0$ and $\sigma = 1$ (d) Exponential distribution with $\lambda = 1$

(e) Cauchy distribution with $m = 0$ and $\gamma = 1$ (f) Logistic distribution with $\mu = 0$ and $s = 1$

Figure 3.1: Probability density functions of the uniform, normal, lognormal, exponential, Cauchy, and logistic distribution with selected parameters

These distinguishing values for a rv are called its *moments*. In each case, we are required to compute the expected value of the rv raised to some integer power.

3.6.1 Expectation

The first moment of a rv X is called *expectation* (also referred to as *expected value* or *mean*). It represents the probability-weighted average value of that rv and is denoted as either $\mathrm{E}\,X$ or μ.

Definition 3.12. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a rv with

$$\int_{\mathbb{R}} |x| \, P_X(dx) < \infty \, .$$

Then the rv X is *integrable*, and the *expectation* (or *expected value* or *mean*) $\mathrm{E}\,X$ of X is given by the following (Lebesgue-) integral

$$\mathrm{E}\,X = \int_{\mathbb{R}} x \, P_X(dx) \, . \tag{3.24}$$

An alternative way of defining the expectation of a rv X is to use the *Lebesgue-Stieltjes integral* regarding the cdf F_X of X, then $\mathrm{E}\,X = \int_{-\infty}^{\infty} x \, dF_X(x)$. With the definition of the expectation given in (3.24), both the definitions of discrete and continuous rv can easily be obtained.

Theorem 3.8. *Let $X : \Omega \to \mathbb{R}$ be a rv.*

1. *If X is discrete with $P(X \in C) = 1$ for a countable set $C \subset \mathbb{R}$, then the expectation $\mathrm{E}\,X$ of X is given by the weighted average*

$$\mathrm{E}\,X = \sum_{x \in C} x P(X = x), \tag{3.25}$$

provided that

$$\sum_{x \in C} |x| P(X = x) < \infty \, . \tag{3.26}$$

2. *If X is continuous with the pdf $f_X(x)$, then the expectation $\mathrm{E}\,X$ of X is given by the integral*

$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x) \, dx, \tag{3.27}$$

provided that

$$\int_{-\infty}^{\infty} |x| f_X(x) \, dx < \infty \, . \tag{3.28}$$

Proof.

1. Let X be discrete with $P(X \in C) = 1$ for a countable set $C \subset \mathbb{R}$, then

$$P_X(B) = \sum_{x \in B \cap C} P(X = x), \quad \forall\, B \in \mathcal{B}(\mathbb{R}).$$

   With this and (3.24) the claim follows.

2. Let X now be continuous with the pdf $f_X(x)$, then by (3.15) holds

$$P_X(B) = \int_B f_X(x)\, dx, \quad \forall\, B \in \mathcal{B}(\mathbb{R}).$$

   With this and (3.24) the claim follows.

□

Example 3.3. We might think of the expectation of a rv as being analogous to a center of mass. This is illustrated in Figure 3.2. The pdfs of two rvs are indicated with their means on the x-axes. It can be confirmed by visual inspection that the indicated means appear to be "centers of mass" for the two pdfs.

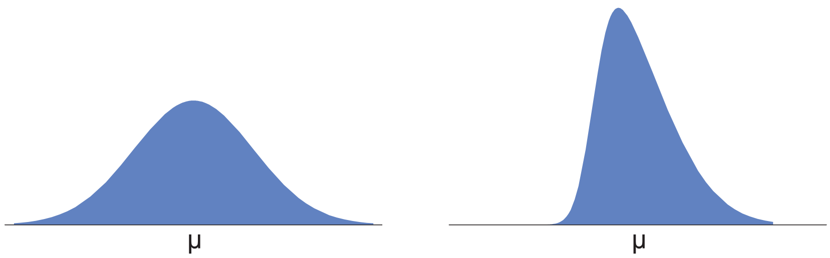


Figure 3.2: Notion of the expected value

Lemma 3.3. *Let X be a rv with pdf according to Definition 3.11. Then its expectation is defined as follows*

(a)	*uniform distribution:*	$\mathrm{E}\, X = (a+b)/2$	(3.29)
(b)	*normal distribution:*	$\mathrm{E}\, X = \mu$	(3.30)
(c)	*lognormal distribution:*	$\mathrm{E}\, X = e^{\mu + \sigma^2/2}$	(3.31)
(d)	*exponential distribution:*	$\mathrm{E}\, X = \lambda^{-1}$	(3.32)
(e)	*Cauchy distribution:*	*undefined*	(3.33)
(f)	*logistic distribution:*	$\mathrm{E}\, X = \mu.$	(3.34)

Properties of the Expectation

Some useful properties of the expectation of a rv are the following:

Theorem 3.9. *Let $X, Y : \Omega \to \mathbb{R}$ be arbitrarily rvs over an arbitrary probability space $(\Omega, \mathcal{A}, P)$. Then the following holds:*

1. Inequality: *If X and Y are integrable and if $X \leq Y$, then it holds*

$$\mathrm{E}\,X \leq \mathrm{E}\,Y, \tag{3.35}$$

especially

$$|\,\mathrm{E}\,X| \leq \mathrm{E}\,|X|. \tag{3.36}$$

2. Linearity: *If X and Y are integrable, then $aX + bY$ is also integrable for arbitrary $a, b \in \mathbb{R}$, and it holds*

$$\mathrm{E}\,(aX + bY) = a\,\mathrm{E}\,X + b\,\mathrm{E}\,Y. \tag{3.37}$$

Proof.

1. The first part follows immediately from the integral representation of the expectation and the monotony property of the integral. The second holds, because if X is integrable, also $|X|$ resp. $-|X|$ are, and since $X \leq |X|$ resp. $-|X| \leq X$, form (3.35) follows that $\mathrm{E}\,X \leq \mathrm{E}\,|X|$ resp. $\mathrm{E}\,|X| = \mathrm{E}\,-|X|) \leq \mathrm{E}\,X$.

2. That $aX + bY$ is integrable follows from $|aX + bY| \leq |a||X| + |b||Y|$, from (3.35), and from the linearity of the Lebesgue integral, since $\mathrm{E}\,|aX+bY| \leq \mathrm{E}(|a||\,\mathrm{E}\,|X| + |b|\,\mathrm{E}\,|Y| < \infty$. The validity of (3.35) follows then also from the linearity of the Lebesgue integral.

□

3.6.2 *k*th Moments

We have already seen how to calculate the expectation of a rv, also termed the first raw moment. In the following, the expected value of a rv raised to an integer power will be investigated.

Definition 3.13. Given a rv X, the *kth moment* $\mu'_k(a)$ of X about a is the expected value of the function $g(X) = (X - a)^k$, provided the expectation exists. Thus

$$\mu'_k(a) = \mathrm{E}\big(g(X)\big) = \mathrm{E}\big((X - a)^k\big), \qquad k = 0, 1, \dots\,. \tag{3.38}$$

If $a = 0$, the kth moment taken about 0 is termed *raw moment*, and is denoted by μ'_k. Is $a = \mu'_1 = \mu$ the expected value of X, we get the *kth central moment*, denoted by μ_k. Note that although the expected value of X is the first raw moment μ'_1, we continue to write μ instead of μ'_1.

The kth moment about 0 (kth raw moment) is

$$\mu'_k = \mathrm{E}(X^k), \qquad k = 0, 1, \dots\,, \tag{3.39}$$

and the kth moment about the mean (kth central moment) reads

$$\mu_k = \mathrm{E}\big((X - \mu)^k\big), \qquad k = 0, 1, \dots\,. \tag{3.40}$$

The central moments μ_k can be expressed in terms of the raw moments μ'_k using (3.40) or the following binomial transform, cf. [Rasch, 1995]:

$$\mu_k = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \mu'_{k-i} \mu^i. \tag{3.41}$$

The first four central moments are then given by

$$\mu_1 = 0 \tag{3.42}$$
$$\mu_2 = \mu'_2 - {\mu'_1}^2 \tag{3.43}$$
$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2{\mu'_1}^3 \tag{3.44}$$
$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6{\mu'_1}^2\mu'_2 - 3{\mu'_1}^4. \tag{3.45}$$

3.6.3 Variance

The second central moment is called variance and gives an indication of the spread of a rv's distribution.

Definition 3.14. The *variance* of a rv X is defined as

$$\operatorname{Var} X = \mu_2 = \operatorname{E}\left((X - \operatorname{E} X)^2\right). \tag{3.46}$$

The positive square root of the variance of a rv X is called its *standard deviation* or *mean square deviation* and is denoted by σ,

$$\sigma = \sqrt{\operatorname{Var} X}. \tag{3.47}$$

Theorem 3.10. *Let $X : \Omega \to \mathbb{R}$ be a rv with $\operatorname{E}(X^2) < \infty$. Then it holds*

$$\operatorname{Var} X = \operatorname{E}(X^2) - (\operatorname{E} X)^2, \tag{3.48}$$

and for arbitrary $a, b \in \mathbb{R}$

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X. \tag{3.49}$$

Proof. From the definition of the variance (3.46) and the linearity of the expectation (3.37) it follows

$$\begin{aligned}
\operatorname{Var} X &= \operatorname{E}\left((X - \operatorname{E} X)^2\right) = \operatorname{E}\left(X^2 - 2X \operatorname{E} X + (\operatorname{E} X)^2\right) \\
&= \operatorname{E}(X^2) - \operatorname{E}(2X \operatorname{E} X) + \operatorname{E}\left((\operatorname{E} X)^2\right) \\
&= \operatorname{E}(X^2) - 2 \operatorname{E} X \operatorname{E} X + (\operatorname{E} X)^2 \\
&= \operatorname{E}(X^2) - (\operatorname{E} X)^2 .
\end{aligned}$$

From (3.48) it follows,

$$
\begin{aligned}
\operatorname{Var}(aX+b) &= \operatorname{E}\left((aX+b)^2\right) - \left(\operatorname{E}(aX+b)\right)^2 \\
&= \operatorname{E}\left((aX)^2 + 2abX + b^2\right) - (a\operatorname{E}X + b)^2 \\
&= a^2\operatorname{E}(X^2) + 2ab\operatorname{E}X + b^2 - a^2(\operatorname{E}X)^2 - 2ab\operatorname{E}X - b^2 \\
&= a^2\left(\operatorname{E}(X^2) - (\operatorname{E}X)^2\right) \\
&= a^2\operatorname{Var}X\,.
\end{aligned}
$$

□

Example 3.4. The variance measures the dispersion of a rv's probability distribution. In Figure 3.3 probability density functions are indicated for two rvs. The one on the left has a higher variance. It is more dispersed than the one on the right.

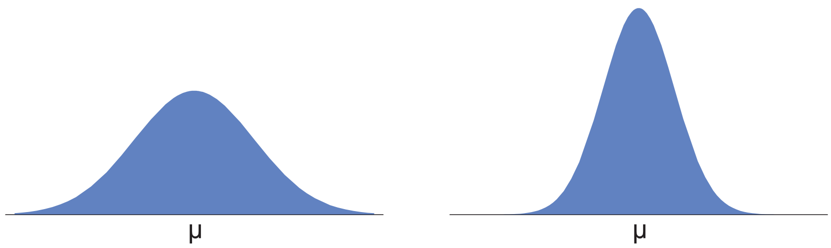


Figure 3.3: Notion of the variance

Lemma 3.4. *Let X be a rv with pdf according to Definition 3.11. Then its variance is defined as follows*

(a)	*uniform distribution:*	$\operatorname{Var}X = (b-a)^2/12$	(3.50)
(b)	*normal distribution:*	$\operatorname{Var}X = \sigma^2$	(3.51)
(c)	*lognormal distribution:*	$\operatorname{Var}X = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$	(3.52)
(d)	*exponential distribution:*	$\operatorname{Var}X = \lambda^{-2}$	(3.53)
(e)	*Cauchy distribution:*	*undefined*	(3.54)
(f)	*logistic distribution:*	$\operatorname{Var}X = (\pi\, s)^2/3.$	(3.55)

3.6.4 Skewness

Skewness is the third standardized moment and measures the degree of symmetry, or more precisely, the lack of symmetry of a rv's distribution. A distribution is referred to as *negatively skewed*, if its left tail is more pronounced than its right tail. Skewness can range from minus infinity to infinity.

Definition 3.15. The *skewness* of a random variable X is defined as

$$\gamma_1(X) = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = \frac{\operatorname{E}[(X-\mu)^3]}{\sigma^3}, \tag{3.56}$$

where μ_i and μ are the ith central moment and the mean, respectively, and σ denotes the standard deviation.

Several other forms of skewness have also been defined; e.g. Pearson [1905] suggested measuring skewness by standardizing the difference between the mean and the mode of a distribution (referred to as the *Pearson mode skewness*),

$$\operatorname{skew} X = \frac{\mu - \operatorname{mode} X}{\sigma}, \tag{3.57}$$

where $\operatorname{mode} X$ gives the mode of the distribution.

Example 3.5. A distribution is symmetric if it looks the same to the left and right of the center point. In Figure 3.4 probability density functions are indicated for two rvs with the same mean and variance. The one on the left is negatively skewed, i.e. its left tail is heavier than its right tail; whereas the one on the right is positively skewed.

Figure 3.4: Notion of the skewness

Lemma 3.5. *Let X be a rv with pdf according to Definition 3.11. Then its skewness is defined as follows*

(a)	*uniform distribution:*	$\gamma_1(X) = 0$	(3.58)
(b)	*normal distribution:*	$\gamma_1(X) = 0$	(3.59)
(c)	*lognormal distribution:*	$\gamma_1(X) = (e^{\sigma^2} - 1)^{1/2}(2 + e^{\sigma^2})$	(3.60)
(d)	*exponential distribution:*	$\gamma_1(X) = 2$	(3.61)
(e)	*Cauchy distribution:*	*undefined*	(3.62)
(f)	*logistic distribution:*	$\gamma_1(X) = 0.$	(3.63)

3.6.5 Kurtosis

Pearson [1905] introduced kurtosis as a measure of how flat the top of a symmetric distribution is when compared to a normal distribution of the same variance.

In older works, kurtosis is sometimes defined as the fourth standardized moment of a distribution,

$$\beta_2(X) = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}. \tag{3.64}$$

More commonly used is the following definition of kurtosis, which is also known as *excess kurtosis*:

Definition 3.16. The *kurtosis* of a rv X is defined as

$$\gamma_2(X) = \beta_2(X) - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mathrm{E}\left[(X-\mu)^4\right]}{\sigma^4} - 3, \tag{3.65}$$

where μ_i and μ are the ith central moment and the mean, respectively, and σ denotes the standard deviation.

Compared to a normal distribution, Pearson [1905] classified more flat-topped distributions, $\gamma_2 < 0$, as *platykurtic*, less flat-topped distributions, $\gamma_2 > 0$, as *leptokurtic*, and equally flat-topped distributions ($\gamma_2 = 0$) as *mesokurtic*. However, kurtosis is actually more influenced by the characteristic of the distribution's tails than the characteristic of the distribution's center [DeCarlo, 1997]. Accordingly, it is often reasonable to describe a leptokurtic distribution as "fat in the tails" and a platykurtic distribution as "thin in the tails".

Example 3.6. In Figure 3.5 probability density functions are indicated for two rvs to illustrate the kurtosis. The one on the left is more peaked at the center and has fatter tails. Hence, it has a greater kurtosis than the one on the right. Note that it is impossible to say which one has the greater variance. The one on the left is more peaked at the center, which might indicate that it has a lower variance. On the other hand it has fatter tails, which might indicate that it has a higher variance. If the effect of the peakedness exactly offsets that of the fat tails, the two probability density functions will have the same variance.

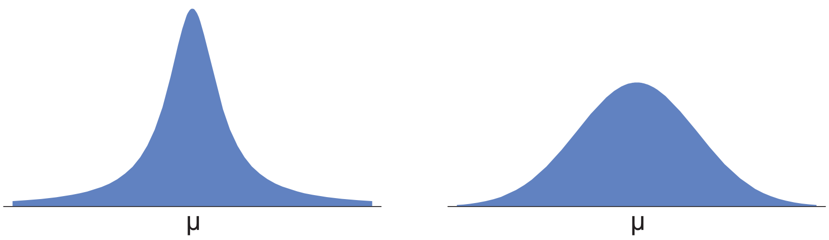


Figure 3.5: Notion of the kurtosis

Lemma 3.6. *Let X be a rv with pdf according to Definition 3.11. Then its kurtosis is defined as follows*

(a) *uniform distribution:* $\gamma_2(X) = -6/5$ (3.66)

(b) *normal distribution:* $\gamma_2(X) = 0$ (3.67)

(c) *lognormal distribution:* $\gamma_2(X) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$ (3.68)

(d) *exponential distribution:* $\gamma_2(X) = 6$ (3.69)

(e) *Cauchy distribution:* *undefined* (3.70)

(f) *logistic distribution:* $\gamma_2(X) = 6/5.$ (3.71)

3.6.6 Moment-Generating Function

Provided it exists, the moment-generating function generates the moments of a pdf, whereby the distribution of the corresponding rv is *uniquely* defined.

Definition 3.17. The *moment-generating function* of a rv X is defined as the expected value of $g(X) = e^{tX}$, i.e.

$$M(t) = \mathrm{E}(e^{tX}), \qquad -\infty < t < \infty \tag{3.72}$$

provided that the expected value exists.

If $M(t)$ is differentiable at zero, then the nth moment about the origin is generated as follows,

$$\mathrm{E}(X^n) = M^{(n)}(0) = \frac{d^n}{dt^n}\bigg|_{t=0} M_X(t). \tag{3.73}$$

For a rv X with continuous probability density function $f(x)$ the moment generating function is given by,

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} (1 + tx + \frac{1}{2!}t^2x^2 + \dots) f(x) dx \\ &= 1 + t\mu_1' + \frac{1}{2!}t^2\mu_2' + \dots, \end{aligned} \tag{3.74}$$

where μ_k' is the kth raw moment. Sometimes it is simpler to work with the logarithm of the moment-generating function, which is called the *cumulant-generating function*, and is defined by $K(t) = \ln M(t)$.

3.7 Functions of Random Variables

Since we will later need linear transformations of rvs and sums of normal rvs, these will be presented in the following. The issue is treated in-depth for example in the monograph of Springer [1979].

Let X be a rv and $Y = g(X)$. Then also Y is a rv since, for any outcome ω, $Y(\omega) = g(X(\omega))$. Often we know $F_X(x)$ and wish to calculate $F_Y(y)$ and $f_Y(y)$. The distribution function then must satisfy

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y).$$

Thus, we need to find all intervals on the x axis such that $g(x) \leq y$ in order to calculate the above probability from $F_X(x)$.

Linear Transformations

An important operation on a rv is its linear transformation, $\varphi(x) = ax + b$ with $\varphi : \mathbb{R} \to \mathbb{R}$ and $a, b \in \mathbb{R}$.

Theorem 3.11. *Let $X : \Omega \to \mathbb{R}$ be an arbitrarily rv and $a, b \in \mathbb{R}$ arbitrary constants with $a \neq 0$. Then $aX + b$ is a rv, and*

1. *the cdf of $aX + b$ is given by*

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & \text{if } a > 0, \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right), & \text{if } a < 0, \end{cases} \tag{3.75}$$

2. *if X is absolutely continuous with pdf f_X, then $aX + b$ is also absolutely continuous with pdf*

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right). \tag{3.76}$$

Proof. Omitted.

Sums and Differences of Normal Random Variables

Lemma 3.7. *Let X_1 and X_2 be independent rvs with $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then the sum $X_1 + X_2$ and the difference $X_1 - X_2$ are normally distributed with*

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2), \tag{3.77}$$

resp.

$$X_1 - X_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)\,. \tag{3.78}$$

Proof. Omitted. □

3.8 Random Variate Generation by the Inversion Method

The following theorem can be used to generate random variates with an arbitrary continuous distribution function F provided that the inverse function F^{-1} is explicitly known.

Theorem 3.12 (Devroye [1986, p. 28])**.** *Let F be a continuous distribution function on $\mathbb{R}$ with inverse F^{-1} defined by*

$$F^{-1}(u) = \inf\{x \mid F(x) = u, 0 < u < 1\}.$$

If U is a uniform $[0, 1]$ random variable, then $F^{-1}(U)$ has the distribution function F. Moreover, if X has the distribution function F, then $F(X)$ is uniformly distributed on $[0, 1]$.

Proof. The first claim follows after noting that for all $x \in \mathbb{R}$,

$$P(F^{-1}(U) \leq x) = P(\inf\{y \mid F(y) = U\} \leq x) = P(U \leq F(x)) = F(x).$$

The second statement follows from the fact that for all $0 < u < 1$,

$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

□

Hence, a random variate with distribution function F can be generated by means of a uniform $[0, 1]$ rv with the transform $X = F^{-1}(U)$. The expense to generate X depends on the expense the computation of the inverse causes. Theorem 3.12 will be used extensively by the directed mutation operators presented in Chapter 7.

4 Evolution Strategies

This chapter presents Evolution Strategies in detail, preluded by an intuitive flow-chart. In Section 4.2 a general Evolution Strategy is introduced together with the standard notation. The following Sections 4.3 to 4.7 deal with the different operators used within. Section 4.8 is devoted to the adaptation of strategy parameters and a state-of-the-art Evolution Strategy with covariance matrix adaptation is outlined in Section 4.9.

4.1 Introduction

In Section 2.2 Evolution Strategies (ESs) have been introduced as representatives of guided random optimization algorithms, mimicking the principles of natural evolution to find optimal solutions to an optimization problem. Accordingly, an ES comprises of a selection, recombination, and mutation operator. These three operators are successively processed for all considered ESs, as illustrated by the evolutionary loop in Figure 4.1.

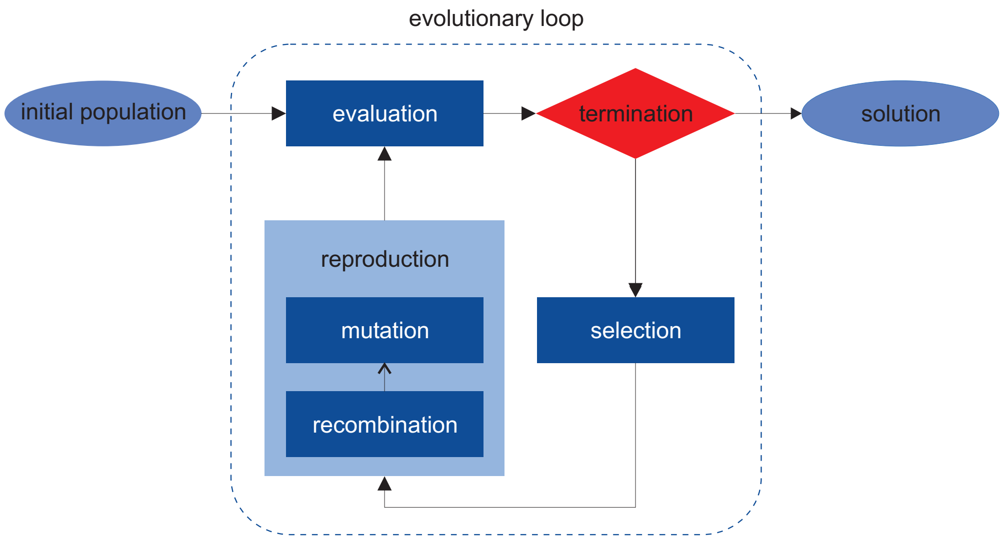


Figure 4.1: Flow-chart of an Evolution Strategy

4.2 The $(\mu/\rho \overset{+}{,} \lambda)$-ES: Algorithm and Notation

A more detailed view of the algorithm shown in Figure 4.1 is given by the pseudo-code of the $(\mu/\rho \overset{+}{,} \lambda)$-ES in Listing 4.1. In the following we briefly outline the algorithm, postponing most of the details to the next sections.

Listing 4.1: Pseudocode of the $(\mu/\rho \overset{+}{,} \lambda)$-ES

```
begin
  g := 0;
  init( P_μ^(0) := { a_m^(0) } = { ( x_m^(0), s_m^(0), F(x_m^(0)) ) } );
  repeat
    for l:=1 to λ do begin
      R_l  := reproduction( P_μ^(g), ρ );
      s_l  := sRecombination(R_l);
      s̃_l  := sMutation(s_l);
      x_l  := xRecombination(R_l);
      x̃_l  := xMutation(x_l, s̃_l);
      f̃_l  := F(x̃_l);
    end;
    P̃_λ^(g) := { ã_l^(g) } = { ( x̃_l, s̃_l, f̃_l ) };
    case selectionType of
      ,: P_μ^(g+1) := selection( P̃_λ^(g), μ );
      +: P_μ^(g+1) := selection( P̃_λ^(g), P_μ^(g), μ );
    end;
    g := g + 1;
  until terminationCriterion;
end;
```

The evolutionary optimization initiates by generating a first population of individuals at generation $g = 0$. The population $\mathbf{P}$ at generation g consists of μ parental individuals and λ descendants and can be described by $\mathbf{P}^{(g)}_{\mu+\lambda} = \{\mathbf{I}^{(g)}_i\}_{i=1,\ldots,\mu+\lambda}$. An individual i in the population consists of a set of decision and strategy variables, $I_i \in \mathcal{I}$, where $\mathcal{I}$ is the state space. During the evolutionary optimization process the population evolves over several generations g, until some termination criterion is fulfilled. Within each generation, the variation operators recombination and mutation are applied and selection takes place. The recombination operator exchanges information in the population in order to spread good properties of solutions. The mutation operator adds random variations to the variables and thus increases diversity in the population. Selection is the counterpart to the variation operators. It decreases on average the diversity in the population by selecting fitter individuals. Thereby selection guides the search into promising regions of the object parameter space and increases the fitness in the population.

The way the offspring population is generated is reflected in the $(\mu/\rho \overset{+}{,} \lambda)$ notation. Besides the number of parents μ and offspring λ, the number of parents involved in the procreation of one offspring is stated by ρ. The special case $\rho = 1$ represents cloning, i.e. no recombination takes place. The parameter is then usually omitted and the notation reduces to $(\mu \overset{+}{,} \lambda)$. The case $\rho > 1$ results in strategies with recombination. The symbol "$\overset{+}{,}$" refers to the kind of selection to be used, i.e. "+"- or ","-selection, respectively (see Section 4.4). Sometimes another parameter κ is introduced that determines the maximal life span of an individual [Schwefel and Rudolph, 1995]. The strategy specific parameters μ, ρ, λ, and κ are termed *exogenous strategy parameters* which are kept constant during the evolutionary run.

The quality of the whole ES depends highly on the interplay of these operators. Some combinations of recombination and mutation operators may even lead to divergence [Kursawe, 1995].

Mutation is considered the main operator in ESs [Bäck and Schwefel, 1993], in detail treated in Section 4.7. It most often evolves itself during the optimization, called *self-adaptation* (see Section 4.8). Self-adaptation is necessary for fast convergence as shown by Rechenberg [1973].

4.3 Representation

Evolution strategies are population based algorithms. Thus, the strategy operates on a population $\mathbf{P}$ that consists of a set of individuals $\mathbf{I}$. Besides the object parameter set $\mathbf{x}$, $\mathbf{x} \in \mathcal{X}$, and its objective function value $f(\mathbf{x})$, $f \in \mathcal{F}$ (also referred to as fitness), an individual usually also contains a set $\mathbf{s}$, $\mathbf{s} \in \mathcal{S}$, of *endogenous* (i.e. evolvable) strategy parameters

$$\mathbf{I} := (\mathbf{x}, \mathbf{s}, f(\mathbf{x})), \quad \text{with } \mathbf{I} \in \mathcal{I}. \tag{4.1}$$

The occurrence of the endogenous strategy parameter set $\mathbf{s}$ is a peculiarity of ESs. It encodes certain properties of the genetic operators, especially those of the mutation operator (see Section 4.7). The endogenous strategy parameter set is not *directly* involved in the calculation of the fitness of the individual; however, it is passed to the offspring depending on the fitness value of the individual. During the optimization process it can evolve and is an inevitable ingredient in self-adaptive ESs (cf. Section 4.8).

Hence, the state space of an individual is completely defined by the tuple $(\mathbf{x}, \mathbf{s}, f(\mathbf{x}))$. The total of these elements composes the state space $\mathcal{I}$,

$$\mathcal{I} = \mathcal{X} \times \mathcal{S} \times \mathcal{F}, \tag{4.2}$$

where evolution takes place.

As already mentioned, the individuals form a population $\mathbf{P}$. Within this population, μ parents $\mathbf{I}_m$, $m = 1, \ldots, \mu$, and λ offspring $\mathbf{I}_l$, $l = 1, \ldots, \lambda$ exist. The parameters μ and λ are *exogenous* strategy parameters, i.e. they are unchanged by the ES. The distinct populations of the parents and offspring at generation g are denoted by $\mathbf{P}_\mu^{(g)}$ and $\mathbf{P}_\lambda^{(g)}$, respectively

$$\mathbf{P}_\mu^{(g)} := \left\{\mathbf{I}_m^{(g)}\right\} = \left(\mathbf{I}_1^{(g)}, \ldots, \mathbf{I}_\mu^{(g)}\right) \tag{4.3}$$

$$\mathbf{P}_\lambda^{(g)} := \left\{\mathbf{I}_l^{(g)}\right\} = \left(\mathbf{I}_1^{(g)}, \ldots, \mathbf{I}_\lambda^{(g)}\right). \tag{4.4}$$

4.4 Selection Operator

In nature, evolution is a highly parallel process carried out with large populations of individuals. If the growth of the population is limited by some constraints, selection takes place. In ESs, the growth of the population is also limited. A selection pressure is induced by the ratio of the offspring to the parents, λ/μ. Hence, in each generation g, only a fraction of the population is selected to serve as parental population in the next generation $\mathbf{P}_\mu^{(g+1)}$,

$$\mathrm{sel} : \mathbf{I}^\gamma \rightarrow \mathbf{I}^\mu. \tag{4.5}$$

The selection operator is a *deterministic* procedure, which chooses the μ best individuals from the set of γ individuals $(\mathbf{I}_1, \ldots, \mathbf{I}_\gamma)$ according to their fitness value $f(\mathbf{x})$. Note that the fitness is only determined by the object parameters. Using the same notation as Beyer [2001], selection reads

$$\mathtt{sel}(I_1, \ldots, I_\gamma) := (I_{1;\gamma}, \ldots, I_{\mu;\gamma}), \qquad \gamma \geq \mu. \tag{4.6}$$

The symbol $(\cdot)_{m;\gamma}$ represents the individual with the mth best fitness value, $(\cdot)_{1;\gamma}$ the best, and $(\cdot)_{\gamma;\gamma}$ the worst individual, respectively. The advantage of this notation is the unification of the two optimization types, minimization and maximization. It is a generalization of the $(\cdot)_{m:\gamma}$ notation known from the theory of order statistics, which expresses the order relation

$$F_{1:\gamma} \leq \cdots \leq F_{m:\gamma} \leq \cdots \leq F_{\gamma:\gamma}. \tag{4.7}$$

Two versions of selection techniques are commonly used. Depending on whether or not the parental population is included in the selection pool, we distinguish *plus* selection, denoted by $(\mu + \lambda)$, and *comma* selection, denoted by (μ, λ), respectively.

In the case of plus selection, both the current population and their parents are copied to the selection pool which is thus of size $\gamma = \mu + \lambda$. This version of selection is *elitist* by preserving always the best solutions obtained so far. Hence, a deterioration of the fitness in the parental population is avoided and constant improvement of the fitness is ensured.

In contrast to plus selection, the comma selection only takes the current population into account. Therefore the fitness of the population can decreases between two generations. While at first this seems unfavorable for an optimization algorithm, in this way the plus selection ensures a constant change in the parental population.

Both variants have rather specific application areas. Based on simulation results, Schwefel [1987] recommends comma selection in the field of real-valued parameter optimization. In the case of combinatorial optimization problems Beyer [1992] recommends plus selection, due to the principal failure of self-adaptation of the mutation strength.

Introducing a maximal life span of an individual, $1 \leq \kappa \leq \infty$, both selection techniques can be unified in the (μ, κ, λ)-selection [Bäck et al., 1997b]. Then $\kappa = 1$ corresponds to the (μ, λ)-strategy and $\kappa = \infty$ to the $(\mu + \lambda)$-strategy, respectively.

4.5 Reproduction Operator

The reproduction operator selects the $1 \leq \rho \leq \mu$ parents which will procreate one offspring individual,

$$\text{rep} : \mathbf{I}^{\mu} \rightarrow \mathbf{I}^{\rho}. \tag{4.8}$$

The vector $\mathbf{R}$ of selected parents is defined as follows:

$$\mathbf{R} := \begin{cases} (\mathbf{I}_{i_1}, \ldots, \mathbf{I}_{i_r}, \ldots, \mathbf{I}_{i_\rho}), & \text{if } \rho < \mu \\ (\mathbf{I}_1, \ldots, \mathbf{I}_m, \ldots, \mathbf{I}_\mu), & \text{if } \rho = \mu \end{cases} \tag{4.9}$$

with

$$\begin{aligned} &\forall\, r \in \{1, \ldots, \rho\} : \quad i_r := \text{rand}\{1, \ldots, \mu\}, && \text{if } \rho < \mu \\ &m = 1, \ldots, \mu, && \text{if } \rho = \mu. \end{aligned} \tag{4.10}$$

In the case of $\rho = \mu$ all parents are involved in the production of the offspring, i. e. $\mathbf{R} = \mathbf{P}_\mu$. Whereas for $\rho < \mu$ only ρ parent individuals will procreate the offspring. They are chosen at random out of $\mathbf{P}_\mu$ with uniform probability $1/\mu$. For the sake of simplicity, selected individuals are returned to the pool again whereby "inbreeding" can occur. Of course this can be prevented using an appropriate selection mechanism. Anyway, to attain the maximum progress rate it is recommended to use all parents for recombination [Beyer, 2001].

4.6 Recombination Operator

In nature, most species recombine their genetic information by mating two parents. This allows to spread genetic information in the population and the offspring might benefit from combining preferable properties of their parents. In the ES context, mating is denoted as recombination. It is more flexible in the sense that an arbitrary number of parents can be involved in the recombination of the offspring.

The recombination operator creates one offspring by mixing the characteristics of the ρ parental individuals selected for reproduction,

$$\text{rec} : \mathbf{I}^\rho \to \mathbf{I}. \tag{4.11}$$

In particular, the case $\rho > 2$ is termed *multirecombination*. Two kinds of recombination are distinguished, *intermediate recombination* and *discrete recombination*, also called *dominant recombination*. Often different types of recombination operators are applied to the distinct subparts of the individuals.

The operator is defined on a ρ-tuple of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_\rho)$. The symbol $\mathbf{v}$ stands for the actual subpart of the individual's representation, i.e. the object parameters $\mathbf{x}$ or the strategy parameters $\mathbf{s}$.

Intermediate ρ-recombination: The descendant $\mathbf{r}$ is defined by the center of gravity of the selected parents

$$\mathbf{r} := \frac{1}{\rho} \sum_{i=1}^{\rho} \mathbf{v}_i. \tag{4.12}$$

The cases $\rho = 2$ and $\rho = \mu$ are simply called *intermediate recombination* and *global intermediate recombination*, respectively.

Discrete ρ-recombination: Every component of the descendant $\mathbf{r}$ is defined by random selection from the ρ corresponding components of the selected parents

$$\mathbf{r} := \sum_{i=1}^{d} \left(\mathbf{e}_i^T \mathbf{v}_{m_i}\right) \mathbf{e}_i, \tag{4.13}$$

with $d = \dim(\mathbf{x})$, $m_i := \text{rand}\{1, \ldots, \rho\}$, and the symbol $\mathbf{e}_i$ representing the unit vector in the ith dimension. Thus, the ith component of the descendant is specified exclusively by the ith component of a randomly selected parent. The cases $\rho = 2$ and $\rho = \mu$ are called *discrete recombination* and *global discrete recombination*, respectively.

4.7 Mutation Operator

Mutation is the process of adding random changes to the genetic information. This guarantees the collective genetic information to be constantly varied and enables a population to adapt to changing environmental conditions.

The mutation operator

$$\text{mut} : \mathbf{I} \to \mathbf{I} \tag{4.14}$$

is defined as

$$\text{mut} := \text{mut}_\text{o} \circ \text{mut}_\text{s} \tag{4.15}$$

such that first the strategy parameters undergo a mutation and the object parameters are then mutated with these strategy parameters.

A comprehensive discussion about mutation is given by Beyer [2001]. According to that, mutation operators should fulfill some general requirements:

1. *Reachability*

   Since the operator shall perform arbitrary search moves in the state space of the individual, it should be realized so that it can transfer the individual from its current state to any other state in finite time. The fulfillment of this requirement is a necessary condition for the functioning of any EA.

2. *Scalability*

   The mutation strength, i.e. the length of the search steps should be tunable for the calibration of the locally optimal mutation strength. To what extent this demand can be fulfilled depends on the granularity of the parameter space.

3. *Absence of biases*

   The mutation distribution should be chosen according to the maximum entropy principle. Following the central limit theorem, for unconstrained, real-valued search spaces this leads to a Gaussian normal distribution.

4. *Symmetry*

   This demand is strongly linked to the third requirement, but not equivalent to it. It reflects the supposed isotropy of the parameter space. Hence, under this assumption, the state change induced by the mutations should be zero.

Beyer points out that the given demands should be seen as guiding principles for the design of mutation operators. Only the first requirement expresses a certain minimal condition which should always be fulfilled, whereas the others are neither absolutely important nor definitely necessary. Since step size adaptation plays a crucial role in approaching the optimum [Rudolph, 1996], also the second requirement has some importance and should thus be fulfilled by a mutation operator.

According to the third requirement, the vector of object parameters $\mathbf{x}$ of a recombined individual is usually mutated by adding a normally distributed random vector $\mathbf{z}$

$$\tilde{\mathbf{x}} := \mathbf{x} + \mathbf{z}, \quad \text{with } \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}), \tag{4.16}$$

where $\mathcal{N}(\mathbf{0}, \mathbf{C})$ is a multivariate normal distribution with expectation $\operatorname{E}\mathbf{z} = \mathbf{0}$ and covariance matrix $\mathbf{C}$. Depending on the grade of variability granted to the normal distribution, three mutation strategies are commonly distinguished: isotropic mutation, scaled mutation, and correlated mutation.

4.7.1 Isotropic Mutation

In the simplest possible case, an *isotropic mutation* distribution is used to generate the random vector $\mathbf{z}$,

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) = \sigma \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{4.17}$$

Every component z_i is statistically independent of other components and thus the covariance matrix has only diagonal entries. Since the components also all have the same standard deviation $\sigma_i = \sigma$, it is the identity matrix $\mathbf{I}$. Hence, the previous equation can be written in a scalar form as

$$z_i \sim \sigma \mathcal{N}(0, 1), \quad i = 1, \dots, n. \tag{4.18}$$

A plot of the probability density is shown in Figure 4.2. Iso-probability contours are circles for $n = 2$ variables and (hyper-)spheres for $n \geq 3$.

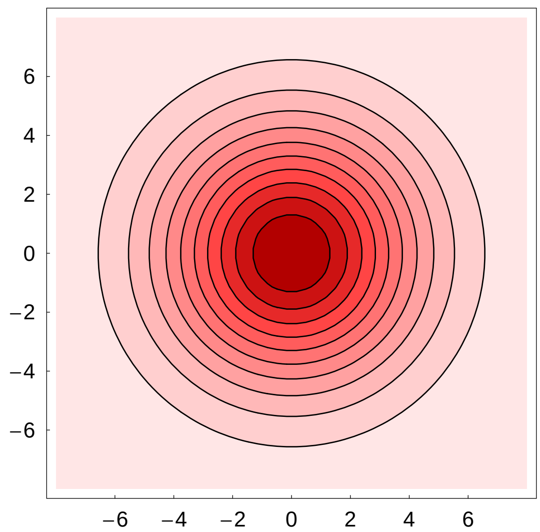


Figure 4.2: Iso-probability contours for isotropic mutations with $\sigma = 3$

4.7.2 Scaled Mutation

More flexibility is achieved by providing each decision variable with an individual step size. The resulting covariance matrix is diagonal, $\mathbf{C} = \text{diag}(\sigma_i^2, \dots, \sigma_n^2)$ and *scaled mutation* then reads

$$z_i \sim \sigma_i \mathcal{N}(0, 1), \quad i = 1, \dots, n. \tag{4.19}$$

By means of the separate mutation strengths, stretched or compressed variants of an isotropic probability distribution parallel to the coordinate axis can be realized. A probability density plot of this mutation is given in Figure 4.3. Iso-probability contours are (hyper-)ellipses with their principal axis parallel to the coordinate axis.

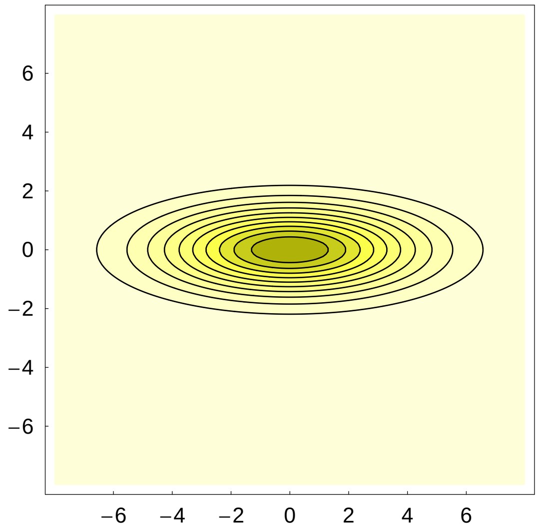


Figure 4.3: Iso-probability contours for scaled mutations with $\sigma_1 = 3$ and $\sigma_2 = 1$

4.7.3 Correlated Mutation

The whole flexibility of the multivariate normal distribution is obtained by allowing non-zero correlation coefficients in the covariance matrix. Because the decision variables are now mutated with correlated random variables, this mutation is referred to as *correlated mutation*. Thus, the random vector $\mathbf{z}$ is given by

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}). \tag{4.20}$$

An example of the probability density of this mutation is given in Figure 4.4. Iso-probability contours are linear transformations (rotation, scale) of an isotropic probability distribution. Now $n(n+1)/2$ strategy parameters need to be adapted, n step sizes (diagonal elements $c_{ii} = \sigma_i^2$), and due to the symmetry of the covariance matrix $n(n-1)/2$ correlation coefficients (non-diagonal elements $c_{ij} = c_{ji}, i \neq j$).

In practice, instead of sampling random numbers directly from $\mathcal{N}(\mathbf{0}, \mathbf{C})$ as implied by (4.20), contemporary ES use the decomposition $\mathbf{C} = (\mathbf{ST})^T(\mathbf{ST})$, where $\mathbf{S} = \mathbf{D}^{1/2}$ is the diagonal matrix of standard deviations ($s_{ii} = \sigma_i$) [Rudolph, 1992]. The matrix

$$\mathbf{T} = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \mathbf{R}_{ij}(\alpha_k) \tag{4.21}$$

is the product of the $n(n-1)/2$ elementary rotation matrices $\mathbf{R}_{ij}(\alpha_k)$, with angles $\alpha_k \in (0, 2\pi]$. The elementary rotation matrices $\mathbf{R}_{ij}$ are obtained from the identity matrix by replacing four elements as follows, $r_{ii} = r_{jj} = \cos\alpha$, $r_{ij} =$

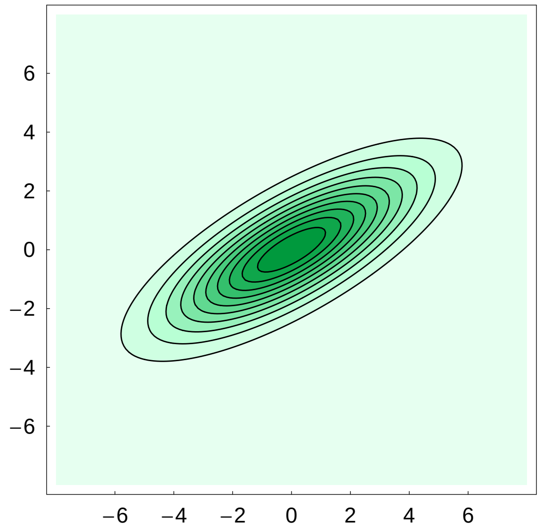


Figure 4.4: Iso-probability contours for correlated mutations with $\sigma_1 = 3$, $\sigma_2 = 1$ and $\alpha = \pi/3$

$-r_{ji} = -\sin\alpha$, yielding

$$\mathbf{T}_{ij}(\alpha_j) = \begin{pmatrix} 1 & 0 & & & & \cdots & & & & & 0 \\ 0 & 1 & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & 1 & & & & & & & \\ & & & & \cos\ \alpha_j & & & -\sin\ \alpha_j & & & \\ & & & & & 1 & & & & & \\ \vdots & & & & & & \ddots & & & & \vdots \\ & & & & & & & 1 & & & \\ & & & & \sin\ \alpha_j & & & \cos\ \alpha_j & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & 1 & 0 \\ 0 & & & & & \cdots & & & & 0 & 1 \end{pmatrix}. \tag{4.22}$$

Then with

$$z' \sim \mathbf{T}^T \mathbf{S}^T \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4.23}$$

random vectors with the same probability density function as in (4.20) can be generated. More details of this mutation variant can be found in [Rudolph, 1992] or [Bäck, 1996].

4.7.4 Problem Dependence

Which mutation distribution is chosen for an optimization is in the end problem dependent. It is always a trade-off between minimizing the learning effort and being able to adapt to mis-scaling or correlated decision variables. The increasing flexibility from isotropic mutation via scaled mutation to correlated mutation results in the number of strategy parameters growing from 1 via n to $n(n+1)/2$ with the concurrent adaptation effort. However, adaptation of the mutation distribution has been shown to be necessary for any efficient optimization algorithm as constant distributions are far from being optimal [Bäck et al., 1997a]. Several techniques are available for this purpose, as described in the following Section 4.8.

4.8 Adaptation of the Strategy Parameters

This section is dedicated to the adaptation of strategy parameters controlling the properties of the variation operators, in particular the parameters of the mutation operator. After the need of a mutation strength control is motivated, the famous 1/5th-success rule for controlling the mutation strength in (1+1)-ES is presented. Next, the ideas of self-adaptation are rendered and an advanced adaptation technique utilizing nonlocal search space information is outlined.

4.8.1 Motivation

The necessity of controlling endogenous strategy parameters already becomes evident when running a simple (1+1)-ES with fixed uniform mutations on the sphere function. Rudolph [1996] shows for this scenario, that the convergence rate of an ES with fixed mutation strength declines to the asymptotics of a pure random search. Beyer and Schwefel [2002] give an example where an ES with isotropic normal mutations has lost its evolvability and stagnates after a period of improvements.

4.8.2 One-Fifth Success Rule

In 1964 Rechenberg developed the first evolution strategy, the (1+1)-strategy [Rechenberg, 1964]. One parent creates a single offspring that is perturbed by an isotropic mutation, see Section 4.7.1. If the mutation is successful, i.e. the objective function value of the offspring is better than that of the parent, the parent is replaced by the offspring. For two simple objective functions, the sphere function (see Section C.1) and the corridor function[1], Rechenberg [1973] computed theoretically the ideal progress rate. The progress rate φ is a measure describing the expected local approach to the optimum. As well as the success probability P_s that measures the probability by which an offspring replaces a parent, the progress rate depends on the step size σ. Based on that fact, a σ control rule can be framed. Rechenberg concluded that the success probability P_s should be one fifth and formulated the adaptation of the step size σ as a function of the success in his famous *1/5th-success rule*:

> To obtain nearly optimal (local) performance of the (1+1)-ES in real-valued search spaces, the mutation strength has to be adjusted

[1] A linear objective function with 2(n - 1) inequality constraints

in such a way that the (measured) success rate is about 1/5.

An implementation of the 1/5th-success rule was given by Schwefel [1975]. He proposed to compute the success probability P_s always after generating n offspring, where n is the number of decision variables. After it is measured over the last $10n$ computed offspring, Rechenberg's adaptation rule should be applied the following way

$$\sigma^{(g+1)} = \begin{cases} c\,\sigma^{(g)}, & \text{if } p_s < 1/5 \\ \sigma^{(g)}, & \text{if } p_s = 1/5 \\ \sigma^{(g)}/c, & \text{if } p_s > 1/5, \end{cases} \tag{4.24}$$

where $\sigma^{(g)}$ is the step size of generation g and c is a multiplicative factor. The optimal value of the factor c depends on the objective function to be optimized, the dimensionality of the search space n, and on the number of generations G that are to be executed. Under the condition that n is sufficiently large, $n \geq 30$, G can be approximated by $G = n$ and Schwefel [1975] recommended using $0.85 \leq c \leq 1$. In general, the success rate P_s should be decreased if the objective function is noisy [Rechenberg, 1973], or constraints or local minima exist [Schwefel, 1995].

The 1/5th-success rule can be characterized by utilizing the global information of success probability in order to deterministically change the mutation strength, obtained by collecting statistical data over a number of generations. Obviously, the 1/5th-success rule is limited in its scope. For one thing it is restricted to the control of only one strategy parameter; for another thing it is highly dependent on the actual fitness landscape. Therefore, it is usually used in (1+1)-ES with isotropic mutation only.

4.8.3 Self-Adaptation

In the previous subsection we presented with the 1/5th-success rule a first heuristic for tuning the endogenous strategy parameter σ. To avoid such external control mechanisms, Rechenberg [1973] proposed that the values of multiple strategy parameters could instead result from a "learning population". He argued that this approach is more flexible since it can be applied with non-isotropic mutation distributions and more robust since it can handle discontinuities in the derivatives. Schwefel [1981] extended this approach to the correlated mutation setting and denoted his adaptation principle as "self-adaptation". While originally developed for mutation operator control in ESs, self-adaptation is meanwhile associated with a great variety of different operators. It found its way also to the field of GA, e.g. for the adaptation of the mutation and recombination rates [Bäck, 1992, Bäck and Schütz, 1996]. This more flexible, *evolutionary* control method will be explained next.

Motivation

The principal idea of self-adaptation bases on coupling the endogenous (i.e. evolvable) strategy parameters with the object parameters in the individuals. Hence, every individual carries its own set of strategy parameters in addition to its object parameters, as already hinted in Section 4.3. These endogenous strategy parameters are similar to the object parameters subject to variation. They may undergo recombination and mutation, as is apparent from the pseudo-code of the $(\mu/\rho \overset{+}{,} \lambda)$-ES in Listing 4.1. The altered strategy parameters configure then the mutation operator which is in turn applied to the individual's object parameters. Since the strategy parameters are not regarded during the calculation of the individual's fitness, they are selected indirectly via the fitness of the individual's object parameters. Hence, they have a higher probability of survival when they lead to object parameter variations that result on average in fitter object parameters. Using this methodology, it is plausible to expect the individuals to learn optimal strategy parameter settings during the evolutionary process. More detailed presentations of the concept of self-adaptation can be found in [Eiben et al., 1999] or [Bäck et al., 1997a, Sec. C7.1].

Implementation

The mutation strength σ is varied by a logarithmic normal mutation,

$$\tilde{\sigma} = \sigma \exp(\tau \mathcal{N}(0,1)). \tag{4.25}$$

The choice of the multiplicative logarithmic normal mutation of the strategy parameters is motivated by heuristic arguments [Schwefel, 1977]. The main reason is that this process preserves positive values. This is mandatory since σ represents the standard deviation of the mutation distribution, which is per definition positive. Then, the median should equal one. By this mutation technique this is guaranteed, since on average a multiplication by a certain value occurs with the same probability as a multiplication with its reciprocal value. Hence, under absence of selection the process would be neutral with respect to the mutation strength. Last, taking the natural logarithm on both sides of (4.25), we get $\ln \tilde{\sigma} = \ln \sigma + \tau \mathcal{N}(0,1)$. On a logarithmic scale the strategy parameters are thus mutated similarly to the way, mutations of the object parameters are performed.

The rotation angles α_j of the correlated mutation can be varied like the object parameters. Simply a normally distributed random number z_j is added.

Depending on the different mutation methods presented in Section 4.7, the set of strategy parameters incorporated into the individuals varies (see Section 4.3). The mutation of the strategy parameters then reads:

- *Isotropic mutation,* $(n_\sigma = 1, n_\alpha = 0)$
  As the standard deviation is identical for all objective parameters, only one strategy parameter σ has to be updated. All objective parameters are then mutated by adding normally distributed random numbers with that mutation strength,

$$\tilde{\sigma} = \sigma \exp(\tau_0 \, z_0), \quad \text{with } z_0 = \mathcal{N}(0,1) \tag{4.26}$$

$$\tilde{x_i} = x_i + \tilde{\sigma} \, z_i, \qquad \text{with } z_i = \mathcal{N}(0,1), \tag{4.27}$$

  where $\tau_0 = \frac{1}{\sqrt{n}}$.

- *Scaled mutation,* $(n_\sigma = n, n_\alpha = 0)$
  Every objective parameter has an individual standard deviation σ_i, thus n strategy parameters have to be updated. The objective parameters are then mutated by adding normally distributed random numbers with the corresponding mutation strengths,

$$\tilde{\sigma}_i = \sigma_i \exp(\tau' z_0 + \tau z_i), \quad \text{with } z_0, z_i = \mathcal{N}(0,1) \tag{4.28}$$
$$\tilde{x}_i = x_i + \tilde{\sigma}_i z_i, \quad \text{with } z_i = \mathcal{N}(0,1), \tag{4.29}$$

  where $\tau' = \frac{1}{\sqrt{2n}}$ and $\tau = \frac{1}{\sqrt{2\sqrt{n}}}$.

- *Correlated mutation,* $(n_\sigma = n, n_\alpha = \frac{n(n-1)}{2})$
  A covariance matrix has to be updated that is completely defined by n standard deviations and $n(n-1)/2$ correlation coefficients. The objective parameters are then mutated by adding normally distributed random numbers with the corresponding correlation,

$$\tilde{\sigma}_i = \sigma_i \exp(\tau_0 z_0 + \tau z_i), \quad \text{with } z_0, z_i = \mathcal{N}(0,1) \tag{4.30}$$
$$\tilde{\alpha}_j = \alpha_j + \beta z_j, \quad \text{with } z_j = \mathcal{N}(0,1) \tag{4.31}$$
$$\tilde{\mathbf{x}} = \mathbf{x} + \mathcal{N}(\mathbf{0}, \mathbf{C}(\tilde{\sigma}, \tilde{\alpha})), \tag{4.32}$$

  where $\mathcal{N}(\mathbf{0}, \mathbf{C}(\tilde{\sigma}, \tilde{\alpha}))$ denotes the correlated mutation vector, $\tau' = \frac{1}{\sqrt{2n}}$, $\tau = \frac{1}{\sqrt{2\sqrt{n}}}$, and $\beta \approx 0.0873 \approx 5°$.

The variables τ_0, τ, τ', and β are exogenous strategy parameters, called *learning rates*. The settings are recommended by Schwefel [1977] as reasonable heuristic settings. However, depending on the actual topology of the objective function, the optimal values of these parameters might differ from the proposed ones.

4.8.4 Nonlocal adaptation approaches

The major disadvantage of self-adaptation on the level of the individuals as introduced in Section 4.8.3 is its "opportunistic" behavior. Selection in ESs is a temporal local business. The individual's fitness is judged on its short term success only. However, local progress need not be positively correlated to global progress. The optimization process may thus get trapped in a local optimum and miss the global optimum. It may even exhibit so-called *premature convergence*.

Clearly, there is no general solution to these problems, but countermeasures can be taken to reduce the probability of such events. One such possibility is to utilize nonlocal – with respect to time – information. This can be achieved either by considering a history of the evolutionary process or by aggregation of state variables. A simple example of this approach has already been discussed in detail in Section 4.8.2: the 1/5th-success rule. Two more advanced techniques are the ideas of cumulative path-length control and the covariance matrix adaptation.

Cumulative Path-Length Control

Cumulative path length control is a deterministic nonlocal adaptation technique for the mutation strength control proposed by Ostermeier et al. [1994].

The basic idea is to exploit information of the so-called evolution path $\mathbf{p}$. In the simplest case, this is just the vector sum of the actually realized evolution steps over a number of g generations. By means of a comparison of $\mathbf{p}$ with the expected length of a path that would result from random selection, it is decided how the step size has to be adjusted.

The basic idea presented is to some extend a reminiscence of the 1/5th-success rule, as given in subsection 4.8.2. The main difference lies in the way how the nonlocal information is utilized. While with the 1/5th-success rule only the number of individuals hitting the local success domain are counted, evolution path related techniques elaborate search space information in a more advanced manner. They are also termed *Cumulative Step Size Adaptation* methods (CSA). Since it is an essential part of the Covariance Matrix Adaptation-ES, it is in more detail discussed in subsection 4.9.4.

Covariance Matrix Adaptation

In contrast to the stochastic update mechanism for the covariance matrix $\mathbf{C}$, as presented in Section 4.8.3, the Covariance Matrix Adaptation (CMA) tunes the matrix in a deterministic fashion. It is presented in the next section.

4.9 Covariance Matrix Adaptation-ES

The Covariance Matrix Adaptation-ES (CMA-ES) uses correlated mutations and a deterministic technique for the adaptation of the covariance matrix $\mathbf{C}$. It was proposed by Hansen and Ostermeier in 1996 [Hansen and Ostermeier, 1996] while first ideas date back to the year 1992 [Ostermeier, 1992]. In some sense, whereas CMA-ES has just been introduced as a deterministic correlated mutation algorithm from the ES point of view, from the point of view of numerical optimization it could also be seen as a stochastic steepest-descent technique.

In the sequel, we refer to articles from Hansen and Ostermeier [2001], from Hansen and Kern [2004], and another one from Hansen [2005].

In CMA-ES, individuals $\mathbf{x}_k$ are sampled from a multivariate normal distribution,

$$\mathbf{x}_k^{(g+1)} \sim \mathcal{N}\left(\mathbf{m}^{(g)}, \left(\sigma^{(g)}\right)^2 \mathbf{C}^{(g)}\right), \quad \text{for } k = 1, \ldots, \lambda, \tag{4.33}$$

where $\mathbf{x}_k^{(g+1)} \in \mathbb{R}^n$ is the kth offspring at generation $g+1$, and $\mathbf{m}^{(g)} \in \mathbb{R}^n$, $\sigma^{(g)} \in \mathbb{R}^+$, and $\mathbf{C}^{(g)} \in \mathbb{R}^{n \times n}$ are the expected value, overall standard deviation, and the covariance matrix at generation g, respectively.

In practice, a sample of distribution (4.33) can be obtained by the equivalence in distribution [Hansen, 2005]

$$\mathbf{x}_k^{(g+1)} \sim \mathcal{N}\left(\mathbf{m}^{(g)}, \left(\sigma^{(g)}\right)^2 \mathbf{C}^{(g)}\right) \sim \mathbf{m}^{(g)} + \sigma^{(g)}\mathbf{B}^{(g)}\mathbf{D}^{(g)}\mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{4.34}$$

using the eigendecomposition of $\mathbf{C}$,

$$\mathbf{C} = \mathbf{B}\mathbf{D}^2\mathbf{B}^T, \tag{4.35}$$

where $\mathbf{B}$ is an orthonormal matrix of eigenvectors from $\mathbf{C}$ and $\mathbf{D}^2 = \mathbf{D}\,\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$. Hence, a sample of the distribution can be obtained by first sampling a vector from the uncorrelated multivariate normal distribution, which then undergoes a linear transformation. In turn it is multiplied by the global step size and translated by the distribution's expectation.

Selection and recombination are performed as follows,

$$\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu} w_i\, \mathbf{x}_{i:\lambda}^{(g+1)}, \quad \text{with} \sum_{i=1}^{\mu} w_i = 1,\, w_1 \geq w_2 \geq \cdots \geq w_\mu > 0, \tag{4.36}$$

where the new expectation $\mathbf{m}^{(g+1)}$ of the distribution is the weighted average of the μ selected from the λ generated offspring. The positive weight coefficients are denoted by $w_{i=1,\ldots,\mu} \in \mathbb{R}^+$. For $w_{i=1,\ldots,\mu} = 1/\mu$, (4.36) calculates the mean value of the μ selected offspring. $\mathbf{x}_{i:\lambda}^{(g+1)}$ represents the ith best out of the λ individuals.

To update of the covariance matrix two mechanisms are used, the rank-μ-update and the rank-one-update. While the first extracts information from the μ best solutions at a given generation, the latter interprets the development of successive iterations via the *evolution path*. The step size σ is updated also by means of an evolution path method. For explanations of the various constants used within the CMA-ES and reasonable settings, the reader is referred to [Hansen, 2005].

4.9.1 Rank-μ-Update

To get a reliable estimator of the covariance matrix, the variance effective selection mass μ_{eff} must be large enough. However, for the sake of fast search, the population size λ and thus also μ_{eff} must be small. As a remedy, information from previous generations is reused. A reliable estimator then reads

$$\mathbf{C}^{(g+1)} = (1 - c_{\text{cov}})\mathbf{C}^{(g)} + c_{\text{cov}} \sum_{i=1}^{\mu} w_i \,\text{OP}\left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}\right), \tag{4.37}$$

where c_{cov} is a learning rate with $0 < c_{\text{cov}} \leq 1$ that implements exponential smoothing, assigning recent generations a higher weight. The operator OP gives the outer product of a vector with itself. As initial covariance matrix the identity matrix is chosen, $\mathbf{C}^{(0)} = \mathbf{I}$. The name rank-$\mu$-update of this covariance matrix update arises from the fact that the sum of outer products in (4.37) is with probability one of rank $\min(\mu, n)$ [Hansen et al., 2003].

4.9.2 Rank-One-Update

This covariance matrix update mechanism takes a different viewpoint. The matrix is updated repeatedly in every generation using a single selected step only. Therefore the so-called *evolution path* is introduced.

The evolution path is a sequence of successive steps that the strategy has taken over the last generations. Hence, it can be described by a summation of consecutive steps, also referred to as *cumulation*. In the construction of the evolution path, the step size is disregarded. Thus, an evolution path of the last three steps of the distribution's mean $\mathbf{m}$ reads

$$\frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}} + \frac{\mathbf{m}^{(g)} - \mathbf{m}^{(g-1)}}{\sigma^{(g-1)}} + \frac{\mathbf{m}^{(g-1)} - \mathbf{m}^{(g-2)}}{\sigma^{(g-2)}}. \tag{4.38}$$

Using again exponential smoothing as in (4.37), the evolution path can be calculated by

$$\mathbf{p}_c^{(g+1)} = (1 - c_c)\mathbf{p}_c^{(g)} + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\, \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}, \tag{4.39}$$

where $\mathbf{p}_c^{(g)} \in \mathbb{R}^n$ is the evolution path at generation g, $c_c \leq 1$ determines the decay of information from previous steps, and the factor $(c_c(2 - c_c)\mu_{\text{eff}})^{1/2}$ is a normalization constant for $\mathbf{p}_c^{(g+1)}$. The initial evolution path is set to the zero vector, $\mathbf{p}_c^{(0)} = \mathbf{0}$.

The rank-one-update of the covariance matrix $C^{(g+1)}$ via the evolution path $\mathbf{p}_c^{(g+1)}$ then reads

$$\mathbf{C}^{(g+1)} = (1 - c_{\text{cov}})\mathbf{C}^{(g)} + c_{\text{cov}}\, \mathbf{p}_c^{(g+1)} \mathbf{p}_c^{(g+1)^T}. \tag{4.40}$$

4.9.3 Combining Rank-One-Update and Rank-μ-Update

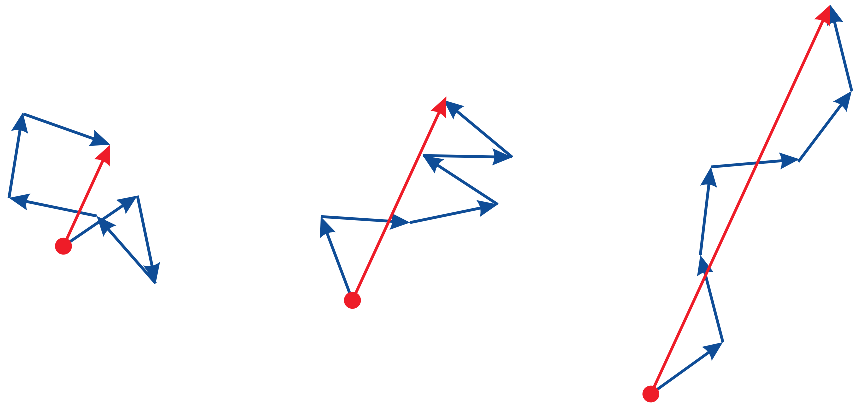
In the final CMA update of the covariance matrix, (4.37) and (4.40) are combined with relative weighting μ_{cov}.

$$\begin{aligned} \mathbf{C}^{(g+1)} = (1 - c_{\text{cov}})\mathbf{C}^{(g)} &+ \frac{c_{\text{cov}}}{\mu_{\text{cov}}} \mathbf{p}_c^{(g+1)} \mathbf{p}_c^{(g+1)^T} \\ &+ c_{\text{cov}} \left(1 - \frac{1}{\mu_{\text{cov}}}\right) \sum_{i=1}^{\mu} w_i \,\mathrm{OP}\left(\frac{\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}\right) \end{aligned} \tag{4.41}$$

4.9.4 Cumulative Step Size Adaptation

There are two reasons to introduce an explicit step size control in addition to the adaptation rule (4.41) for the covariance matrix $\mathbf{C}^{(g+1)}$. First, by (4.41) the optimal overall step size cannot be well approximated and second, the maximal reliable learning rate for the covariance matrix update in (4.41) is too slow to obtain competitive change rates for the overall step size.

The control of the step size $\sigma^{(g)}$ is realized via the evolution path (see Section 4.9.2). In each generation, the evolution path is compared to a path that



(a) Short evolution path (b) Medium evolution path (c) Long evolution path

Figure 4.5: Evolution paths from different typical situations. Each consists of six single steps of comparable length. The overall length of the evolution paths is remarkably different and evaluated for step size control

would result from random selection. Comparing the two paths lengths provides an indication of an efficient mutation strength. This technique is termed *cumulative step size adaptation* (CSA) or *cumulative path length control* and can be used independently of the covariance matrix update. To exploit the length of an evolution path is motivated by the following heuristics:

- If the evolution path is short, the single steps annihilate each other (Figure 4.5(a)). When steps are "anti-correlated", the step size should be decreased.
- If the evolution path is long, the single steps point to similar directions (Figure 4.5(c)). When steps are "correlated", the step size should be increased since the same distance could have been covered by fewer but longer steps.
- If the evolution path is of medium length, the path of selected steps is approximately as long as the expected path length under random selection. As this is the desired situation the step size remains unchanged.

To calculate the evolution path, the same techniques as in (4.39) are applied,

$$\mathbf{p}_\sigma^{(g+1)} = (1 - c_\sigma)\mathbf{p}_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{\text{eff}}}\mathbf{C}^{(g)^{-\frac{1}{2}}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}, \tag{4.42}$$

where $C^{(g)^{-\frac{1}{2}}}$ rescales the step $\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}$ within the coordinate system given by $\mathbf{B}^{(g)}$. The step size update by means of the comparison of the evolution path length to the expected path length under random selection then reads

$$\sigma^{(g+1)} = \sigma^{(g)} \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|\mathbf{p}_\sigma^{(g+1)}\|}{\mathrm{E}\,\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1\right)\right), \tag{4.43}$$

where $d_\sigma \approx 1$ is a damping parameter.

Part II

Directed Evolutionary Algorithms

5 Motivation

Since the early work of Rechenberg [1973] and Schwefel [1974] the design of mutation operators turned out to be one of the most critical points in ES. These early works relied on just one single mutation strength, i.e. step size, for all problem dimensions (isotropic mutation, see subsection 4.7.1) and were concerned mainly with determining the optimal step size for a faster search. To put it in a more general light, besides the one step size, the covariance matrix of the mutation operator's distribution was considered to be the identity matrix. Soon Schwefel extended this approach and proposed to self-adapt one step size per variable, i.e. to use a diagonal covariance matrix with positive entries (scaled mutation, subsection 4.7.2). Consequently, as the most general case, he later suggested self-adapting of the whole covariance matrix (correlated mutation, subsection 4.7.3).

Anyway, all of these methods rely on normally distributed mutations and relatively little effort has been put into examining different distributions as mutation operators. One such example is the so-called Fast Evolution Strategy by Yao and Liu [1997], where a Cauchy distribution is proposed as mutation operator. Nevertheless, Rudolph [1998] later proved that the order of local convergence is identical to that of normal mutations. An example where a Laplace distribution is applied can be found in [Montana and Davis, 1989]. However, just to exchange the mutation distribution seems in general to be a questionable idea.

The intention of directed mutation on the other hand is to introduce a different mutation *principle*. Directed mutation will impart true directionality to the search. This means that for every problem dimension a tendency towards the positive or negative domain can be established by the mutation distribution. Thus, hopefully the mutation distribution will adopt favorable directions over the generations and sustain further advance into it. An example of a directed mutation operator in a two dimensional setting is illustrated in Figure 5.1. The shown distribution tends to favor the positive direction on the x_1 axis and the negative on the x_2 axis, respectively.

Directed mutation will abandon the *random mutation hypothesis* – a fundamental tenet postulating that mutations occur at random, regardless of fitness consequences to the resulting offspring. This seems to be justified by the fact that the ES knows its optimization history and is thus able to extrapolate the evolution path to some extent. Under the assumption of a local similar objective function it is obviously reasonable to generate a bigger portion of offspring along the successful path. In this sense directed mutation is another example of the nonlocal adaptation approaches presented in subsection 4.8.4.

To achieve directed mutation it is necessary to give up the symmetry demand posed in Section 4.7; at least symmetry with respect to the ordinate, if not even

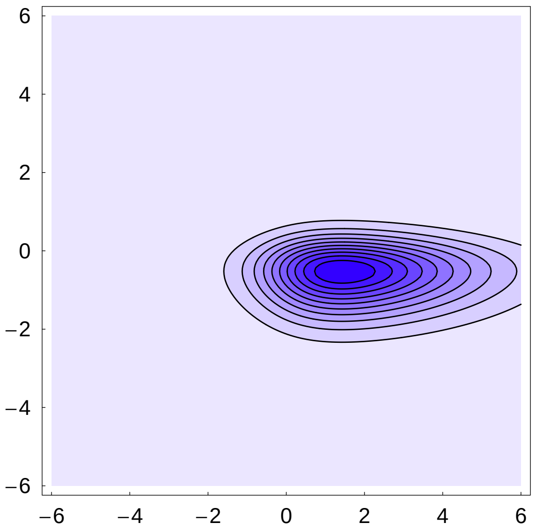
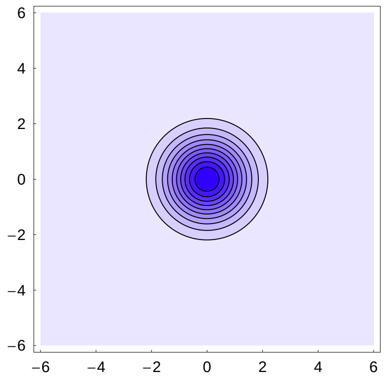


Figure 5.1: Iso-probability contours of a skew distribution

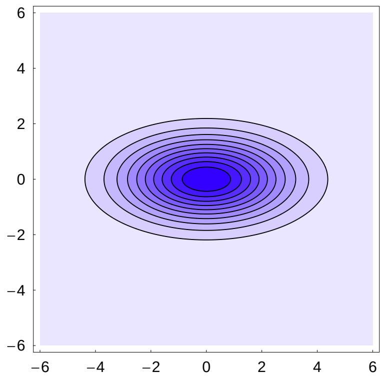
symmetry with respect to the distribution's mean. Stripping away this demand results in a great new flexibility of the mutation operator. Some iso-probability contours of a directed mutation operator with selected parameters are shown in Figure 5.2. It is apparent that the isotropic as well as the scaled mutation are included as marginals. Beyond, arbitrary skew distributions can be realized.

The basis of any directed mutation operator is an appropriate customizable distribution. Usually introducing skewness into any distribution concurs with expectations unequal to zero. There are obviously several methods to reach this goal whereof some are presented in the sequel. Two essential variants are the constructive approach and the skewing function approach.

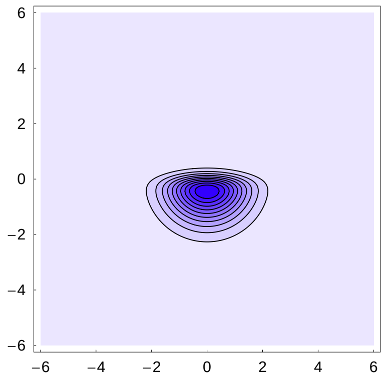
Roughly, this part is organized as follows: in the sequel of this chapter the general principle of directed mutation is rendered, postponing the mathematical and implementational details to the next chapters. Also, several alternative techniques recently found in the literature are subsumed. The next chapter is devoted to the biological foundations and questions that arise from this. In Chapter 7 the first and older constructive approach is presented in detail. Two distinct variants are discussed that differ only slightly in construction but vastly in operation. The next chapter introduces the more substantive skewing function approach with its most prominent member, the skew-normal distribution (see Section 8.4). In Chapter 9 the theoretical fundamentals of directed mutation are utilized. First the realization of directed mutation operators on the basis of the previously presented distributions is discussed. With directed mutation operators being available, then Directed Evolution Strategies as a whole are introduced. The last chapter of this part is devoted to a directed variant of the Covariance Matrix Adaptation-ES. It is shown how this most powerful ES can further be enhanced by the directed mutation principle. Due to its functioning, adaptation of the shape parameters can be realized via an intragenerational update.



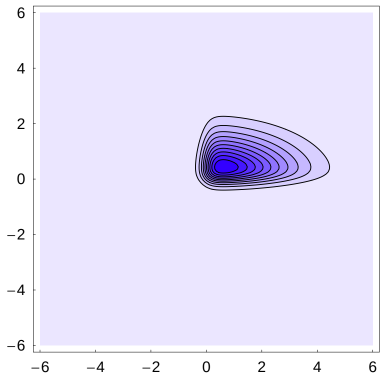
(a) No skewness and equal mutation strengths resulting in isotropic mutation



(b) No skewness and different mutation strengths resulting in scaled mutation



(c) Symmetry with respect to the x_1 axis and negative skewness with respect to the x_2 axis under equal mutation strengths



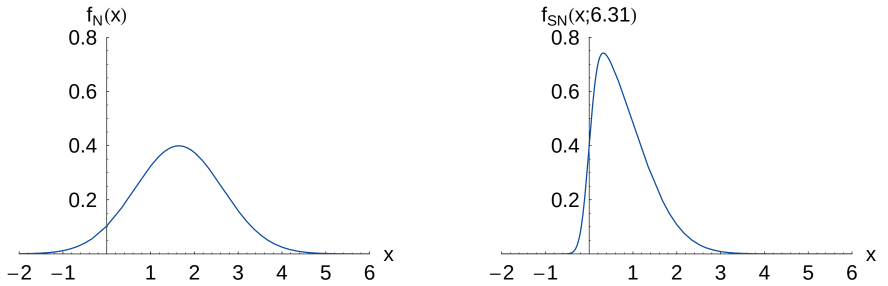
(d) Positive skewness with respect to both, the x_1 and the x_2 axis under different mutation strengths

Figure 5.2: Several iso-probability contours of a skew distribution

5.1 Principle

The main idea of directed mutation is to impart true directionality to the search by generating random numbers that lie preferably in the direction the optimum is expected. This usually implies distributions with expectations unequal to zero, contrasting with conventional mutation operators. Using this method the optimization strategy should adopt most promising directions over the generations.

An obvious approach is to use distributions with non-zero mean. Suppose we want to favor the positive domain intensively, say 95% of the generated random numbers should be positive. Using a non-central normal distribution, the distribution $\mathcal{N}(1.64, 1)$ will obtain this, depicted in Figure 5.3(a).



(a) Non-central normal distribution, $\mathcal{N}(1.64, 1)$

(b) Skew-normal distribution, SN(x; 6.31)

Figure 5.3: Directed mutation by means of a non-central versus a skew distribution; generating 95% positive and 5% negative random numbers

However, the distribution itself is still symmetric. As it is merely translated by some amount, the random numbers are in fact sampled around a new location. The major appearance of this approach is surely the translation rather than the directionality. Since the aim is to enhance the frequency of positive numbers but not their expectation this strategy seems not to be very useful. Further interpreting of the pdf in Figure 5.3(a) reveals another undesired behavior. Ideally, negative and positive random numbers are scattered around the origin with different frequencies but both with declining probability for increasing distance. In the example this is fulfilled only in the negative domain while in the positive domain the probability of creation is at first even increasing with increasing distance. In general, to adapt variable means is a different concept, for example tackled by the Covariance Matrix Adaptation presented in subsection 4.9.

An alternative to generate positive random numbers with higher frequency is to use skew distributions as illustrated in Figure 5.3(b). With respect to the previous explanations, this pdf is apparently more sensible. Also here 95% of the sampled random numbers will be positive, but the distribution's expectation is with $\mathrm{E}\,X = 0.78$ not even half the expectation of the translated variant. Thus the change of location is comparably small.

To be able to tune the proportion of positive to negative random numbers customizable skew distributions are needed. These are the basis of the directed mutation operators presented in the sequel of this part.

5.1.1 On the Moments and their Convergence

As described in Section 3.6, moments are statistics that give an indication of several characteristics of a distribution. Since the distributions to be tackled next are parameterized, the actual value of the moments usually depends on these control parameters. Under this condition the concept of the moments' convergence needs to be generalized. Of interest is now the convergence of the moments with respect to the control parameters. Even if the moments are convergent for an arbitrary fixed parameter setting, this need not be the case when the parameters tend to infinity. Thus, convergence is not only demanded for a particular setting but rather for all possible modulations, including the limit cases.

It has now repeatedly been intimated that true directionality usually concurs with expectations unequal to zero[1]. First of all this means that the mutation operator is not compliant to the conventional evolution strategy any longer postulating expected values of zero, thus maintaining the current position under the absence of selection (cf. the symmetry demand in Section 4.7). Anyway, more important is to ensure the expectation's convergence, i.e. forcing the mutation operator to continue mutating *near by* the current position. As argued before, this has to be guaranteed for all degrees of skewness. Mutation operators based on the Ξ-distributions proposed by Hildebrand [2001] violate this demand. Diverging expected values for increasing skewness parameters can result in wide jumps in the search space.

More important is the second moment, the variance. Since it can be seen as a measure of the mutation strength it is a strategy parameter in most evolution strategies, cf. the subsections *Isotropic Mutation*, *Scaled Mutation*, and *Correlated Mutation* on page 48ff. Because of this it should not implicitly be modified by the skewness parameter. In the ideal case the variance is shape invariant. At least convergence is necessary and a small spread between its minimum and maximum is desired to limit the impact of the skewness on the mutation strength. Again, the Ξ-distribution violates this demand.

5.2 Alternative Approaches

Besides the directed mutation operators resulting from the two major construction approaches explained in the Chapters 7 and 8, several other forms can be found in the literature. Three of these are in advance recapitulated.

[1] In fact it is possible to design skew distributions with vanishing first two moments for arbitrary skewness. However, these distributions look quite strange and are of no relevance for directed mutation.

5.2.1 Adaptation of a Non-Central Normal Distribution

As stated above, an obvious approach to impart directionality to the search is to use symmetric distributions with non-zero mean, i.e. translated distributions. Examples for this can be found in [Ostermeier, 1992] or [Voigt and Anheyer, 1994], where the expectation vector is used as additional strategy parameter. For a dynamic context this approach has been proposed by Weicker [2001]. The task there is not only to optimize a given function, but also to track the moving optimal region. In an early paper this task has been identified to be difficult for the usual strategies [Weicker and Weicker, 1999].

Therefore, an offset vector v is introduced which determines the new position of an offspring. The mutation of an individual then reads

$$\sigma = \sigma \, \exp(n^{-1/2}\mathcal{N}(0,1)) \tag{5.1}$$
$$v_i = v_i + \mathcal{N}(0,1) \tag{5.2}$$
$$x_i = x_i + v_i + \mathcal{N}(0,\sigma). \tag{5.3}$$

This mutation scheme is applied to two dynamic test functions, the *moving circle problem* and the *moving corridor problem*. Both realize a tracking region with “good” fitness while all other points are assigned to a “bad” fitness. The fitness within the tracking region is determined by the distance to the optimum.

As the focus lay on dynamic problems, the results are not discussed in detail here. However, it can be concluded from the experiments that ES with either isotropic mutation or scaled mutation exhibit a substantially increasing number of invalid individuals with increasing problem dynamics.

5.2.2 Directed Mutations by Means of Direction Vectors

Ghozeil and Fogel [1996] proposed to shift the representations of the individuals from Cartesian to polar coordinates. Offspring then can be generated by adding a direction vector with some given step size, (r, θ), to the parental individual. For the mutation, independent step sizes σ_r and σ_θ are used. Adaptation of σ_r is done in the usual self adaptive way by

$$\sigma_r' = \sigma_r \exp(\mathcal{N}(0,1)). \tag{5.4}$$

Mutation on the direction is done as follows: the direction θ is defined as an n-dimensional vector in Cartesian coordinates which is perturbed by an n-dimensional normally distributed random vector with zero mean and self adaptive standard deviation σ_θ. The result is normalized again to give the new direction vector. Ghozeil and Fogel [1996] also investigated a variant where r was set equal to σ_r and applied both methods to four test functions.

Their directed mutation outperformed the scaled mutation with standard deviation $\sigma = 1.224(f(\mathbf{x}))^{1/2}/n$ only on the Rosenbrock function but performed worse on the other three functions. However, compared to a scaled mutation with self adaptive step sizes it performed slightly better on all four functions.

5.2.3 Polymorphic Mutation

Anticipating some aspects of the skewing function approach (see Chapter 8), the polymorphic mutation was the first directed mutation operator with a skew distribution given in closed form [Berlik, 2003a,b]. Its pdf is defined by

$$f_{\mathrm{PM}}(x;\lambda) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \, \frac{1}{1+e^{-\lambda x}} \, e^{-\frac{x^2}{2\sigma^2}}, \quad \forall\, x, \lambda \in \mathbb{R}, \tag{5.5}$$

where the three functional parts in (5.5) are the well known kernel of the normal distribution on the right, a term introducing the skewness in the middle, and a normalization factor on the left.

Most interesting of course is the term in the middle. It is a sigmoid function that enforces one side of the entire distribution and attenuates at the same time the other one. In fact, it is a distribution itself, namely the logistic distribution (3.23). By the real-valued parameter λ the intensity of skewness can be adjusted. For some values of λ density functions are plotted in Figure 5.4. The normal distribution is contained as a marginal for $\lambda = 0$, meaning symmetry.

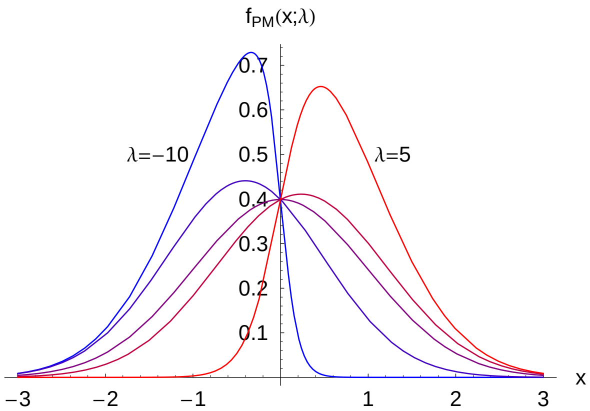


Figure 5.4: Probability density functions $f_{\mathrm{PM}}(x;\lambda)$ of the polymorphic distribution with skewness parameter $\lambda \in \{-10, -1, 0, 1/2, 5\}$

However, although the function looks quite sweet-tempered and the logistic distribution is the simplest conceivable continuous sigmoid function (confer Section 8.2), it is mathematically difficult to handle. For example, the function cannot be integrated analytically and the moments cannot be calculated. Furthermore, no efficient method for random variate generation is known, such that it disqualifies as mutation distribution and is not further investigated.

6 Biological Evolution

> *"If under changing conditions of life organic beings present individual differences in almost every part of their structure, and this cannot be disputed; if there be, owing to their geometrical rate of increase, a severe struggle for life at some age, season or year, and this certainly cannot be disputed; then, considering the infinite complexity of the relations of all organic beings to each other and to their conditions of life, causing an infinite diversity in structure, constitution, and habits, to be advantageous to them, it would be a most extraordinary fact if no variations had ever occurred useful to each being's own welfare, in the same manner as so many variations have occurred useful to man. But if variations useful to any organic being ever do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance, these will tend to produce offspring similarly characterised. This principle of preservation, or the survival of the fittest, I have called natural selection. It leads to the improvement of each creature in relation to its organic and inorganic conditions of life; and consequently, in most cases, to what must be regarded as an advance in organisation. Nevertheless, low and simple forms will long endure if well fitted for their simple conditions of life."*

On the Origin of Species, Charles Darwin [1859]

It is now almost 150 years ago that Darwin originally published these revolutionary ideas. When they were controversial discussed first, nowadays they are widely accepted and found their way from the scientific community to society in general. Darwin clearly anticipated the importance of his paradigm in the field of species of organic nature but realized little of the consequences it would cause besides.

> *"A grand and almost untrodden field of inquiry will be opened, on the causes and laws of variation, on correlation, on the effects of use and disuse, on the direct action of external conditions, and so forth."*

On the Origin of Species, Charles Darwin [1859]

The applicability of the paradigm beyond the field of genetics in more abstract constructs turned out later. Its worth in an engineering context was not realized until the latter half of the 20th century. Evolution inspired algorithms for optimization and machine learning were e.g. proposed by Box [1957], Box and Draper [1969], Friedman [1959], and Bremermann [1962].

In this chapter several aspects concerning biology are treated. Therefore we start with a short survey on the field. The presented material is compiled from von Sengbusch [1989] and Futuyma [1998, 2005]. Then, the controversy of directed mutation in biology is outlined and some thoughts about the term *directed mutation* are sketched.

6.1 Survey

Evolution is the process that created all the great variety of organisms. Mankind recognized rather early that some kind of heredity of features exists, while the connection between heredity and evolution was not recognized until much later. Till the last century the belief lasted that the variety of species existed right from the beginning of life and that the features of species are constant. Nowadays, it is regarded as assured that each species has developed from more primitive predecessors, that complex systems evolved from simpler ones, and that the adaptation of organisms is continuously improved.

By and by proofs for changes during earth history accumulated and it became evident that the formation of living things has to be seen as a historical process. Different attempts to explain the mechanisms and reasons for these changes were proposed. It was the *theory of selection* by Charles Darwin [1859] that gave an interpretation that is today - after decades of controversial discussions - regarded as a fact. It has to be viewed as one of the most important foundations of general biology. In short, it states the following:

1. Species are mutable. They developed to its current expression in a continuous succession of generations beginning with the origin of life.

2. Individuals of one species differ from one another. Within a species each feature can be found in considerable variations.

3. Each individual undergoes a natural selection. Only those who are adapted best to their environment have a chance to survive and to propagate.

As the two equally important causes of evolution are seen changes and restructuring of genetic information by means of mutations and recombination on the one hand, and selection on the other. In contrast to selection, mutations are seen as non-directional events in evolution. Regarding evolution as the sum of subtle subsequent steps, the directed tendencies can be identified. However, selection is not oriented on the future. It is rather oriented on a given present state. Structures and functions that have no advantage in the present situation are not developed any further and can actually be lost. Hence, characteristics that could be advantageously in the future have no selection advantage.

Evolution results in an accumulation of valuable, i.e. effectively used genetic information causing an increasing complexity of advantageous structures and an improved success in reproduction, in other words an increased fitness. Such tendencies are regarded as an advance. Advance is always based on some change, but not every change is advantageous. Tendencies resulting in a poorer fitness are called regression. Evolution is an opportunistic principle, with only a very small percentage of changes actually leading to improvements. Whether the particular change is of advantage depends on the respective situation. Better or of advantage in this context means more efficient, more often or more complex than before. With the wide variety of different environmental factors many different trends of evolution emerged, resulting in a diversification of organisms as a consequence.

Since quite a long time now, research into evolution is not concerned any longer with the question whether evolution did take place, but is interested instead in the details of how evolution occurred. Since it is mainly dependent on circumstantial evidence like fossils, the exact time of events can never be specified with absolute certainty. Taking into account all existing evidence, it is however possible to deduce the most probable sequence of events. On the other hand, the physical and chemical laws have most likely always remained the same and it is reasonable to assume that the diversity of environmental influences was formerly not larger than it is today. With leaving the period of the origin of living systems aside, we can presume that inheritance follows rules that can also be elucidated by studying living organisms. Thus partial developments can today, too, be observed or understood by experiments. Evolution is a continuous process, what has been illustrated by numerous examples. It is the mainspring of today's great variety of organisms.

6.2 Directed Mutation in Biology

The term *directed mutation* has a slightly different meaning in the field of biology than that presented in the previous chapter where we emphasized the directional aspect of directed mutation in the scenario of optimization. By means of distributions with adjustable skewness a *tendency* has been introduced into the mutation operator. In contrast to that in the biological setting the attention is focused primarily on whether advantageous mutations occur depending on the environment or not. Directed mutation thus assumes that the occurrence rate of mutation is *not* independent of the fitness that a particular one confers to the mutant.

After the famous experiments of Luria and Delbrück [1943] and later Lederberg and Lederberg [1951] it was until some years ago seen as a fundamental tenet of modern evolutionary theory that adaptively directed mutations do not occur [Futuyma, 1998]. However, recently new evidence is reported supporting the directed mutation hypothesis (see e.g. [Cairns et al., 1988]). Anyway, since the controversy is hard to follow for non-microbiologist, we will not unfold the theme any further. For some mathematical consequences arising from the controversy see [Zheng, 2003].

6.3 About the Term "Directed Mutation"

There are two aspects concerning the naming of directed mutation that should further be explained. The one is the naming conflict with the asymmetric mutation, the other the connotation the term incurred in the area of natural evolution.

We prefer to use the term *directed* rather then *asymmetric* to designate the mutation type. First, the term *directed* better describes the phenomenological view on the mutation. Usually we are more interested in this then the intrinsic property of the underlying distribution – that without doubt is asymmetric. Second, if really the characteristic of the distribution should be described, it would be more accurate to speak of *skew mutations*, as the third moment is commonly referred to as skewness, not asymmetry.

Besides the question whether directed mutation occurs in biology or not (recall the previous section), the connotation the term recently incurred should be mentioned. Some people impute that directed mutations indicate the impact of an intelligent guiding instance in the evolutionary process. We are not concerned with this debate and can think of directed mutation just as a tool evaluating non-local search space information.

7 Constructive Approach

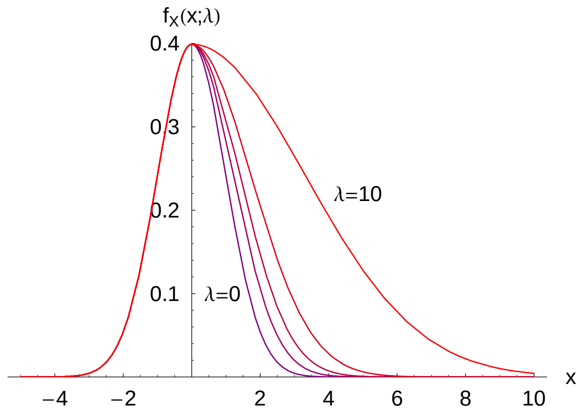
This chapter is dedicated to the first kind of skew distributions utilized in evolution strategies. The guiding principle is to use a piecewise defined function whereof one half equals the normal distribution and the other half is scaled to some extent with respect to the abscissa. The integral of the complete function is subsequently normed to one again. This construction principle, in more detail presented in Section 7.1, traces back to Hildebrand [1996]. He used it to build several skew distributions for his *asymmetric mutation*. Two of them are elucidated in what follows. In Section 7.2 the $\Xi_{c,\sigma}$ distribution is presented. An extended version, the $\Xi^{\gamma}_{c,\sigma}$ distribution, is discussed in Section 7.3. As these distributions have some immanent drawbacks, with the Naïve Skew-Normal distribution an alternative using the same construction principle is developed in Section 7.4.

7.1 Construction Principle

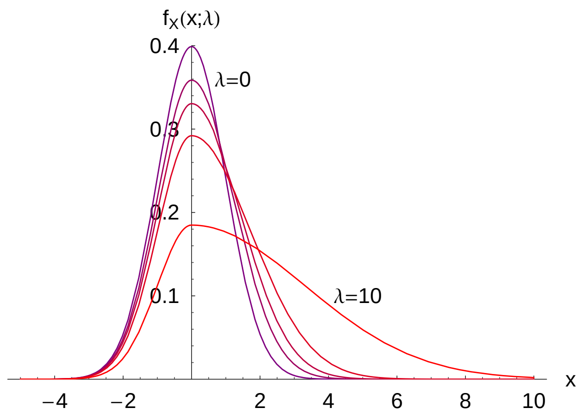
All skew distribution functions presented in this chapter have the construction principle in common: the functions are piecewise defined. Dedicated functions are applied for the negative and positive semi-infinite support. As one part, the corresponding piece of the normal distribution is chosen, the other part is defined by a scaled variant of the normal distribution with respect to the abscissa, illustrated in Figure 7.1(a). The integral of the complete function has subsequently to be normed to one again, see Figure 7.1(b). The extension of the scaling, i.e. the grade of skewness, is determined by a control parameter λ, where a positive (negative) skewness parameter λ results in positive (negative) skewness of the distribution. For $\lambda = 0$, the distribution as a whole gets back to the normal distribution.

Depending on whether positive or negative skewed distributions are desired, the left or right half of the normal distribution has to be replaced by the scaled normal distribution. Thus, in sum four cases have to be distinguished in the definition of the distribution.

The aim of every distribution function used in mutation operators is to sample random variables from it. There are several possibilities to do so as discussed for example in [Devroye, 1986] or [Hörmann et al., 2004]. The approach used to generate random numbers of the distributions presented in this chapter is called *inversion method*, see subsection 3.8. Therefore the inverse of the distribution function has to be known explicitly.



(a) Non-normed interim result



(b) Normed variant

Figure 7.1: Construction principle of skewed distributions with skewness parameter $\lambda \in \{0, 1/2, 1, 2, 10\}$

7.2 $\Xi_{c,\sigma}$ Distribution

This section introduces the most relevant distribution to Hildebrand's asymmetric mutation, the $\Xi_{c,\sigma}$ distribution [Hildebrand, 2001]. The parameter controlling the skewness here is denoted by c. The second parameter σ regulates the mutation strength. Presented in the following subsections are the density (7.2.1), distribution (7.2.2), inverse of the distribution with random variate generation (7.2.3 and 7.2.4), and the moments (7.2.5).

7.2.1 Density

The probability density function of a $\Xi_{c,\sigma}$ distributed random variable X is given by

$$\xi_{c,\sigma}(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma}\left(1+\sqrt{1-c}\right)}\, e^{-\frac{x^2}{2\sigma(1-c)}} & \text{if } c < 0, x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma}\left(1+\sqrt{1-c}\right)}\, e^{-\frac{x^2}{2\sigma}} & \text{if } c < 0, x \geq 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma}\left(1+\sqrt{1+c}\right)}\, e^{-\frac{x^2}{2\sigma}} & \text{if } c \geq 0, x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma}\left(1+\sqrt{1+c}\right)}\, e^{-\frac{x^2}{2\sigma(1+c)}} & \text{if } c \geq 0, x \geq 0. \end{cases} \tag{7.1}$$

Some graphs are depicted in Figure 7.2. Skewness is introduced by stretching one function half a time. This will impact the moments, shown in subsection 7.2.5. Note the way the mutation strength σ is incorporated. As the variable σ is not squared in the exponent of the terms $e^{-x^2/2\sigma}$, it takes the role of a variance. This can easily lead to some confusion, cf. in particular the definition of the variance in (7.4).

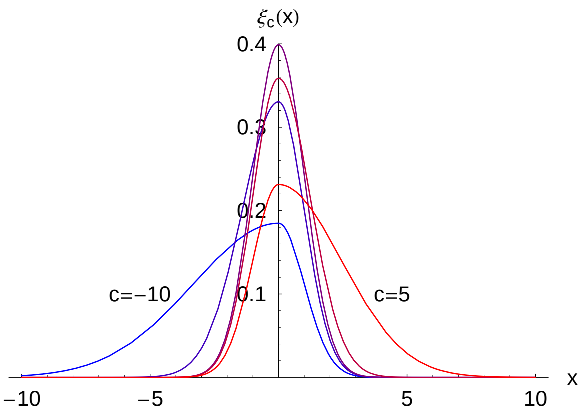


Figure 7.2: Probability density functions $\xi_{c,\sigma}(x)$ of the $\Xi_{c,\sigma}$ distribution with skewness parameter $c \in \{-10, -1, 0, 1/2, 5\}$ and $\sigma = 1$

7.2.2 Distribution

The definition of the cumulative distribution function of a $\Xi_{c,\sigma}$ distributed random variable X reads

$$\Xi_{c,\sigma}(x) = \begin{cases} \frac{\sqrt{1-c}}{1+\sqrt{1-c}} \left(1 + \operatorname{erf} \frac{x}{\sqrt{2\sigma(1-c)}}\right) & \text{if } c < 0, x < 0 \\ \frac{1}{1+\sqrt{1-c}} \left(\sqrt{1-c} + \operatorname{erf} \frac{x}{\sqrt{2\sigma}}\right) & \text{if } c < 0, x \geq 0 \\ \frac{1}{1+\sqrt{1+c}} \left(1 + \operatorname{erf} \frac{x}{\sqrt{2\sigma}}\right) & \text{if } c \geq 0, x < 0 \\ \frac{1}{1+\sqrt{1+c}} \left(1 + \sqrt{1+c} \operatorname{erf} \frac{x}{\sqrt{2\sigma(1+c)}}\right) & \text{if } c \geq 0, x \geq 0, \end{cases} \tag{7.2}$$

where erf(·) denotes the error function, see Appendix B. Some graphs are illustrated in Figure 7.3.

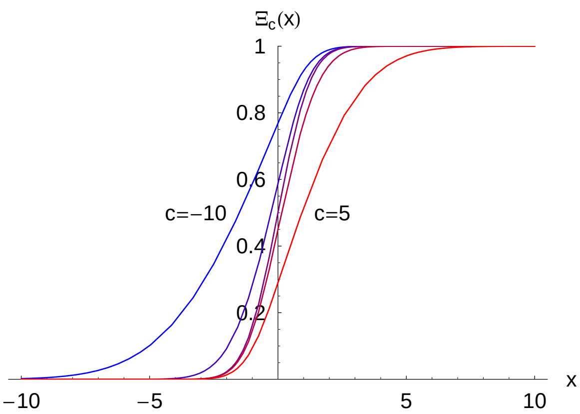


Figure 7.3: Cumulative distribution functions of the $\Xi_{c,\sigma}$ distribution with skewness parameter $c \in \{-10, -1, 0, 1/2, 5\}$ and $\sigma = 1$

7.2.3 Inverse of the Distribution

Generating $\Xi_{c,\sigma}$ distributed random variables is done using the inversion method. Hence, the inverse of the distribution function is needed that is given by Hildebrand [2001] as

$$\bar{\Xi}_{c,\sigma}(y) = \begin{cases} \sqrt{2\sigma(1-c)} \operatorname{erf}^{-1} \left(y\left(1 + \frac{1}{\sqrt{1-c}}\right) - 1\right) & \text{if } c < 0, y < \frac{\sqrt{1-c}}{1+\sqrt{1-c}} \\ \sqrt{2\sigma} \operatorname{erf}^{-1} \left(y\left(1 + \sqrt{1-c}\right) - \sqrt{1-c}\right) & \text{if } c < 0, y \geq \frac{\sqrt{1-c}}{1+\sqrt{1-c}} \\ \sqrt{2\sigma} \operatorname{erf}^{-1} \left(y\left(1 + \sqrt{1+c}\right) - 1\right) & \text{if } c \geq 0, y < \frac{1}{1+\sqrt{1+c}} \\ \sqrt{2\sigma(1+c)} \operatorname{erf}^{-1} \left(y + \frac{y-1}{\sqrt{1+c}}\right) & \text{if } c \geq 0, y \geq \frac{1}{1+\sqrt{1+c}}, \end{cases} \tag{7.3}$$

where $\operatorname{erf}^{-1}(\cdot)$ denotes the inverse error function.

7.2.4 Random Variate Generation

As already stated, random numbers of the $\Xi_{c,\sigma}$ distribution are generated using the inversion method described in Section 3.8. According to Theorem 3.12, uniform distributed random numbers have to be multiplied with the $\Xi_{c,\sigma}$ distribution's inverse (7.3). Note that this method is demanding and cumbersome. Two case differentiations, one calculation of the transcendent inverse error function (see Appendix B.2), and several arithmetic operations are necessary to generate a single $\Xi_{c,\sigma}$ distributed random variable.

7.2.5 Moments

This subsection provides formulas of the moments of a $\Xi_{c,\sigma}$ distributed random variable. Regard the fat tails in Figure 7.2 which already indicate diverging moments as $c \rightarrow \pm\infty$.

Expectation The formula for the expected value of the $\Xi_{c,\sigma}$ distribution takes the form

$$\mathrm{E}(X) = \frac{c\sqrt{2\sigma}}{\sqrt{\pi}\left(1 + \sqrt{1 + |c|}\right)}, \tag{7.4}$$

depicted in Figure 7.4.

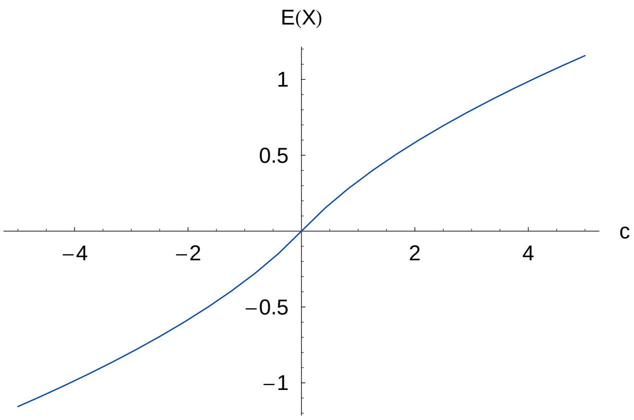


Figure 7.4: Expectation of a $\Xi_{c,1}$ distributed random variable vs. skewness parameter

The limits are

$$\lim_{c \rightarrow -\infty} \big(\mathrm{E}(X)\big) = -\infty \tag{7.5a}$$

$$\lim_{c \rightarrow \infty} \big(\mathrm{E}(X)\big) = \infty. \tag{7.5b}$$

Variance The variance of the $\Xi_{c,\sigma}$ distribution is given by

$$\mathrm{Var}(X) = \sigma\left(\left(2-\sqrt{1+|c|}+|c|\right)-\frac{2c^2}{\pi\left(1+\sqrt{1+|c|}\right)^2}\right). \tag{7.6}$$

Keep in mind that using the given definitions the variance is proportional to σ, $\mathrm{Var}(X) \propto \sigma$, and not to σ^2. The standard deviation is thus proportional to the unusual term $\sqrt{\sigma}$. A graph of the variance is depicted in Figure 7.5.

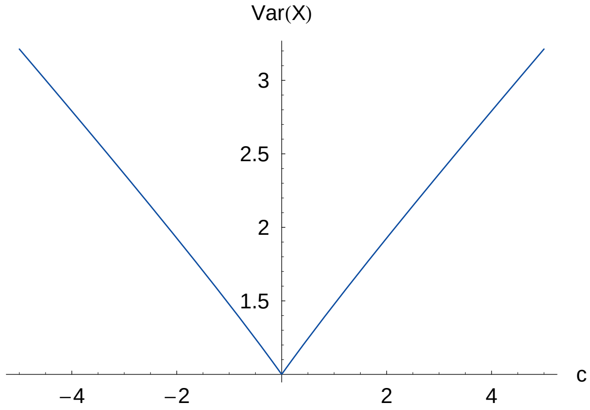


Figure 7.5: Variance of a $\Xi_{c,1}$ distributed random variable vs. skewness parameter

The limits are

$$\lim_{c\to-\infty}\bigl(\mathrm{Var}(X)\bigr) = \infty \tag{7.7a}$$

$$\lim_{c\to\infty}\bigl(\mathrm{Var}(X)\bigr) = \infty. \tag{7.7b}$$

Skewness The skewness of the $\Xi_{c,\sigma}$ distribution reads

$$\gamma_1(X) = \begin{cases} -\frac{\sqrt{2}c\left(-4c^2+\pi\left(\sqrt{1-c}(c-2)+c^2+2c-2\right)\right)}{\left(1+\sqrt{1-c}\right)^3\left(-4-\sqrt{1-c}(\pi-4)-\pi c+2c+2\pi\right)^{3/2}} & \text{if } c<0 \\ \frac{\sqrt{2}c\left(4c^2+\pi\left(\sqrt{1+c}(c+2)-c^2+2c+2\right)\right)}{\left(1+\sqrt{1+c}\right)^3\left(-4-\sqrt{1+c}(\pi-4)+\pi c-2c+2\pi\right)^{3/2}} & \text{if } c\geq 0 \end{cases} \tag{7.8}$$

and is illustrated in Figure 7.6. The limits are

$$\lim_{c\to-\infty}\bigl(\gamma_1(X)\bigr) = -\frac{\sqrt{2}\,(4-\pi)}{(\pi-2)^{3/2}} \tag{7.9a}$$

$$\lim_{c\to\infty}\bigl(\gamma_1(X)\bigr) = \frac{\sqrt{2}\,(4-\pi)}{(\pi-2)^{3/2}}. \tag{7.9b}$$

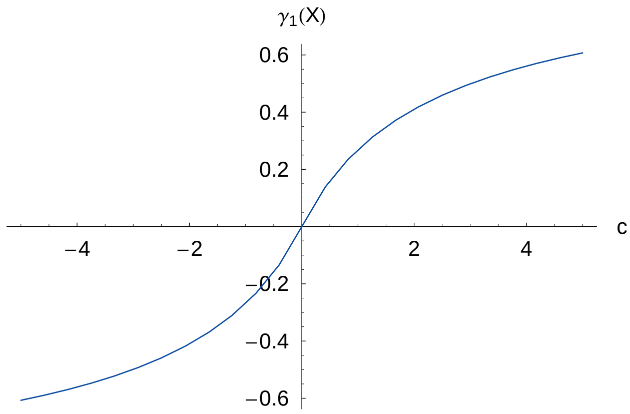


Figure 7.6: Skewness of a $\Xi_{c,1}$ distributed random variable vs. skewness parameter

7.3 $\Xi^{\gamma}_{c,\sigma}$ Distribution

Hildebrand [2001] also proposed another skew distribution that extends the $\Xi_{c,\sigma}$ variant by a further control parameter γ. This parameter is an exogenous strategy parameter that affects the variance. As the $\Xi^{\gamma}_{c,\sigma}$ distribution offers no principally new properties we present only its density and distribution.

7.3.1 Density

The probability density function of the $\Xi^{\gamma}_{c,\sigma}$ distribution reads

$$\xi^{\gamma}_{c,\sigma}(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma^{\gamma}}\left(1+(1-c)^{\gamma}\right)}\, e^{-\frac{x^2}{2(\sigma(1-c))^{\gamma}}} & \text{if } c < 0, x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^{\gamma}}\left(1+(1-c)^{\gamma}\right)}\, e^{-\frac{x^2}{2\sigma^{\gamma}}} & \text{if } c < 0, x \geq 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^{\gamma}}\left(1+(1+c)^{\gamma}\right)}\, e^{-\frac{x^2}{2\sigma^{\gamma}}} & \text{if } c \geq 0, x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^{\gamma}}\left(1+(1+c)^{\gamma}\right)}\, e^{-\frac{x^2}{2(\sigma(1+c))^{\gamma}}} & \text{if } c \geq 0, x \geq 0. \end{cases} \tag{7.10}$$

7.3.2 Distribution

The cumulative distribution function of the $\Xi^{\gamma}_{c,\sigma}$ distribution is given by

$$\Xi^{\gamma}_{c,\sigma}(x) = \begin{cases} \frac{\sqrt{(1-c)^{\gamma}}}{1+\sqrt{(1-c)^{\gamma}}} \left(1 + \operatorname{erf} \frac{x}{\sqrt{2(\sigma(1-c))^{\gamma}}}\right) & \text{if } c < 0, x < 0 \\ \frac{1}{1+\sqrt{(1-c)^{\gamma}}} \left(\sqrt{(1-c)^{\gamma}} + \operatorname{erf} \frac{x}{\sqrt{2\sigma^{\gamma}}}\right) & \text{if } c < 0, x \geq 0 \\ \frac{1}{1+\sqrt{(1+c)^{\gamma}}} \left(1 + \operatorname{erf} \frac{x}{\sqrt{2\sigma^{\gamma}}}\right) & \text{if } c \geq 0, x < 0 \\ \frac{1}{1+\sqrt{(1+c)^{\gamma}}} \left(1 + \sqrt{(1+c)^{\gamma}} \operatorname{erf} \frac{x}{\sqrt{2(\sigma(1+c))^{\gamma}}}\right) & \text{if } c \geq 0, x \geq 0. \end{cases} \tag{7.11}$$

7.4 Naïve Skew-Normal Distribution

The naïve skew-normal (NSN) distribution is built similar to the $\Xi_{c,\sigma}$ distribution [Berlik, 2004a]. The main difference lies in the scaling operation. One half of the normal density is compressed instead of being expanded. This avoids fat tails (see Figure 7.8) and guarantees convergent expectation (7.22) and variance (7.25). Also, the variance σ^2 occurs with its "correct" exponent in the density function. Since the mean of the distribution is always set to zero when used as mutation function, it is omitted in the following definitions. As is true for the Ξ distribution, also the NSN distribution is symmetric in the case $\lambda = 0$ and then meets the normal distribution. The NSN distribution is termed as it is in allusion to the skew-normal distribution, reflecting the ingenuous construction principle.

7.4.1 Kernel Function

The distribution function's active kernel is a compressed part of the normal distribution. To get negative skewness for negative skewness parameters λ, the active kernel has to be located on the positive domain. The fix part on the negative domain then becomes the (constant) fat tail. Hence, the active kernel takes the form

$$f(x;\lambda) = e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} \quad \text{if } \lambda < 0, x \geq 0. \tag{7.12}$$

For positive skewness an analog argumentation holds,

$$f(x;\lambda) = e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} \quad \text{if } \lambda \geq 0, x < 0. \tag{7.13}$$

Supplemented by the fix parts the interim "distribution" reads

$$f(x;\lambda) = \begin{cases} e^{-\frac{x^2}{2\sigma^2}} & \text{if } \lambda < 0, x < 0 \\ e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} & \text{if } \lambda < 0, x \geq 0 \\ e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} & \text{if } \lambda \geq 0, x < 0 \\ e^{-\frac{x^2}{2\sigma^2}} & \text{if } \lambda \geq 0, x \geq 0, \end{cases} \tag{7.14}$$

illustrated in Figure 7.7.

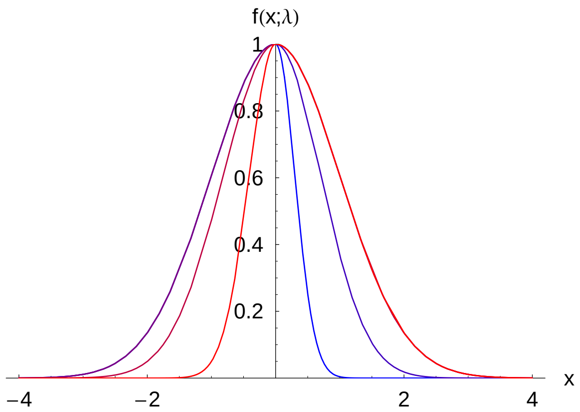


Figure 7.7: Kernel of the naïve skew-normal distribution

To transform (7.14) into a probability density function its integral has to be normalized to one. Therefore the area depending on λ is calculated

$$\begin{aligned} \text{for } \lambda < 0: \quad \int_{-\infty}^{\infty} f(x;\lambda)dx &= \int_{-\infty}^{0} e^{-\frac{x^2}{2\sigma^2}}dx + \int_{0}^{\infty} e^{-\frac{(1-\lambda)x^2}{2\sigma^2}}dx \\ &= \sqrt{\frac{\pi}{2}}\sigma + \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{1-\lambda}}\sigma \\ &= \sqrt{\frac{\pi}{2}}\frac{1+\sqrt{1-\lambda}}{\sqrt{1-\lambda}}\sigma, \end{aligned} \tag{7.15a}$$

$$\begin{aligned} \text{and for } \lambda \geq 0: \quad \int_{-\infty}^{\infty} f(x;\lambda)dx &= \int_{-\infty}^{0} e^{-\frac{(1+\lambda)x^2}{2\sigma^2}}dx + \int_{0}^{\infty} e^{-\frac{x^2}{2\sigma^2}}dx \\ &= \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{1+\lambda}}\sigma + \sqrt{\frac{\pi}{2}}\sigma \\ &= \sqrt{\frac{\pi}{2}}\frac{1+\sqrt{1+\lambda}}{\sqrt{1+\lambda}}\sigma. \end{aligned} \tag{7.15b}$$

7.4.2 Density

With the above results (7.15a) and (7.15b) the density function of the NSN distribution can be defined.

Definition 7.1. A random variable X is said to be *naïve skew-normal* with skewness (or shape) parameter λ, $\lambda \in \mathbb{R}$, written $X \sim \mathcal{NSN}(\lambda)$, if its probability density function is

$$f_{\mathrm{NSN}}(x;\lambda) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma}\, e^{-\frac{x^2}{2\sigma^2}} & \text{if } \lambda < 0, x < 0 \\ \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma}\, e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} & \text{if } \lambda < 0, x \geq 0 \\ \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma}\, e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} & \text{if } \lambda \geq 0, x < 0 \\ \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma}\, e^{-\frac{x^2}{2\sigma^2}} & \text{if } \lambda \geq 0, x \geq 0. \end{cases} \tag{7.16}$$

Graphs of the NSN density for several degrees of skewness are shown in Figure 7.8.

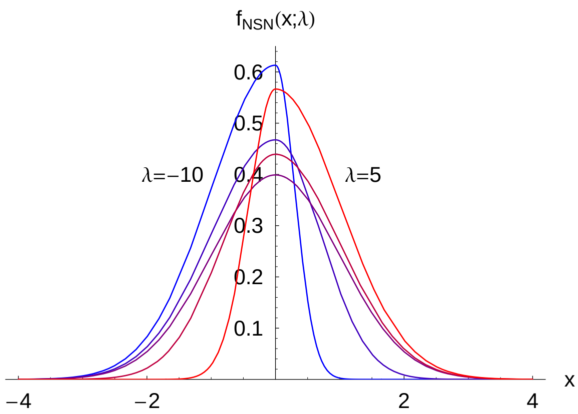


Figure 7.8: Probability density functions $f_{\mathrm{NSN}}(x;\lambda)$ of the naïve skew-normal distribution with skewness parameter $\lambda \in \{-10, -1, 0, 1/2, 5\}$

7.4.3 Distribution

To obtain the cumulative distribution function of the naïve skew-normal distribution the density given in (7.16) has to be integrated. Again, four separate cases are treated resulting in:

$$\text{If } \lambda < 0, x < 0: \quad \int f_{\text{NSN}}(x;\lambda)dx = \frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}} \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma} + c_1, \tag{7.17a}$$

$$\text{If } \lambda < 0, x \geq 0: \quad \int f_{\text{NSN}}(x;\lambda)dx = \frac{1}{1+\sqrt{1-\lambda}} \operatorname{erf} \frac{\sqrt{1-\lambda}\,x}{\sqrt{2}\,\sigma} + c_2, \tag{7.17b}$$

$$\text{If } \lambda \geq 0, x < 0: \quad \int f_{\text{NSN}}(x;\lambda)dx = \frac{1}{1+\sqrt{1+\lambda}} \operatorname{erf} \frac{\sqrt{1+\lambda}\,x}{\sqrt{2}\,\sigma} + c_3, \tag{7.17c}$$

$$\text{If } \lambda \geq 0, x \geq 0: \quad \int f_{\text{NSN}}(x;\lambda)dx = \frac{\sqrt{1+\lambda}}{1+\sqrt{1+\lambda}} \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma} + c_4, \tag{7.17d}$$

where $\operatorname{erf}(\cdot)$ denotes the error function (see Appendix B). The next step is to calculate the integration constants. They are determined by the intended limits,

for $\lambda < 0, x < 0:$

$$\lim_{x\to-\infty} \left(\frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}} \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma} + c_1 \right) \overset{!}{=} 0$$

$$\Leftrightarrow \qquad c_1 = \frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}, \tag{7.18a}$$

for $\lambda < 0, x \geq 0:$

$$\lim_{x\to\infty} \left(\frac{1}{1+\sqrt{1-\lambda}} \operatorname{erf} \frac{\sqrt{1-\lambda}\,x}{\sqrt{2}\,\sigma} + c_2 \right) \overset{!}{=} 1$$

$$\Leftrightarrow \qquad c_2 = \frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}, \tag{7.18b}$$

for $\lambda \geq 0, x < 0:$

$$\lim_{x\to-\infty} \left(\frac{1}{1+\sqrt{1+\lambda}} \operatorname{erf} \frac{\sqrt{1+\lambda}\,x}{\sqrt{2}\,\sigma} + c_3 \right) \overset{!}{=} 0$$

$$\Leftrightarrow \qquad c_3 = \frac{1}{1+\sqrt{1+\lambda}}, \tag{7.18c}$$

for $\lambda \geq 0, x \geq 0:$

$$\lim_{x\to\infty} \left(\frac{\sqrt{1+\lambda}}{1+\sqrt{1+\lambda}} \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma} + c_4 \right) \overset{!}{=} 1$$

$$\Leftrightarrow \qquad c_4 = \frac{1}{1+\sqrt{1+\lambda}}. \tag{7.18d}$$

From the partial results in (7.17a)-(7.17d) follows with the integration constants as given in (7.18a)-(7.18d) the distribution function of the naïve skew-normal distribution.

Definition 7.2. A random variable X is said to be *naïve skew-normal* with skewness (or shape) parameter λ, $\lambda \in \mathbb{R}$, written $X \sim \mathcal{NSN}(\lambda)$, if its cumulative distribution function is defined as

$$F(x;\lambda) = \begin{cases} \frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}} \left(1 + \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma}\right) & \text{if } \lambda < 0, x < 0 \\ \frac{1}{1+\sqrt{1-\lambda}} \left(\sqrt{1-\lambda} + \operatorname{erf} \frac{\sqrt{1-\lambda}\,x}{\sqrt{2}\,\sigma}\right) & \text{if } \lambda < 0, x \geq 0 \\ \frac{1}{1+\sqrt{1+\lambda}} \left(1 + \operatorname{erf} \frac{\sqrt{1+\lambda}\,x}{\sqrt{2}\,\sigma}\right) & \text{if } \lambda \geq 0, x < 0 \\ \frac{1}{1+\sqrt{1+\lambda}} \left(1 + \sqrt{1+\lambda}\, \operatorname{erf} \frac{x}{\sqrt{2}\,\sigma}\right) & \text{if } \lambda \geq 0, x \geq 0, \end{cases} \tag{7.19}$$

where $\operatorname{erf}(\cdot)$ denotes the error function.

Illustrated in Figure 7.9 are graphs of the NSN distribution for several degrees of skewness.

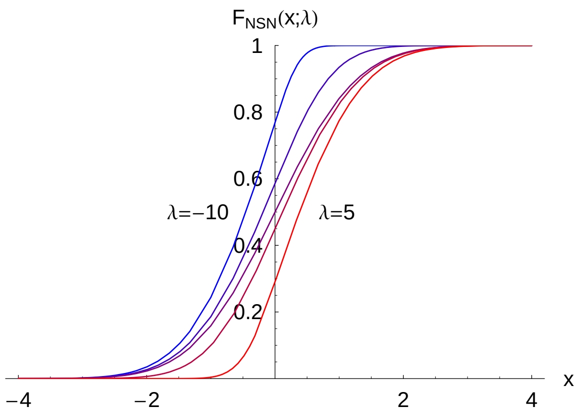


Figure 7.9: Cumulative distribution functions $F_{\text{NSN}}(x;\lambda)$ of the naïve skew-normal distribution with skewness parameter $\lambda \in \{-10, -1, 0, 1/2, 5\}$

7.4.4 Moments

Next, formulas for the first three moments of a $\mathcal{NSN}(\lambda)$ distributed random variable X are derived.

Expectation

The expectation of a random variable X distributed according to (7.19) can be calculated as follows:

If $\lambda < 0$:

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x f(x;\lambda) dx \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma} \left(\int_{-\infty}^{0} x\, e^{-\frac{x^2}{2\sigma^2}} dx + \int_{0}^{\infty} x\, e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} dx \right) \qquad (7.20a) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma} \left(-\sigma^2 + \frac{\sigma^2}{1-\lambda} \right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{\left(1+\sqrt{1-\lambda}\right)\sqrt{1-\lambda}}
\end{aligned}$$

If $\lambda \geq 0$:

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x f(x;\lambda) dx \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma} \left(\int_{-\infty}^{0} x\, e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} dx + \int_{0}^{\infty} x\, e^{-\frac{x^2}{2\sigma^2}} dx \right) \qquad (7.20b) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma} \left(-\sigma^2 + \frac{\sigma^2}{1+\lambda} \right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{\left(1+\sqrt{1+\lambda}\right)\sqrt{1+\lambda}}
\end{aligned}$$

Formulas 7.20a and 7.20b can be consolidated. The expectation of a $\mathcal{NSN}(\lambda)$ distributed random variable X is then given by

$$\mathrm{E}(X) = \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{\left(1+\sqrt{1+|\lambda|}\right)\sqrt{1+|\lambda|}}, \qquad \forall\, \lambda \in \mathbb{R}, \qquad (7.21)$$

depicted in Figure 7.10. For the limits holds

$$\lim_{\lambda\to-\infty} \big(\mathrm{E}(X)\big) = -\sqrt{\frac{2}{\pi}} \qquad (7.22a)$$

$$\lim_{\lambda\to\infty} \big(\mathrm{E}(X)\big) = \sqrt{\frac{2}{\pi}}. \qquad (7.22b)$$

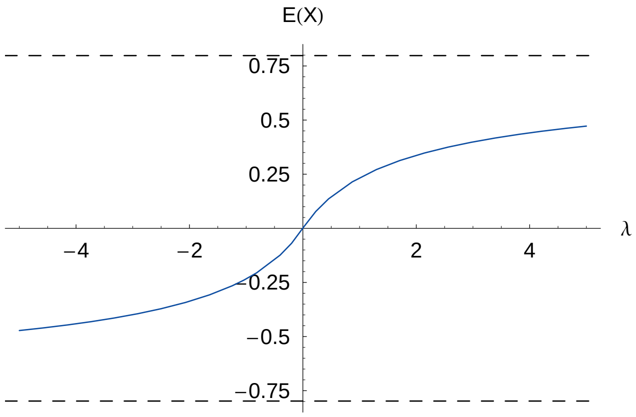


Figure 7.10: Expectation of a $\mathcal{NSN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

Variance

The variance of a random variable X distributed according to (7.19) is given by:

If $\lambda < 0$:

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \, f(x;\lambda) dx \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\,\sigma} \left(\int_{-\infty}^{0} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1-\lambda+\sqrt{1-\lambda}} \right)^2 e^{-\frac{x^2}{2\sigma^2}} dx \right. \\
&\quad \left. + \int_{0}^{\infty} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1-\lambda+\sqrt{1-\lambda}} \right)^2 e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} dx \right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\,\sigma} \left(\frac{(\pi - \lambda(\pi-2))\,\sigma^3}{\sqrt{2\,\pi}\,(1-\lambda)} + \frac{(\pi - 2\,\lambda)\,\sigma^3}{\sqrt{2\,\pi}\,(1-\lambda)^{3/2}} \right) \\
&= \frac{-2\lambda + \pi + \sqrt{1-\lambda}\,(-(\pi-2)\lambda + \pi)}{\pi\,\left(1+\sqrt{1-\lambda}\right)(1-\lambda)}\,\sigma^2 \qquad (7.23a)
\end{aligned}$$

If $\lambda \geq 0$:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \, f(x;\lambda) dx$$

$$= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma} \left(\int_{-\infty}^{0} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1+\lambda+\sqrt{1+\lambda}} \right)^2 e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} dx \right.$$

$$\left. + \int_{0}^{\infty} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1+\lambda+\sqrt{1+\lambda}} \right)^2 e^{-\frac{x^2}{2\sigma^2}} dx \right)$$

$$= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\,\sigma} \left(\frac{(\pi+2\,\lambda)\,\sigma^3}{\sqrt{2\,\pi}\,(1+\lambda)^{3/2}} + \frac{(\pi+\lambda(\pi-2))\,\sigma^3}{\sqrt{2\,\pi}\,(1+\lambda)} \right)$$

$$= \frac{2\lambda+\pi+\sqrt{1+\lambda}\,((\pi-2)\lambda+\pi)}{\pi\,\left(1+\sqrt{1+\lambda}\right)(1+\lambda)}\,\sigma^2 \tag{7.23b}$$

Also formulas 7.23a and 7.23b can be consolidated. The variance of a $\mathcal{NSN}(\lambda)$ distributed random variable X is then given by

$$\mathrm{Var}(X) = \frac{2\,|\lambda| + \pi + \sqrt{1+|\lambda|}\,((\pi-2)\,|\lambda| + \pi)}{\pi\,\left(1+\sqrt{1+|\lambda|}\right)(1+|\lambda|)}\,\sigma^2, \quad \forall\,\lambda \in \mathbb{R}, \tag{7.24}$$

depicted in Figure 7.11. For the limits holds

$$\lim_{\lambda\to-\infty} \big(\mathrm{Var}(X)\big) = \frac{(\pi\,-2)\,\sigma^2}{\pi} \tag{7.25a}$$

$$\lim_{\lambda\to\infty} \big(\mathrm{Var}(X)\big) = \frac{(\pi\,-2)\,\sigma^2}{\pi}. \tag{7.25b}$$

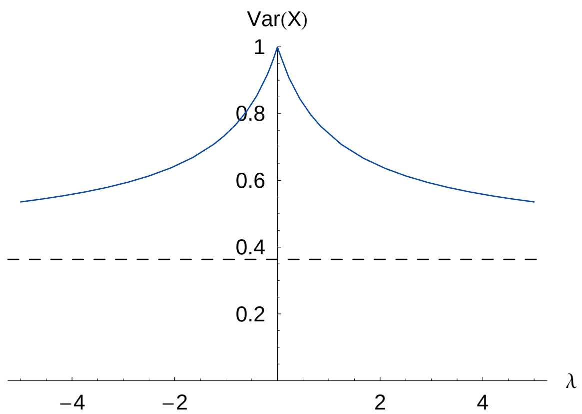


Figure 7.11: Variance of a $\mathcal{NSN}(\lambda)$ distributed random variable vs. skewness parameter. The limit is indicated with a dashed line.

Skewness

The skewness of a random variable X distributed according to (7.19) can be calculated as follows:

If $\lambda < 0$:

$$\begin{aligned}
\gamma_1(X) = \frac{\mu_3}{\sigma^3} &= \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^3 f(x;\lambda) dx \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma^4} \left(\int_{-\infty}^{0} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1-\lambda+\sqrt{1-\lambda}} \right)^3 e^{-\frac{x^2}{2\sigma^2}} dx \right. \\
&\quad \left. + \int_{0}^{\infty} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1-\lambda+\sqrt{1-\lambda}} \right)^3 e^{-\frac{(1-\lambda)x^2}{2\sigma^2}} dx \right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1-\lambda}}{\left(1+\sqrt{1-\lambda}\right)\sigma^4} \left(\frac{\left((\pi-4)(1-\lambda)^{3/2} - 3\,\pi(1-\lambda) - 6\,\lambda + 4\right)\,\sigma^4}{\pi\,(1-\lambda)^{3/2}} \right. \\
&\quad \left. - \frac{\left(\sqrt{1-\lambda}\,(4+2\,\lambda-3\,\pi)+\pi-4\right)\,\sigma^4}{\pi\,(1-\lambda)^2} \right) \\
&= \sqrt{\frac{2}{\pi^3}} \frac{\lambda\left(-(4-\pi)\,\lambda + (3\,\pi-8)\sqrt{1-\lambda} - 2\,\pi + 8\right)}{\left(1+\sqrt{1-\lambda}\right)(1-\lambda)^{3/2}}
\end{aligned} \tag{7.26a}$$

If $\lambda \geq 0$:

$$\begin{aligned}
\gamma_1(X) = \frac{\mu_3}{\sigma^3} &= \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^3 f(x;\lambda) dx \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma^4} \left(\int_{-\infty}^{0} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1+\lambda+\sqrt{1+\lambda}} \right)^3 e^{-\frac{(1+\lambda)x^2}{2\sigma^2}} dx \right. \\
&\quad \left. + \int_{0}^{\infty} \left(x - \sqrt{\frac{2}{\pi}} \frac{\lambda\,\sigma}{1+\lambda+\sqrt{1+\lambda}} \right)^3 e^{-\frac{x^2}{2\sigma^2}} dx \right) \\
&= \sqrt{\frac{2}{\pi}} \frac{\sqrt{1+\lambda}}{\left(1+\sqrt{1+\lambda}\right)\sigma^4} \left(- \frac{\left((\pi-4)(1+\lambda)^{3/2} - 3\,\pi(1+\lambda) + 6\,\lambda - 4\right)\,\sigma^4}{\pi\,(1+\lambda)^{3/2}} \right. \\
&\quad \left. + \frac{\left(\sqrt{1+\lambda}\,(4-2\,\lambda+3\,\pi)+\pi-4\right)\,\sigma^4}{\pi\,(1+\lambda)^2} \right) \\
&= \sqrt{\frac{2}{\pi^3}} \frac{\lambda\left((4-\pi)\,\lambda + (3\,\pi-8)\sqrt{1+\lambda} - 2\,\pi + 8\right)}{\left(1+\sqrt{1+\lambda}\right)(1+\lambda)^{3/2}}
\end{aligned} \tag{7.26b}$$

Again, formulas 7.26a and 7.26b can be consolidated. The skewness of a $\mathcal{NSN}(\lambda)$ distributed random variable X is then given by

$$\gamma_1(X) = \sqrt{\frac{2}{\pi^3}}\, \frac{\lambda\left((4-\pi)\,|\lambda| + (3\,\pi - 8)\,\sqrt{1+|\lambda|} - 2\,\pi + 8\right)}{\left(1+\sqrt{1+|\lambda|}\right)\,(1+|\lambda|)^{3/2}}, \quad \forall\, \lambda \in \mathbb{R}, \tag{7.27}$$

illustrated in Figure 7.12. For the limits holds

$$\lim_{\lambda\to-\infty}\big(\gamma_1(X)\big) = -\lambda\,\frac{(4-\pi)\,\sigma^3}{\pi^{3/2}} \tag{7.28a}$$

$$\lim_{\lambda\to\infty}\big(\gamma_1(X)\big) = \lambda\,\frac{(4-\pi)\,\sigma^3}{\pi^{3/2}}. \tag{7.28b}$$

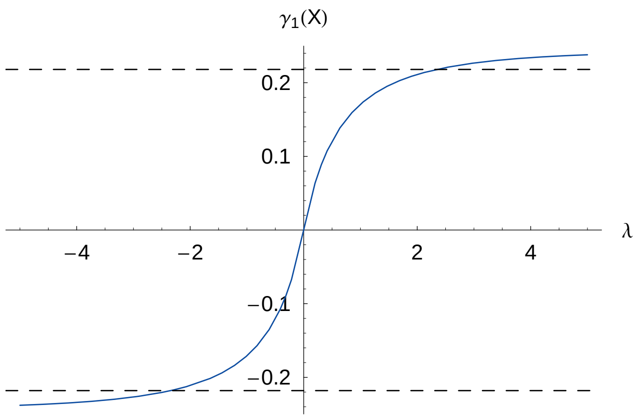


Figure 7.12: Skewness of a $\mathcal{NSN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

The above results are subsumed in the following lemma.

Lemma 7.1. *Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{NSN}(\lambda)$ distributed random variable. Then for the first three moments holds*

$$\mathrm{E}(X) = \sqrt{\frac{2}{\pi}}\, \frac{\lambda\,\sigma}{\left(1+\sqrt{1+|\lambda|}\right)\sqrt{1+|\lambda|}}, \quad \forall\, \lambda \in \mathbb{R} \tag{7.29}$$

$$\mathrm{Var}(X) = \frac{2\,|\lambda| + \pi + \sqrt{1+|\lambda|}\,((\pi-2)\,|\lambda| + \pi)}{\pi\left(1+\sqrt{1+|\lambda|}\right)(1+|\lambda|)}\,\sigma^2, \quad \forall\, \lambda \in \mathbb{R} \tag{7.30}$$

$$\gamma_1(X) = \sqrt{\frac{2}{\pi^3}}\, \frac{\lambda\left((4-\pi)\,|\lambda| + (3\,\pi - 8)\,\sqrt{1+|\lambda|} - 2\,\pi + 8\right)}{\left(1+\sqrt{1+|\lambda|}\right)\,(1+|\lambda|)^{3/2}}, \quad \forall\, \lambda \in \mathbb{R}. \tag{7.31}$$

Proof. Follows directly from the above calculations. □

7.4.5 Inverse of the Distribution

$\mathcal{NSN}(\lambda)$ distributed random variables can be generated using the inversion method, just like $\Xi_{c,\sigma}$ distributed random numbers. The necessary inverse distribution is defined as

$$F^{-1}_{\text{NSN}}(y;\lambda) = \begin{cases} \sqrt{2}\,\sigma\,\text{erf}^{-1}\left(y\left(1+\frac{1}{\sqrt{1-\lambda}}\right)-1\right) & \text{if } \lambda<0, y<\frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}} \\ \frac{\sqrt{2}\,\sigma}{\sqrt{1-\lambda}}\,\text{erf}^{-1}\left(y\left(1+\sqrt{1-\lambda}\right)-\sqrt{1-\lambda}\right) & \text{if } \lambda<0, y\geq\frac{\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}} \\ \frac{\sqrt{2}\,\sigma}{\sqrt{1+\lambda}}\,\text{erf}^{-1}\left(y\left(1+\sqrt{1+\lambda}\right)-1\right) & \text{if } \lambda\geq 0, y<\frac{1}{1+\sqrt{1+\lambda}} \\ \sqrt{2}\,\sigma\,\text{erf}^{-1}\left(y\left(1+\frac{1}{\sqrt{1+\lambda}}\right)-\frac{1}{\sqrt{1+\lambda}}\right) & \text{if } \lambda\geq 0, y\geq\frac{1}{1+\sqrt{1+\lambda}}, \end{cases} \tag{7.32}$$

where $\text{erf}^{-1}(\cdot)$ denotes the inverse error function.

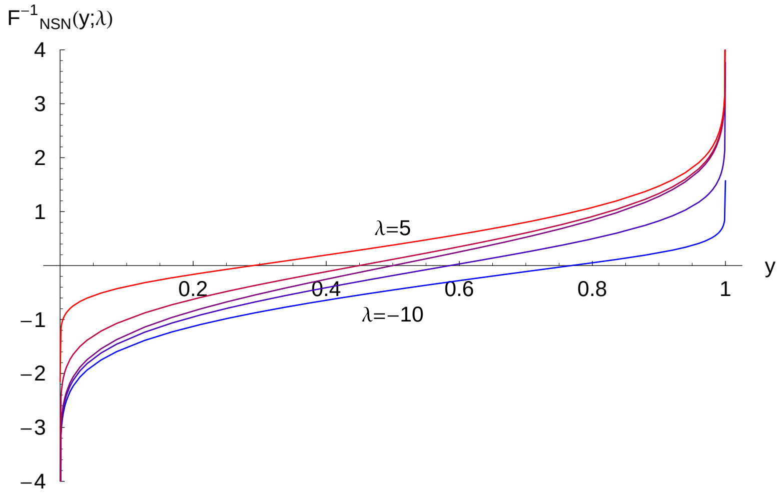


Figure 7.13: Inverse cumulative distribution functions $F^{-1}_{\text{NSN}}(y;\lambda)$ of the naïve skew-normal distribution with skewness parameter $\lambda \in \{-10,-1,0,1/2,5\}$

7.4.6 Random Variate Generation

Since $\mathcal{NSN}(\lambda)$ distributed random variables are generated using the same method as described in subsection 7.2.4, the same drawbacks as mentioned there occur here, too.

7.5 Standardized Naïve Skew-Normal Distribution

From Figure 7.11 we can see that the variance of the NSN distribution is convergent, but still spreads about $0.64\,\sigma^2$, cf. (7.25). To make the variance shape invariant (recall the discussion from subsection 5.1.1), a linear transformation has to be applied to the NSN distribution leading to the standardized naïve skew-normal distribution (SNSN). This transformation obviously depends on the skewness parameter λ. Taking into account that by Theorem 3.10 $\operatorname{Var}(a + bX) = b^2 \operatorname{Var}(X)$ holds and the standardized variance shall be $(s(\lambda))^2 \operatorname{Var}(X) \overset{!}{=} 1$, from the definition of the NSN distribution's variance (7.24) follows

$$s(\lambda) = \sqrt{\frac{\pi(1 + |\lambda|)}{4\left(\sqrt{1 + |\lambda|} - 1\right) + |\lambda|\,(\pi - 2) + \pi\left(2 - \sqrt{1 + |\lambda|}\right)}}. \tag{7.33}$$

Normalization will be treated for the more relevant skew-normal distribution in greater detail in Section 8.5. Since the SNSN distribution offers no principally new properties and suffers due to its construction principle from the problems already described, we present no further details here. For a more accurate discussion on the SNSN distribution see [Berlik, 2004a].

8 Skewing Function Approach

In Chapter 7 we presented several skew distributions constructed by piecewise defined functions. It was shown that the problem of diverging moments arising with the $\Xi_{c,\sigma}$ distribution could be solved with the naïve skew normal distribution. However, there remain some construction principle immanent problems. The distributions are all of limited mathematical tractability. Furthermore, random variate generation by means of the inversion method is expensive due to the relative complex functions. Last, the function definitions are quite unaesthetic.

All this will be fixed by the construction principle presented next. The fundamental idea is to multiply an arbitrary symmetric probability density function with a skewing function. This operation amplifies one side of the pdf and attenuates the other at the same amount such that on average the pdf remains the same up to a constant factor. Loosely speaking, the two operations annihilate each other in total. Using this method skew distributions occur that are of striking simplicity and beauty.

The remainder of this chapter is organized as follows: first, the construction principle is discussed in more detail in Section 8.1 and subsequently some appropriate skewing functions are outlined (Section 8.2). Also distributions can serve as skewing functions. This issue is treated separately in Section 8.3. The most prominent member of skew distributions resulting from this approach, the skew-normal distribution, is introduced in Section 8.4. A standardized variant is provided in the proximate section and last some related families of skew distributions are outlined in Section 8.6.

8.1 Construction Principle

The principal idea of all distributions treated in this chapter is to use sigmoid functions to perturb arbitrary symmetric pdfs. By symmetric we mean symmetry with respect to the ordinate, i.e. functions fulfilling $f(x) = f(-x)$. We will use "symmetric pdf" and $f(x) = f(-x)$ interchangeably in the sequel.

A special case of sigmoid functions are monotone, i.e. non periodic skewing functions. Multiplying a symmetric distribution with a monotone skewing function results in a weighting with a pleasant property. As a pair of corresponding points $f(x)$ and $f(-x)$ of the pdf is weighted with factors s and $(1-s)$ respectively, the sum of every weighted pair complements again to $f(x)$; since for the sum holds $s\,f(x) + (1-s)f(-x) = s\,f(x) + f(-x) - s\,f(-x) = s\,f(x) + f(x) - s\,f(x) = f(x)$, using the symmetry of the pdf in the second last step. Thus, the skewing function cancels out in the weighted pdf on average by just halving the pdf's mass. This can simply be fixed introducing a factor of two, resulting in a skewing that does not change the pdf's mass in total, though the allocation of the mass.

A definition of skewing functions is for example given by Wang et al. [2004b] which is here specialized in monotone functions.

Definition 8.1. A function π is called a *skewing function*, if it satisfies

$$0 \leq \pi(x) \leq 1 \quad \text{and} \quad \pi(-x) = 1 - \pi(x). \tag{8.1}$$

We will call a skewing function a *monotone skewing function* if it is additionally monotone.

The construction principle outlined above is formalized in the following lemma.

Lemma 8.1. *Let Y be a random variable with probability density function f_0 symmetric about the ordinate, and π a monotone skewing function. Then*

$$f(x; \lambda) = 2\, f_0(x)\, \pi(\lambda\, x), \quad x, \lambda \in \mathbb{R}, \tag{8.2}$$

is a probability density function of a random variable X for any λ.

Proof. It has to be shown that (8.2) is a pdf. Per definition of the pdf and the skewing function holds that (8.2) is positive. With this and

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x;\lambda)dx &= \int_{-\infty}^{\infty} 2\, f_0(x)\, \pi(\lambda\, x)dx \\
&= \int_{-\infty}^{\infty} f_0(x)\, \pi(\lambda\, x)dx + \int_{-\infty}^{\infty} f_0(x)\, (1 - \pi(-\lambda\, x))\, dx \\
&= \int_{-\infty}^{\infty} f_0(x)\, \pi(\lambda\, x)dx + \int_{-\infty}^{\infty} f_0(x)dx - \int_{-\infty}^{\infty} f_0(x)\, \pi(-\lambda\, x)dx \\
&= \int_{-\infty}^{\infty} f_0(x)\, \pi(\lambda\, x)dx + \int_{-\infty}^{\infty} f_0(x)dx - \int_{-\infty}^{\infty} f_0(-x)\, \pi(\lambda\, x)dx \\
&= \int_{-\infty}^{\infty} f_0(x)\, \pi(\lambda\, x)dx + \int_{-\infty}^{\infty} f_0(x)dx - \int_{-\infty}^{\infty} f_0(x)\, \pi(\lambda\, x)dx \\
&= \int_{-\infty}^{\infty} f_0(x)dx = 1
\end{aligned} \tag{8.3}$$

the claim follows, since the reflection $f_0(x)\, \pi(-\lambda\, x) = f_0(-x)\, \pi(\lambda\, x)$ does not change the integral and $f_0(-x) = f_0(x)$ holds due to the symmetry of $f_0(x)$ about the ordinate. □

Note that since every monotone increasing skewing function is also a cdf, the claim of Lemma 8.1 also follows immediately from Azzalini's probabilistic proof of Lemma 8.3 which is more stringent and tailored to the domain. Due to symmetry, Lemma 8.3 holds as well for monotone decreasing skewing functions.

8.2 Skewing Functions

Several functions can be used to perturb a given distribution. In what follows we discuss the hyperbolic tangent in subsection 8.2.1, the inverse tangent in subsection 8.2.2, the inverse cotangent in subsection 8.2.3, the logistic distribution in subsection 8.2.4, and the error function in subsection 8.2.5. Note that not all of these functions comply with the requirements of Definition 8.1, as not all of them are unipolar, i.e. range from zero or one. Since some are bipolar sigmoid functions or have different limits they first have to be translated or scaled accordingly to serve as skewing functions. However, for the sake of clarity the original definitions are given, only extended with a parameter λ controlling the shape.

8.2.1 Hyperbolic Tangent

The hyperbolic tangent $\tanh x$ is defined for all values of the argument x by

$$\tanh x = \frac{\sinh x}{\cosh x}. \tag{8.4}$$

It is restricted in range and takes values only between -1 to 1 (cf. also [Spanier and Oldham, 1987, Ch. 30]). The function is often used as activation function in artificial neural networks.

To use it as a skewing function it is here parameterized with the shape parameter λ. The parameterized hyperbolic tangent can be defined in terms of the exponential function,

$$\tanh(x;\lambda) = \tanh(\lambda\,x) = \frac{e^{2\lambda\,x}-1}{e^{2\lambda\,x}+1} = \frac{e^{\lambda\,x}-e^{-\lambda\,x}}{e^{\lambda\,x}+e^{-\lambda\,x}}, \quad x,\lambda \in \mathbb{R}. \tag{8.5}$$

For positive shape parameters function (8.5) approaches $+1$ as $x \to \infty$ and -1 as $x \to -\infty$. For $\lambda = 0$ it is constantly 0, and for negative shape parameters it approaches -1 as $x \to \infty$ and $+1$ as $x \to -\infty$. Several graphs are illustrated in Figure 8.1.

8.2.2 Inverse Tangent

The inverse tangent $\arctan x$ is the inverse function of the tangent, defined for all values of the argument x by

$$\arctan x = \int_0^x \frac{\mathrm{d}t}{1+t^2}. \tag{8.6}$$

It is a multivalued function, taking its principal values only in the restricted range $-\pi/2$ to $\pi/2$ (cf. also [Spanier and Oldham, 1987, Ch. 35]). For its use as skewing function it is here parameterized with the shape parameter λ. The parameterized inverse tangent may also be defined in terms of the complex

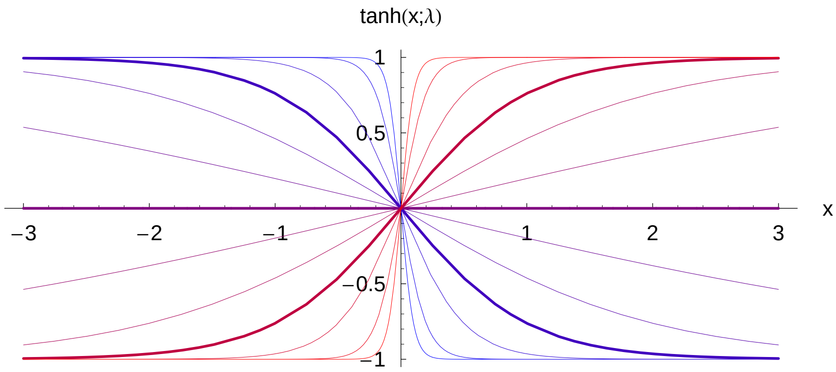


Figure 8.1: Parameterized hyperbolic tangent for selected parameters, $\lambda \in [-10, 10]$. The bold blue, violet, and red line indicate $\tanh(x; -1)$, $\tanh(x; 0)$, and $\tanh(x; 1)$, respectively.

logarithm,

$$\arctan(x; \lambda) = \arctan(\lambda\, x) = \frac{i}{2}\log(1 - i\,\lambda\, x) - \frac{i}{2}\log(1 + i\,\lambda\, x), \quad x, \lambda \in \mathbb{R}. \tag{8.7}$$

For positive shape parameters function (8.7) approaches $\pi/2$ as $x \to \infty$ and $-\pi/2$ as $x \to -\infty$. For $\lambda = 0$ it is constantly 0, and for negative shape parameters it approaches $-\pi/2$ as $x \to \infty$ and $+\pi/2$ as $x \to -\infty$. Several graphs are illustrated in Figure 8.2.

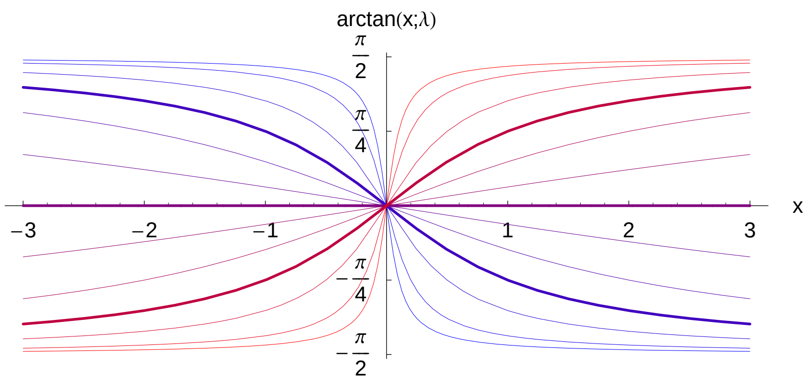


Figure 8.2: Parameterized inverse tangent for selected parameters, $\lambda \in [-10, 10]$. The bold blue, violet, and red line indicate $\arctan(x; -1)$, $\arctan(x; 0)$, and $\arctan(x; 1)$, respectively.

8.2.3 Inverse Cotangent

The inverse cotangent $\operatorname{arccot} x$ is the inverse function of the cotangent, defined for all values of the argument x by

$$\operatorname{arccot} x = \frac{\pi}{2} - \arctan x. \tag{8.8}$$

It is a multivalued function, taking its principal values only in the restricted range 0 to π. There is at least one other possible convention for defining the inverse cotangent, taking $\operatorname{arccot} x$ to have the range $[-\pi/2, \pi/2]$, (cf. also [Spanier and Oldham, 1987, Ch. 35]). To use it as a skewing function it is here parameterized with the shape parameter λ. The parameterized inverse cotangent is defined by

$$\operatorname{arccot}(x;\lambda) = \frac{\pi}{2} - \arctan(x;\lambda), \quad x, \lambda \in \mathbb{R}. \tag{8.9}$$

For positive shape parameters function (8.9) approaches 0 as $x \to \infty$ and π as $x \to -\infty$. For $\lambda = 0$ it is constantly $\pi/2$, and for negative shape parameters it approaches π as $x \to \infty$ and 0 as $x \to -\infty$. Note that positive shape parameters will lead to negatively skewed distributions. Several graphs are illustrated in Figure 8.3.

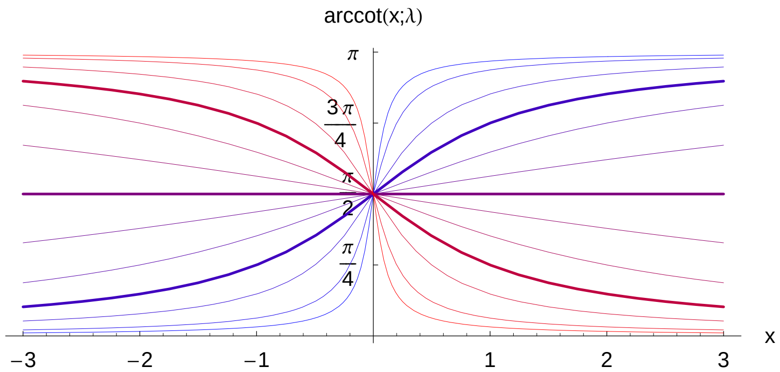


Figure 8.3: Parameterized inverse cotangent for selected parameters, $\lambda \in [-10, 10]$. The bold blue, violet, and red line indicate $\operatorname{arccot}(x;-1)$, $\operatorname{arccot}(x;0)$, and $\operatorname{arccot}(x;1)$, respectively.

8.2.4 Logistic Distribution

The logistic distribution function is probably the simplest possible sigmoid function since just one single $\exp(\cdot)$ term achieves the two distinct limits. Like the hyperbolic tangent it is often used as activation function in artificial neural networks. The distribution was already defined in (3.23) and is here equipped with a parameter λ controlling the shape,

$$f_{\mathrm{LD}}(x;\lambda) = \frac{1}{1 + e^{-\lambda x}}, \quad x, \lambda \in \mathbb{R}. \tag{8.10}$$

For positive shape parameters function (8.10) approaches 1 as $x \to \infty$ and 0 as $x \to -\infty$. For $\lambda = 0$ it is constantly $1/2$, and for negative shape parameters it approaches 0 as $x \to \infty$ and 1 as $x \to -\infty$. Several graphs are illustrated in Figure 8.4.

The logistic distribution (8.10) is closely related to the hyperbolic tangent (8.5) as it is an adaptation of the latter to the range $[0, 1]$ with doubled skewness

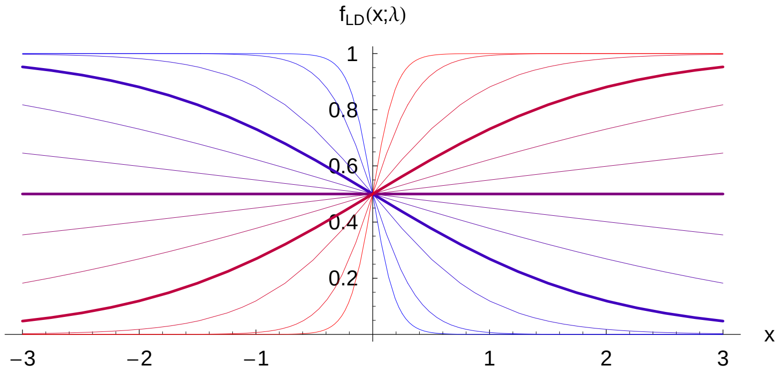


Figure 8.4: Parameterized logistic distribution for selected parameters, $\lambda \in [-10, 10]$. The bold blue, violet, and red line indicate $f_{\mathrm{LD}}(x; -1)$, $f_{\mathrm{LD}}(x; 0)$, and $f_{\mathrm{LD}}(x; 1)$, respectively.

parameter,

$$\begin{aligned}\frac{1}{2}\left(1+\tanh(x;\lambda)\right) &= \frac{1}{2}\left(1+\tanh(\lambda\, x)\right)\\ &= \frac{1}{2}\left(1+\frac{e^{\lambda\, x}-e^{-\lambda\, x}}{e^{\lambda\, x}+e^{-\lambda\, x}}\right)\\ &= \frac{1}{2}\frac{e^{\lambda\, x}+e^{-\lambda\, x}+e^{\lambda\, x}-e^{-\lambda\, x}}{e^{\lambda\, x}+e^{-\lambda\, x}}\\ &= \frac{e^{\lambda\, x}}{e^{\lambda\, x}+e^{-\lambda\, x}}\\ &= \frac{1}{1+e^{-2\,\lambda\, x}}\\ &= f_{\mathrm{LD}^*}(x; 2\lambda).\end{aligned} \tag{8.11}$$

It is sometimes used as "smooth step function" since it tends to the step function in the limits. The mentioned simplicity was the main reason to apply it in the first skew mutation operator, see subsection 5.2.3.

8.2.5 Error Function

Also the integral of the normal density function, i.e. the error function, is a sigmoid function. It will intensively be used in the sequel as part of the skew-normal distribution (see Section 8.4). It is restricted in range and takes values only between -1 to 1 (cf. also Appendix B). For the use as skewing function it is here extended with the shape controlling parameter λ, defined by

$$\operatorname{erf}(x;\lambda) = \operatorname{erf}(\lambda\, x) = \frac{2}{\sqrt{\pi}}\int_0^{\lambda\, x} e^{-t^2}\mathrm{d}t, \quad x, \lambda \in \mathbb{R}. \tag{8.12}$$

For positive shape parameters function (8.12) approaches $+1$ as $x \to \infty$ and

-1 as $x \to -\infty$. For $\lambda = 0$ it is constantly 0, and for negative shape parameters it approaches -1 as $x \to \infty$ and $+1$ as $x \to -\infty$. Several graphs are illustrated in Figure 8.5.

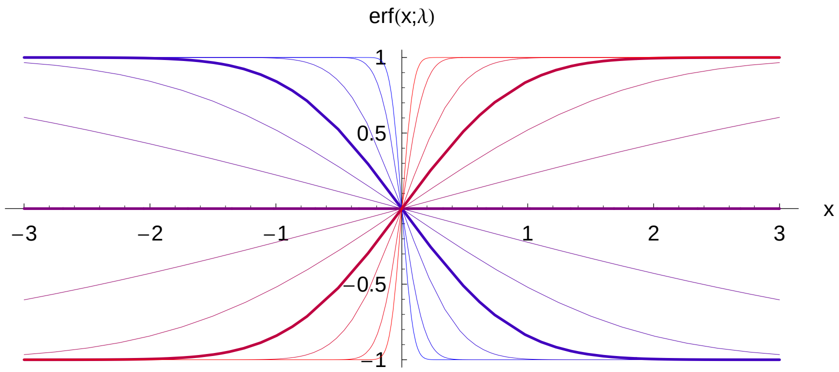


Figure 8.5: Parameterized error function for selected parameters, $\lambda \in [-10, 10]$. The bold blue, violet, and red line indicate $\operatorname{erf}(x; -1)$, $\operatorname{erf}(x; 0)$, and $\operatorname{erf}(x; 1)$, respectively.

8.3 Skewing by Means of Distribution Functions

As we saw in the last section, also distribution functions can serve as skewing functions. One example using the logistic function as skewing function has with the Polymorphic Mutation already been presented in subsection 5.2.3. Azzalini picked up this idea in the following two lemmas.

Lemma 8.2 (Azzalini 1985)**.** *Let Y be a random variable with probability density function f_0 symmetric about the y-axis, and X a random variable with absolutely continuous distribution function G such that G' is symmetric about the y-axis. Then*

$$f(z; \lambda) = 2 f_0(z) G(\lambda z), \qquad -\infty < z < \infty \tag{8.13}$$

is a probability density function of a random variable Z for any real λ.

Proof. Omitted.

Later he extended this basic lemma to a more general form. The skewness there is not just controlled by a real parameter but instead any odd function is allowed as skewness control.

Lemma 8.3 (Azzalini 2005, Lemma 1)**.** *Let f_0 be a probability density function symmetric about the y-axis and G an absolutely continuous distribution function such that G' is symmetric about the y-axis. Then*

$$f(z) = 2 f_0(z) G(w(z)), \qquad -\infty < z < \infty \tag{8.14}$$

is a probability density function for any odd function $w(\cdot)$.

Proof. If $Y \sim f_0$ and $X \sim G'$ are independent random variables, then

$$\frac{1}{2} = \mathrm{P}(X - w(Y) \leq 0) = \mathrm{E}_Y(\mathrm{P}(X - w(Y) \leq 0 \mid Y)) = \int_{\mathbf{R}} G(w(z)) f_0(z) dz$$

on noticing that $w(Y)$ and $X - w(Y)$ also have symmetric distributions about the y-axis. □

Lemma 8.3 constitutes a general rule to manipulate a symmetric *basis density* f_0 through a *perturbation function* $G(w(z))$ to get a new legitimate density f. Note that the set of perturbed densities always includes the basis density, since $w(z) \equiv 0$ gives $f_0 = f$. It will in the next section be used to define the skew-normal distribution.

8.4 Skew-Normal Distribution

The class of distributions that is used to build the most relevant directed mutation operator is called skew-normal (SN) distribution, a term coined by Azzalini [1985]. He was the first who systematically investigated this function; even if its appearance can be traced back in several earlier papers, like for example in [Birnbaum, 1950].

Reasons for the outstanding role the SN distribution plays are summarized in [Azzalini, 2005]. Most relevant for the given context are the following points:

- The normal family is an interior point of this parametric class of probability densities, just as in the practical statistical work the normal family is quite naturally perceived as the "central" form of a range of densities. Contrary to that, for very many other parametric classes of probability densities the normal family is the limiting or the boundary case.
- It retains – at least partly – the mathematical tractability and some formal properties of the standard parametric class, i.e. the normal family.
- Via its stochastic representations it provides a simple mechanism of genesis of variates. Only because of this its application as directed mutation operator is possible. In addition, the availability of a stochastic representation allows simple derivation of some formal properties of the distribution.
- Just one parameter regulates the shape of the distributions with high flexibility and therewith their main characteristic, the skewness.

In the following subsections we will give the definition of the SN distribution, list some helpful properties including the stochastic representation and the moments, and present a method for random variate generation. For a more accurate treatise on it and some extensions, the reader is referred e.g. to [Azzalini, 2005] or [Genton, 2004, 2005]. A small historical review is given by Arnold and Beaver [2002].

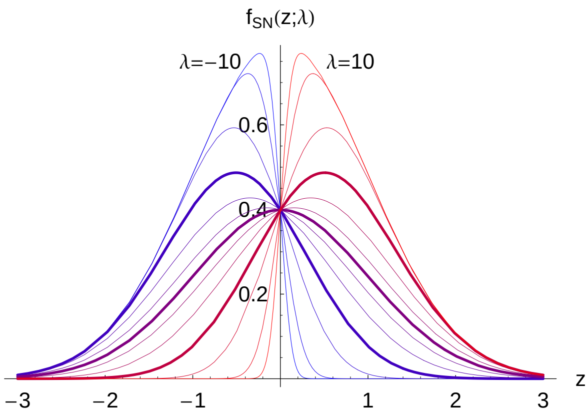


Figure 8.6: Density functions of the skew-normal distribution for selected skewness parameters, $\lambda \in [-10, 10]$. Bold lines indicate the SN(-1), SN(0), and SN(1) distribution.

8.4.1 Probability Density Function

On using Lemma 8.3 with $f_0 = \phi$ and $G = \Phi$, the probability density function and the cumulative distribution function of the standard normal density, respectively, and $w(z) = \lambda\, z$ where $\lambda \in \mathbb{R}$, we get the density of the SN distribution.

Definition 8.2. A random variable Z is said to be *skew-normal* with skewness (or shape) parameter λ, $\lambda \in \mathbb{R}$, written as $Z \sim \mathcal{SN}(\lambda)$, if its probability density function is

$$f_{\mathrm{SN}}(z; \lambda) = 2\phi(z)\Phi(\lambda\, z), \qquad z \in \mathbb{R}, \tag{8.15}$$

where ϕ and Φ represent the probability density function and the cumulative distribution function of the standard normal density, respectively.

Positive (negative) values of the shape parameter indicate positive (negative) skewness of the distribution. In the case $\lambda = 0$ the SN density gets back to the normal density (see Figure 8.6).

8.4.2 Stochastic Representation

The stochastic representation of a $\mathcal{SN}(\lambda)$ distributed random variable Z is given by Azzalini [1986] as

$$Z = \begin{cases} Y & \text{if } \ W \leq \lambda Y \\ -Y & \text{if } \ W > \lambda Y, \end{cases} \tag{8.16}$$

where Y and W are two independent identically distributed standard normal random variables.

That Z being defined to equal Y or $-Y$ conditionally on the event $\{\lambda\, Y > W\}$ is in fact $\mathcal{SN}(\lambda)$ distributed proofs e.g. the following equation by Arnold and

Beaver [2002]:

$$\begin{aligned} P(Z \le z) &= P(Y \le z \mid \lambda Y > W) \\ &= \frac{P(Y \le z, \lambda Y > W)}{P(\lambda Y > W)} \\ &= \frac{1}{P(\lambda Y > W)} \int_{-\infty}^{z} \int_{-\infty}^{\lambda y} \varphi(y)\varphi(w)dw\,dy \\ &= \frac{1}{P(\lambda Y > W)} \int_{-\infty}^{z} \varphi(y)\Phi(\lambda y)dy. \end{aligned} \tag{8.17}$$

Since $P(\lambda Y > W) = P(\lambda Y - W > 0) = 1/2$, because $\lambda Y - W$ has a normal distribution with zero mean, it follows after differentiating (8.17) with respect to z that Z has the skew-normal density (8.15).

8.4.3 Some Properties

The SN class enjoys remarkable properties in terms of mathematical tractability. Some results found by Azzalini [1985, 1986] are recapitulated in the following.

(a) If $\lambda = 0$, we obtain the $\mathcal{N}(0,1)$ density.

(b) If $Z \sim \mathcal{SN}(\lambda)$, then $-Z \sim \mathcal{SN}(-\lambda)$.

(c) As $\lambda \to \infty$, density (8.15) converges pointwise to the half-normal density, i.e. $2\phi(z)$ for $z \ge 0$.

(d) If $Z \sim \mathcal{SN}(\lambda)$, then $Z^2 \sim \chi_1^2$.

(e) For fixed λ, density (8.15) is strongly unimodal, i.e. $\log f(z;\lambda)$ is a concave function of z.

Because of property (d) the even moments of the SN distribution are equal to the even moments of the standard normal distribution. To determine the odd moments the moment generating function of Z can be used, see subsection 3.6.6. It follows immediately using the following lemma.

Lemma 8.4 (Zacks [1981, pp. 53–54]). *If U is a $\mathcal{N}(0,1)$ random variable, then*

$$\mathrm{E}(\Phi(hU + k)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right) \tag{8.18}$$

for any real h, k.

The moment generating function is then given by

$$M(t) = 2\exp(t^2/2)\,\Phi(\delta\, t)\,, \tag{8.19}$$

where

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

Henze [1986] gives an alternative to the above approach by providing a formula to explicitly derive the odd moments. It holds

$$\mathrm{E}\left(Z^{2k+1}\right) = \sqrt{\frac{2}{\pi}}\,\lambda(1+\lambda^2)^{-(k+1/2)}2^{-k}(2k+1)!\sum_{\nu=0}^{k}\frac{\nu!(2\lambda)^{2\nu}}{(2\nu+1)!(k-\nu)!}, \tag{8.20}$$

where $Z \sim \mathcal{SN}(\lambda)$ and $k = 0, 1, 2, \dots$.

8.4.4 Moments

With property (d) and (8.19) or (8.20) the first four moments can be derived as

$$\mathrm{E}(Z) = b\delta, \tag{8.21a}$$

$$\mathrm{Var}(Z) = 1 - (b\delta)^2 \tag{8.21b}$$

$$\gamma_1(Z) = \frac{1}{2}(4-\pi)\,\mathrm{sign}(\lambda)\left(\frac{(\mathrm{E}(Z))^2}{\mathrm{Var}(Z)}\right)^{3/2} \tag{8.21c}$$

$$\gamma_2(Z) = 2(\pi-3)\left(\frac{(\mathrm{E}(Z))^2}{\mathrm{Var}(Z)}\right)^{2} \tag{8.21d}$$

where

$$b = \sqrt{\frac{2}{\pi}}$$

and

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

$\gamma_1(Z)$ and $\gamma_2(Z)$ denote the third and fourth standardized cumulants. As desired, both expectation and variance converge. The limits are

$$\lim_{\lambda\to\pm\infty}\big(\mathrm{E}(Z)\big) = \mathrm{sign}(\lambda)\sqrt{\frac{2}{\pi}} \tag{8.22a}$$

$$\lim_{\lambda\to\pm\infty}\big(\mathrm{Var}(Z)\big) = 1 - \frac{2}{\pi} \tag{8.22b}$$

$$\lim_{\lambda\to\pm\infty}\big(\gamma_1(Z)\big) = \mathrm{sign}(\lambda)\frac{\sqrt{2}(\pi-4)}{(\pi-2)^{3/2}} \tag{8.22c}$$

$$\lim_{\lambda\to\pm\infty}\big(\gamma_2(Z)\big) = \frac{8(\pi-3)}{(\pi-2)^2}. \tag{8.22d}$$

Their graphs are depicted in Figures 8.7–8.10. One can see that the variance is convergent, but still spreads about 0.64. To make the variance shape invariant, a linear transformation has to be applied to the SN distribution leading to the standardized SN distribution, see Section 8.5.

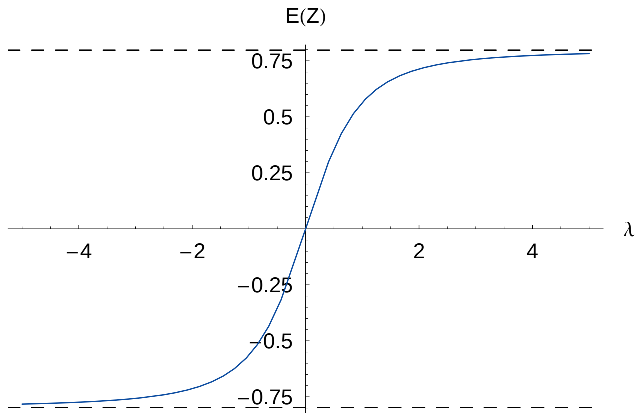


Figure 8.7: Expectation of a $\mathcal{SN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

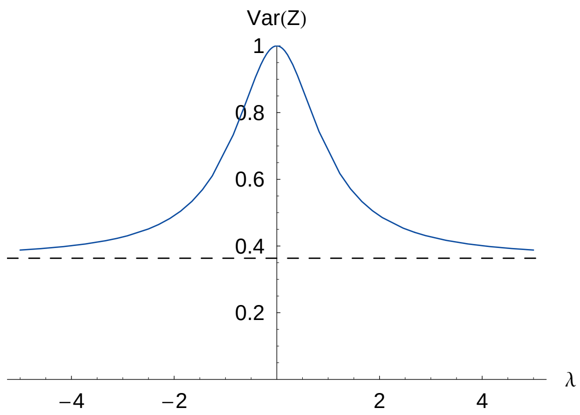


Figure 8.8: Variance of a $\mathcal{SN}(\lambda)$ distributed random variable vs. skewness parameter. The limit is indicated with a dashed line.

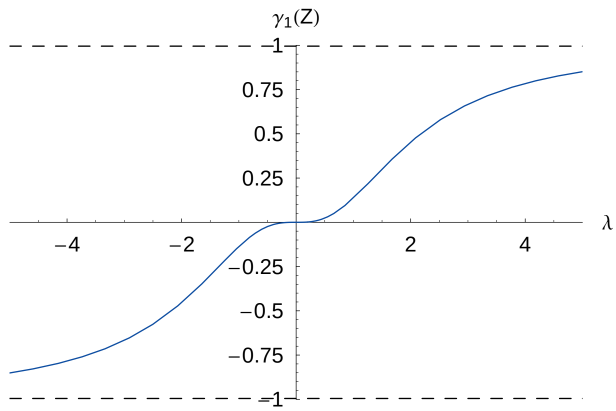


Figure 8.9: Skewness of a $\mathcal{SN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

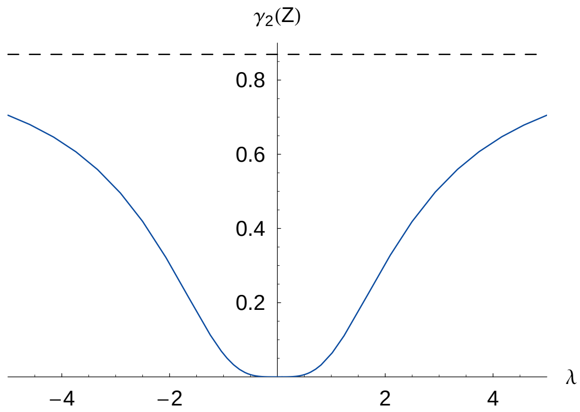


Figure 8.10: Kurtosis of a $\mathcal{SN}(\lambda)$ distributed random variable vs. skewness parameter. The limit is indicated with a dashed line.

8.4.5 Random Variate Generation

Generation of $\mathcal{SN}(\lambda)$ distributed random numbers is simple and fast. A random variable Z with density (8.15) can be generated using its stochastic representation (subsection 8.4.2). Therefore sample Y and W from ϕ and Φ', respectively. Then Z is defined to be equal to Y or $-Y$, conditionally on the event $\{W \leq \lambda Y\}$:

$$Z = \begin{cases} Y & \text{if } \; W \leq \lambda Y \\ -Y & \text{if } \; W > \lambda Y. \end{cases}$$

Thus simply two standard normal random variables are needed to generate one $\mathcal{SN}(\lambda)$ distributed random variable.

8.5 Standardized Skew-Normal Distribution

Using a linear transformation the SN distribution can be changed to a variant where the skewness has not influence on the variance any longer – which thus becomes shape invariant [Berlik and Reusch, 2004]. The transformation that has to be applied depends on the skewness parameter λ. By Theorem 3.10 holds $\operatorname{Var}(a + bZ) = b^2 \operatorname{Var}(Z)$ and the standardized variance shall be $(s(\lambda))^2 \operatorname{Var}(Z) \stackrel{!}{=} 1$. Thus from the definition of the SN distribution's variance (8.21b) follows

$$s(\lambda) = \frac{1}{\sqrt{\operatorname{Var}(Z)}} = \frac{1}{\sqrt{1 - (b\delta)^2}} = \sqrt{\frac{\pi(1 + \lambda^2)}{\pi + (\pi - 2)\lambda^2}}. \tag{8.23}$$

8.5.1 Probability Density Function

In the previous section the linear transformation to be applied to the SN distribution has been calculated. With Theorem 3.11 we can now specify its density. The theorem states that if F is the distribution function of a random variable

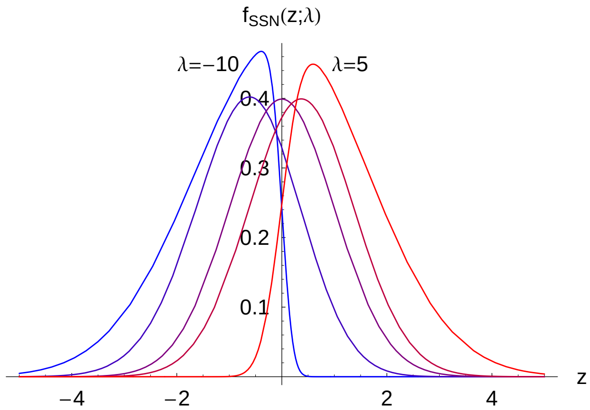


Figure 8.11: Probability density functions $f_{\mathrm{SSN}}(z;\lambda)$ of the Standardized Skew-Normal distribution with skewness parameter $\lambda \in \{-10, -1, 0, 1/2, 5\}$

Z, then $aZ+b$ has distribution function $F\big((z-b)/a\big)$, provided $a>0$. If the corresponding densities exist, they are $f(z)$ and $\frac{1}{a}f\big((z-b)/a\big)$. Thus the density of the Standardized Skew-Normal (SSN) distribution takes the form given in the following definition.

Definition 8.3. A random variable Z is said to be *standardized skew-normal* with skewness (or shape) parameter λ, $\lambda \in \mathbb{R}$, written as $Z \sim \mathcal{SSN}(\lambda)$, if its probability density function is

$$f_{\mathrm{SSN}}(z;\lambda) = \frac{2}{s(\lambda)}\phi\left(\frac{z}{s(\lambda)}\right)\Phi\left(\frac{\lambda z}{s}\right), \qquad z \in \mathbb{R}, \tag{8.24}$$

where ϕ and Φ represent the probability density function and the cumulative distribution function of the standard normal density, respectively.

Due to the standardization the densities with $\lambda \neq 0$ are widened and flattened, see Figure 8.11.

8.5.2 Moments

By Theorems 3.37 and 3.10 holds $\mathrm{E}(a+bZ) = a + b\,\mathrm{E}(Z)$ and $\mathrm{Var}(a+bZ) = b^2\,\mathrm{Var}(Z)$, respectively. With (8.23) we then can deduce the first four moments of the SSN distribution from the moments of the SN distribution (8.21a)–(8.21d):

$$\mathrm{E}(Z) = sb\delta, \tag{8.25a}$$

$$\mathrm{Var}(Z) = 1, \tag{8.25b}$$

$$\gamma_1(Z) = \frac{1}{2}(4-\pi)\big(\mathrm{E}(Z)\big)^3, \tag{8.25c}$$

$$\gamma_2(Z) = 2(\pi-3)\big(\mathrm{E}(Z)\big)^4 \tag{8.25d}$$

where

$$b = \sqrt{\frac{2}{\pi}}$$

and

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

After some algebra we realize that the limits of the SNN class' moments are

$$\lim_{\lambda\to\pm\infty} \big(\mathrm{E}(Z)\big) = \mathrm{sign}(\lambda)\sqrt{\frac{2}{\pi-2}} \tag{8.26a}$$

$$\lim_{\lambda\to\pm\infty} \big(\mathrm{Var}(Z)\big) = 1 \tag{8.26b}$$

$$\lim_{\lambda\to\pm\infty} \big(\gamma_1(Z)\big) = \mathrm{sign}(\lambda)\frac{\sqrt{2}(\pi-4)}{(\pi-2)^{3/2}} \tag{8.26c}$$

$$\lim_{\lambda\to\pm\infty} \big(\gamma_2(Z)\big) = \frac{8(\pi-3)}{(\pi-2)^2}. \tag{8.26d}$$

Graphs of these moments are shown in Figures 8.12–8.15.

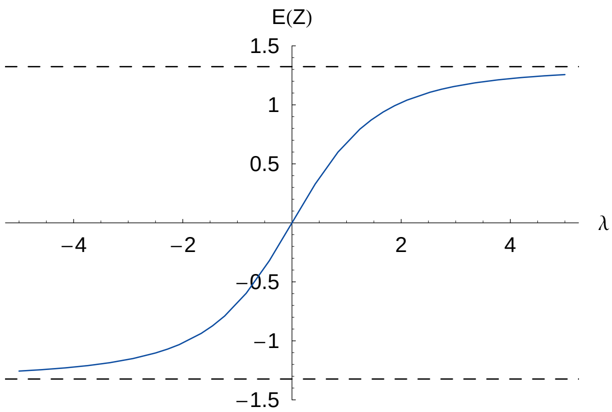


Figure 8.12: Expectation of a $\mathcal{SSN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

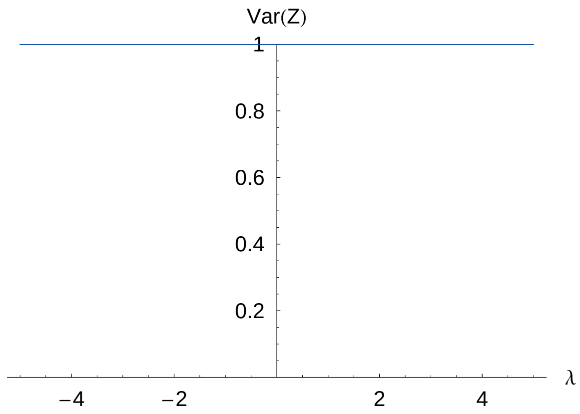


Figure 8.13: Variance of a $\mathcal{SSN}(\lambda)$ distributed random variable vs. skewness parameter

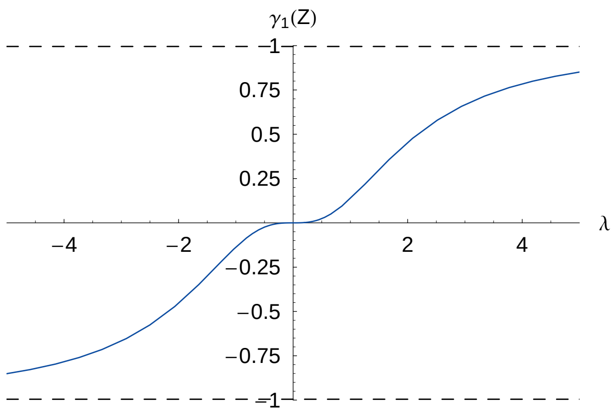


Figure 8.14: Skewness of a $\mathcal{SSN}(\lambda)$ distributed random variable vs. skewness parameter. The limits are indicated with dashed lines.

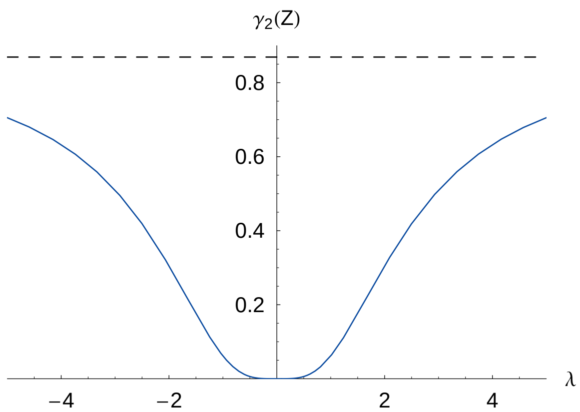


Figure 8.15: Kurtosis of a $\mathcal{SSN}(\lambda)$ distributed random variable vs. skewness parameter. The limit is indicated with a dashed line.

8.6 Further Skew Distributions

With Lemma 8.3 a general rule is given to skew a symmetric distribution that has later been used to define the skew-normal distribution. This lemma is quite universal and thus not only limited to the normal case. For instance Gupta et al. [2002] used it to define skew-uniform, t, Cauchy, Laplace, and logistic distributions. In the following we will outline with the skew-logistic and the skew-Cauchy distribution the variants most relevant to the field of evolutionary algorithms.

8.6.1 Skew-Logistic Model

With the logistic distribution the following skewed version originates.

Definition 8.4. A random variable Z with density function

$$\begin{aligned} f_{SL}(z;\lambda) &= 2f(z)F(\lambda z) \\ &= \frac{2e^{-z/\sigma}}{\sigma(1+e^{-z/\sigma})^2(1+e^{-\lambda z/\sigma})}, \quad z \in \mathbb{R}, \end{aligned} \tag{8.27}$$

where $\lambda, \sigma \in \mathbb{R}$, $\sigma > 0$,

$$f(z) = \frac{e^{-z/\sigma}}{\sigma(1+e^{-z/\sigma})^2}$$

is the logistic density and

$$F(z) = \frac{1}{1+e^{-z/\sigma}}$$

the corresponding distribution function is called a *skew-logistic* random variable with parameter λ.

Neither the moment generating function nor the characteristic function of Z has a closed form. Gupta et al. [2002] give the first four moments of Z as follows,

$$E(Z) = 2\sigma A_1, \tag{8.28a}$$

$$E(Z^2) = \frac{1}{3}(\pi\sigma)^2, \tag{8.28b}$$

$$E(Z^3) = 2\sigma^3 A_3, \tag{8.28c}$$

$$E(Z^4) = \frac{7}{15}(\pi\sigma)^4 \tag{8.28d}$$

where

$$A_i = \int_0^\infty \frac{(\ln z)^i}{((1+z)^2(1+z)^\lambda)} dz, \; i = 1, 3.$$

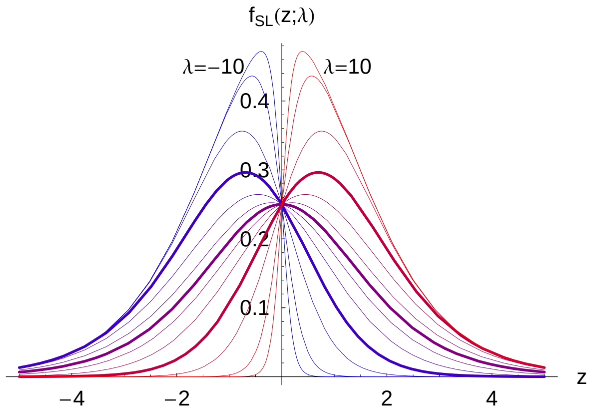


Figure 8.16: Density functions of the skew-logistic distribution for selected skewness parameters, $\lambda \in [-10, 10]$. Bold lines indicate the $f_{SL}(z; -1)$, $f_{SL}(z; 0)$, and $f_{SL}(z; 1)$ density.

8.6.2 Skew-Cauchy Model

Taking the Cauchy distribution as basis, the skewed variant is defined as follows.

Definition 8.5. A random variable Z with density function

$$\begin{aligned} f_{SC}(z;\lambda) &= 2f(z)F(\lambda z) \\ &= \frac{\sigma}{\pi(\sigma^2+z^2)}\left(1+\frac{2}{\pi}\arctan\frac{\lambda z}{\sigma}\right), \quad z \in \mathbb{R}, \end{aligned} \tag{8.29}$$

where $\lambda, \sigma \in \mathbb{R}$, $\sigma > 0$,

$$f(z) = \frac{\sigma}{\pi(\sigma^2+z^2)}$$

is the Cauchy density and

$$F(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{z}{\sigma}$$

the corresponding distribution function is called a *skew-Cauchy* random variable with parameter λ.

The moment generating function of Z does not converge and the characteristic function does not have a closed form.

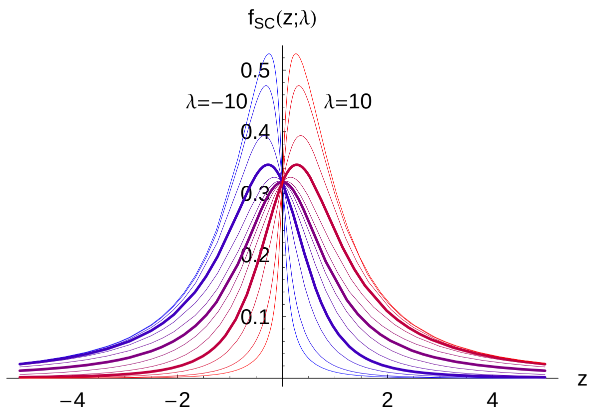


Figure 8.17: Density functions of the skew-Cauchy distribution for selected skewness parameters, $\lambda \in [-10, 10]$. Bold lines indicate the $f_{SC}(z;-1)$, $f_{SC}(z;0)$, and $f_{SC}(z;1)$ density.

9 Directed Evolution Strategies

The two previous chapters laid the theoretical fundamentals of directed mutation. Different construction principles have been expound and customizable skew distributions developed. Especially with the SN distribution a sound basis to construct a directed mutation operator is given. However, some odds and ends are sill missing. The self adaptation of the skewness parameters has to be set up and an appropriate recombination scheme has to be provided. Both will be discussed in the following sections. We are then able to formulate the entire Directed Evolution Strategy (DES) and provide it in pseudo-code. Further, some critical remarks on normalization are given. The discussion of the theory of Directed Evolution Strategies will be closed with a comparison of the different mutation operators' characteristics.

9.1 Building the Directed Mutation Operator

There are two relevant aspects concerning the extension of a mutation operator to Directed Mutation. The first one is of course the mutation distribution, the second is the necessary change in the individual's representation. The first is rather straightforward. All that has to be done is to exchange the operator's distribution by a skew variant, which means in the end to surrogate the random number generator. The representation needs to be updated since the shape parameters are endogenous strategy parameters and hence stored in the individual, see Section 4.3, *Representation*. The set of endogenous strategy parameters $\mathbf{s}$ then comprises the mutation strengths and the shape parameters, thus the state space is given by

$$\mathcal{S} = \left\{\mathbb{R}_0^+\right\}^t \times \mathbb{R}^n, \tag{9.1}$$

with $t = 1$ for isotropic mutation and $t = n$ for scaled mutation.

9.2 Self-Adaptation of the Strategy Parameters

Adaptation of the mutation strengths is done with the method already used in the scaled mutation case, i.e. with a logarithmic normal operator as described in subsection 4.8.3, *Self-Adaptation*, on page 52. Thus, the actual mutation strength σ_i is multiplied with a factor obtained by an exponential transformation of a normally distributed random number, resulting in the new mutation strength $\tilde{\sigma}_i$, cf. (4.28).

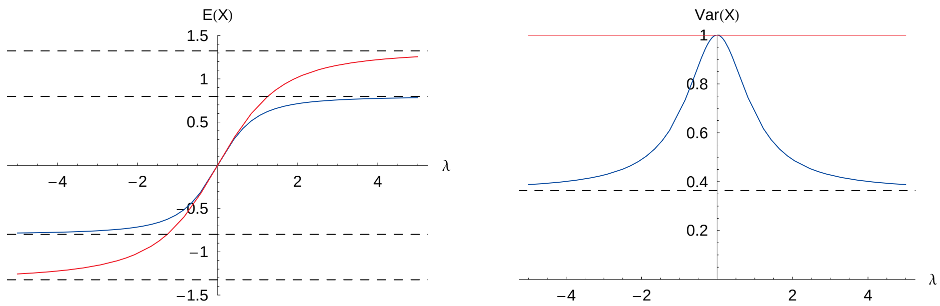
The shape parameters are modified in analogy to the object parameters since no positive values have to be preserved. However, to prevent deadlocks from overfitted shape parameters damping should be introduced. This can be done via a damping factor d which leads to

$$\tilde{\lambda}_i = (1-d)\lambda_i + z_i, \quad \text{with } z_i = \mathcal{N}(0,1) \text{ and } 0 \leq d \leq 1. \tag{9.2}$$

Besides the already introduced learning rates, the variable d is another exogenous strategy parameter. The right choice of d depends on the individual optimization problem. Experimental results show that $d \approx 0.05$ is a good starting point for most problems, see [Berlik and Reusch, 2004]. Obviously, with $d = 1$ no learning of the shape parameters takes place. In contrast, with $d = 0$ no damping takes place as it is the case for the asymmetric mutation [Hildebrand, 2001].

9.3 Normalization

As argued in subsection 5.1.1, *On the Moments and their Convergence*, the moments of skew distributions should be convergent with respect to the shape parameter. Even then there might remain some impact of the shape on the moments. To limit this impact on the variance, i.e. the mutation strength, a small spread of variance modulation was demanded, the ideal case was sketched as a shape invariant variance. Therefore Sections 7.5 and 8.5 provided with the standardized naïve skew-normal and the standardized skew-normal distribution transformed variants with shape invariant variance, depicted in Figure 9.1 for the SN and SSN distributions.



(a) Expectation (b) Variance

Figure 9.1: Expectations and variances of $\mathcal{SN}(\lambda)$ and $\mathcal{SSN}(\lambda)$ distributed random variables vs. skewness parameter. The moments of the SN and SSN distribution are depicted in blue and red, respectively. The limits are indicated with dashed lines.

While this first seems reasonable from a pure technical point of view, on second thought things look quite different, see Figure 9.2. The reason for the decreasing variance is the tendency of the SN distribution towards the half normal distribution for increasing shape parameters. Hence the width of the distribution's variance effective mass is approximately halved. This effect is balanced out by the transformation applied to the SSN distribution, resulting in a variant stretched with respect to the abscissa. However, from the phenomenological point of view, the SN distribution has exactly the shape it should have. Because we can just as well argue that the variance has to be less, since random number generation now takes place more probable in either the positive or negative domain. There however, random numbers are distributed approximately “normal”. Since the inferior domain contributes little to the variance, the overall variance decreases.

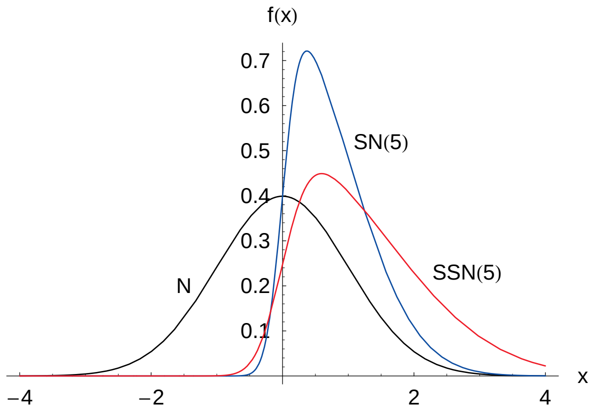


Figure 9.2: Comparison of the normal, $\mathcal{SN}(5)$, and $\mathcal{SSN}(5)$ distribution

Another problem of standardization is the increasing expectation of the distributions witch becomes obvious already in Figure 9.2 and is shown in detail in Figure 9.1(a). As argued earlier, it should like the variance be perturbed as less as possible.

In the end, everyone has to decide on his own which variant fits best to the demands of a given problem. However, anticipating one result of the empirical studies to be presented in Chapter 11, no significant performance difference can be ascertained between SN and SSN mutation operators.

9.4 Coupled Recombination

To use directed mutation also the recombination operator has to be considered. The statement by Bäck [1996] that independent recombination of object variables and strategy parameters is justified by experimental observations could not be approved for the use of directed mutations. When doing so all directed mutation variants yield significantly worse results compared to the conventional variants [Berlik, 2004b].

The reason for this could be seen in the higher grade of localization that arises from the togetherness of object variable, mutation strength, and skewness parameter, as illustrated in Figur 9.3. It is apparent that favorable mutation directions highly depend on the localization of the object variable. Mixing these information from the two points P_1 and P_2 will lead to adverse results. In fact they should be treated as a unit. Thus coupled recombination assures that the strategy parameters are chosen from the same parent where the object variable at hand is taken from when recombining parents. In the case of intermediate recombination (c.f. page 46) the same weight ought to be used for the different components.

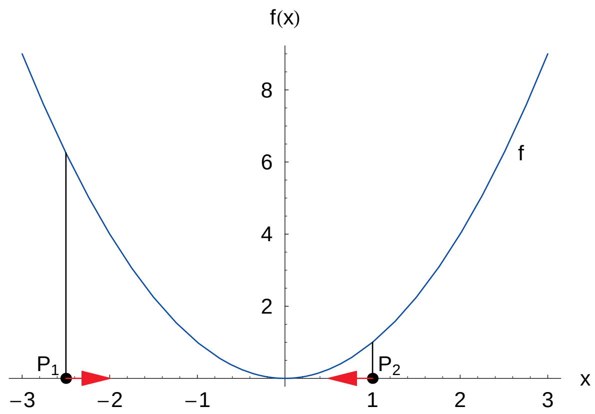


Figure 9.3: Motivation for a coupled recombination scheme with the red arrows indicating optimal mutation directions

9.5 Directed Evolution Strategies

In Section 4.2 pseudo-code of the $(\mu/\rho \overset{+}{,} \lambda)$-ES was given in Listing 4.1. Factoring in the necessary changes for the directed variant, the reworked version takes the form shown in Listing 9.1 with the differences highlighted in red. It is conspicuous how few modifications are made. Essentially, they pertain the new operators and the different representation of the individuals. Obviously, a new mutation operator for the object parameters is used (line 9) that implements skew mutations and needs for this purpose additional input in form of the shape parameters $\tilde{\mathbf{s}}_l$. These are generated together with the mutation strengths by the strategy parameter mutation, (line 8). As explained in the previous section also the recombination operator should be modified. Hence, coupled recombination is introduced, (line 7). Since the representation of the individuals now additionally contains shape information, also an adapted initialization procedure is needed, (line 3).

Listing 9.1: Pseudocode of the $(\mu/\rho \overset{+}{,} \lambda)$-DES. Differences to the conventional $(\mu/\rho \overset{+}{,} \lambda)$-ES are indicated in red.

```
begin
  g := 0;
  init( P_μ^(0) := { a_m^(0) } = { ( x_m^(0), s_m^(0), F(x_m^(0)) ) } );
  repeat
    for l:=1 to λ do begin
      R_l := reproduction( P_μ^(g), ρ );
      (s_l, x_l) := sxCoupledRecombination(R_l);
      s̃_l := sMutation(s_l);
      x̃_l := xMutation(x_l, s̃_l);
      f̃_l := F(x̃_l);
    end;
    P̃_λ^(g) := { ã_l^(g) } = { ( x̃_l, s̃_l, f̃_l ) };
    case selectionType of
      ,: P_μ^(g+1) := selection( P̃_λ^(g), μ );
      +: P_μ^(g+1) := selection( P̃_λ^(g), P_μ^(g), μ );
    end;
    g := g + 1;
  until terminationCriterion;
end;
```

9.6 Comparison

The characteristics of the different mutation operators presented before are summarized in Table 9.1 (cf. also [Berlik and Reusch, 2005]). We see that SN and SSN perform considerably better than the other operators. While all but the asymmetric mutation operator have convergent expectations and variances these two are the only with probability density functions given in closed form. Further, only these provide acceptable, simple and fast random variate generation procedures. Taking into account that during an optimization process a vast amount of random numbers has to be generated this issue is very important and hence also reflected in the point *Usefulness*. The asymmetric mutation is already unusable because of its diverging moments. The advantage of the SN and SSN distributions compared to the NSN and SNSN distributions originates from the random number generation and mathematical tractability. If a shape invariant variance is desired the SSN should be used, causing only a slight overhead compared to the SN mutation operator.

Table 9.1: Comparison of the different directed mutation operators

	Asymmetric	Naïve skew-normal	Standardized naïve skew-normal	Skew-normal	Standardized skew-normal
Convergent expectation	−	+	+	+	+
Convergent variance	−	+	+	+	+
Shape invariant variance	−	−	+	−	+
Mathematical tractability	○	○	○	+	+
Given in closed form	−	−	−	+	+
Random variate generation	○	○	○	+	+
Usefulness	−	○	○	+	+

10 Directed Covariance Matrix Adaptation-ES

So far directed mutation was realized with uncorrelated mutation models only. However, its usefulness has repeatedly been shown for both, for test function optimization as as well as for real-world scenarios, see e.g. [Berlik and Reusch, 2004, Berlik and Fathi, 2005a]. We will here also review some results in Section 11.1, *Results of the Directed Evolution Strategy* and Chapter 12, *Multiobjective Optimization of Screw-Type Machines*.

Now, as already mentioned in the introduction and Chapter 4, there are several even more powerful ES approaches. Since they rely on the flexibility of correlated mutations, their performance depends obviously highly on the choice of the covariance matrix $\mathbf{C}$ which has to be adjusted not only to the problem at hand, but also to the current state of the evolution process. Several methods have been proposed, from the self-adaptation of the mutation parameters in ES (SA-ES) [Schwefel, 1995] to the Covariance Matrix Adaptation-ES (CMA-ES) [Hansen and Ostermeier, 1996], cf. subsection 4.7.3 and Section 4.9. While the first removes the need to manually adjust the covariance matrix, the latter takes into account the history of evolution and deterministically adapts the covariance matrix from the last moves of the algorithm, thereby directing the search to use the most recent descent direction. In [Hansen et al., 2003] an advanced version of the CMA-ES is presented, that is computationally more efficient.

As it was already the case in the ES field, all present approaches use symmetric normally distributed random numbers. The aim of the sequel of this chapter is therefore to accommodate the CMA-ES with a multivariate skew-normal distribution, yielding the Directed Covariance Matrix Adaptation-ES (DCMA-ES). Recent studies have shown remarkable results. However, much further research is necessary and the results are in that sense preliminary.

A conceptually related approach, called Least-Square-CMA-ES (LS-CMA-ES), was presented by Auger et al. [2004]. It is based on quasi-Newton techniques, i.e. relying on local curvature information to find out the next points to sample. Therefore it aims at learning the local Hessian matrix by solving a linear least-square minimization problem. The solution is then found by evaluating the pseudo-inverse of this linear system. The cost of the direct computation of this pseudo-inverse by standard numerical methods is scaling as n^6, indicating already an also high effort of the approximative solution. In contrast the DCMA-ES is computationally by far less expensive.

The rest of this chapter is organized as follows: first we present a multivariate version of the skew-normal distribution and give a hint how to generate corresponding random vectors. Afterwards a possible way of integrating the concept into the CMA-ES framework is proposed. Especially the dualism *intergenera-*

tional versus *intragenerational* update of the shape vector is discussed. Finally, some experimental data concerning the intragenerational shape vector update is provided.

10.1 Multivariate Skew-Normal Distribution

An extension of the skew-normal distribution to the multivariate setting was studied by Azzalini and Dalla Valle [1996] and Azzalini and Capitanio [1999]. The multivariate skew-normal distribution represents a mathematically tractable extension of the multivariate normal density with additional parameters to regulate the skewness. The authors demonstrate that this distribution has a reasonable flexibility in real data fitting, while it maintains some convenient formal properties of the normal density. Recently, an excellent survey on skewed multivariate models has been presented by Arnold and Beaver [2002].

An n-dimensional random vector X is said to have a multivariate skew-normal distribution, denoted by $\mathcal{SN}_{\mathrm{n}}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, if it is continuous with probability density function

$$f_{SN_n}(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = 2\varphi_n(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^T(\boldsymbol{z} - \boldsymbol{\mu})) \tag{10.1}$$

where $\varphi_n(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the probability density function of the n-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{\Omega}$. $\Phi(\cdot)$ is the standard normal distribution function $\mathcal{N}(0, 1)$ and $\boldsymbol{\alpha}$ is a n-dimensional shape vector.

To generate $\mathcal{SN}_{\mathrm{n}}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ distributed random vectors their stochastic representation is used. Let Y have the probability density function $\varphi_n(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Omega})$ and W be a $\mathcal{N}(0, 1)$ distributed random variable. If

$$Z = \begin{cases} Y + \boldsymbol{\mu} & \text{if } W < \boldsymbol{\alpha}^T Y \\ -Y + \boldsymbol{\mu} & \text{otherwise,} \end{cases} \tag{10.2}$$

then $Z \sim \mathcal{SN}_{\mathrm{n}}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, see e.g. [Wang et al., 2004a].

10.2 Update of the Shape Vector

In building a DCMA-ES we are faced with principally the same problems that had to be tackled for DES algorithms, cf. Chapter 9. An appropriate multivariate distribution has to be provided and the self adaptation of the shape parameters has to be set up. The elementary CMA-ES itself has already been discussed in Section 4.9, *Covariance Matrix Adaptation-ES*, and with the multivariate skew-normal distribution the distribution is also given. Hence, only a method for the self adaptation of the skewness parameters is missing. Several techniques to realize this are conceivable. First, self adaptation like in conventional ESs can be used. As the CMA-ES already estimates the step size and the covariance matrix we can further think of an estimation approach for the shape parameters. Last, using the peculiarity of the CMA-ES that per definition a global intermediate recombination is applied (see Section 4.6), also an intragenerational update is possible.

In short, the first approach yields no benefit. Exhaustive tests showed, that self adaptation is too inert to follow the relatively fast proceeding DCM-ES.

As an estimation approach an ad hoc implementation comparable to the mechanics of evolution path calculation was used to adjust the shape vector, cf. (4.42). Shape control then reads

$$\boldsymbol{p}_\alpha^{(g+1)} = (1 - c_\alpha)\boldsymbol{p}_\alpha^{(g)} + \sqrt{c_\alpha(2 - c_\alpha)\mu_{\text{eff}}}\,\boldsymbol{C}^{(g)^{-\frac{1}{2}}} \frac{\boldsymbol{m}^{(g+1)} - \boldsymbol{m}^{(g)}}{\alpha^{(g)}} \tag{10.3}$$

with learning rate

$$c_\alpha = \frac{\mu_{\text{eff}} + 2}{n + \mu_{\text{eff}} + 3} \tag{10.4}$$

and all other constants as given by Hansen and Kern [2004]. Although the learning rate was altered over the whole $[0, \ldots, 1]$ range, no satisfying results were obtained during the test runs.

Further, conventional estimation methods of the skewness parameter have been investigated. However, regression models proposed e.g. by Ferreira and Steel [2004] or estimation as discussed by Gupta and Gupta [2004] are not applicable due to the small sample volume. In fact, by magnitudes too few samples are available.

The possibility to replace respectively supplement the conventional *intergenerational* shape update with an *intragenerational* one is discussed next. The fundamental idea is to use the informational gain arising from the creation of every new sample. This is possible because all offspring originate from the same recombined parent, namely the distribution's mean. Thus, all solutions and their directional information can be compared with the same point. So we track the fitness of every generated offspring within a descendant generation sequence and adapt the shape vector accordingly on-the-fly.

One appropriate heuristic is given as follows: calculate the normalized direction vector from the mean of the current distribution to the actual offspring. If the fitness of this offspring is better than the mean fitness, factor the direction vector into the shape vector. Otherwise take the opposite direction. Additionally, the fitness ratio is weighted exponentially and with the dimension. The definition of the update vector thus reads

$$\boldsymbol{u} = n \exp\left(\frac{\text{fit}\,\boldsymbol{x}}{\text{fit}\,\boldsymbol{x}_{\text{mean}}}\right) \frac{\boldsymbol{x} - \boldsymbol{x}_{\text{mean}}}{|\boldsymbol{x} - \boldsymbol{x}_{\text{mean}}|} \tag{10.5}$$

where n is the dimension, $\boldsymbol{x}$ and $\boldsymbol{x}_{\text{mean}}$ are the actual individual and the mean used to generate the offspring, respectively, and the function $\text{fit}(\cdot)$ gives the fitness of a sample. The intragenerational update of the shape vector $\boldsymbol{p}_\alpha^{[l]}$ depends on the actual individual's fitness.

$$\boldsymbol{p}_\alpha^{[l]} = \begin{cases} \boldsymbol{p}_\alpha^{[l-1]} + \boldsymbol{u} & \text{if } \operatorname{fit} \boldsymbol{x} \succ \operatorname{fit} \boldsymbol{x}_{\text{mean}} \\ \boldsymbol{p}_\alpha^{[l-1]} - \boldsymbol{u} & \text{else} \end{cases} \tag{10.6}$$

with $l \in [1, \ldots, \lambda]$ and $\boldsymbol{p}_\alpha^{[0]} = \mathbf{0}$.

More systematically, first we need to extract the directional information of a solution with respect to the generating distribution's mean in a vector called update vector. Subsequently this vector is optionally weighted, e.g. according to the relative fitness of the solution, and in turn factored in the distribution's shape vector. These steps are described in detail in the next subsections where also the concept of dimensional scaling is outlined and simulation results are provided.

10.2.1 Update Vector

For the shape vector update the directions of solutions with respect to the distribution's mean have to be extracted. Therefore, two alternative update vectors are examined. The *differential update*,

$$\boldsymbol{u}_{\text{dif}}(\boldsymbol{x}_l) = \boldsymbol{x}_l - \boldsymbol{x}_{\text{mean}}, \quad l \in \{1, \ldots, \lambda\}, \tag{10.7}$$

and the *normed update*

$$\boldsymbol{u}_{\text{norm}}(\boldsymbol{x}_l) = \frac{\boldsymbol{u}_{\text{dif}}}{|\boldsymbol{u}_{\text{dif}}|} = \frac{\boldsymbol{x}_l - \boldsymbol{x}_{\text{mean}}}{|\boldsymbol{x}_l - \boldsymbol{x}_{\text{mean}}|}, \quad l \in \{1, \ldots, \lambda\}, \tag{10.8}$$

with $\boldsymbol{x}_{\text{mean}}$ denoting the mean of the selected solutions calculated by the CMA-ES and $\boldsymbol{x}_l$ one of the λ new sampled solutions, respectively.

As (10.7) simply calculates the difference of the two points, not only the direction information is embodied in the update vector $\boldsymbol{u}_{\text{dif}}$ but also a weighting by means of the distance. This might be undesired as the distance measure carries no information with respect to the solution's fitness. Hence, with equation (10.8) a normed alternative is suggested.

10.2.2 Weighting

The weighting determines the way the directional information of the update vector is regarded in the shape vector. Several strategies are conceivable whereof some are presented next. In general, the weighting cannot only specify the absolute impact a specific update vector will have but also if inferior solutions are to be ignored or may be factored in with their inverted update vectors.

Therefore, the fitness ratio of the sampled solution and the mean of the selected solutions is calculated,

$$\eta = \frac{\operatorname{fit} \boldsymbol{x}}{\operatorname{fit} \boldsymbol{x}_{\text{mean}}}. \tag{10.9}$$

If $\eta > 1$ the sampled solution is better than the given mean and we should take advantage of its directional information. In the case $\eta < 1$ the sampled solution is weaker than the given mean. Its directional information can be handled in different ways, as presented in the following paragraphs.

Selective Weighting with the Heaviside Step Function

The simplest weighting function can be set up using a translated Heaviside step function,

$$f_w(\eta) = \mathrm{H}(\eta - 1) = \begin{cases} 0: & \eta \leq 1 \\ 1: & \eta > 1. \end{cases} \tag{10.10}$$

Then only solutions better than the distribution's mean are regarded and no further weighting takes place, see Figure 10.1(a).

Constant Weighting with the Signum Function

To factor in also inferior solutions, a translated signum function can be used,

$$f_w(\eta) = \mathrm{sign}(\eta - 1) = \begin{cases} -1: & \eta < 1 \\ 0: & \eta = 1 \\ 1: & \eta > 1. \end{cases} \tag{10.11}$$

Then inferior solutions with respect to the distribution's mean are weighted with the factor -1, superior ones with the factor 1 respectively, see Figure 10.1(b). Of course (10.11) could have been also defined by the Heaviside step function.

Exponential Weighting

To introduce fitness dependent weights, an exponential function can be used,

$$f_w(\eta) = \exp(\eta - 1). \tag{10.12}$$

Thereby the weighting of points increases with their relative fitness. Superior points are weighted with a factor $f_w > 1$ while inferior points are weighted with a factor $0 < f_w < 1$, see Figure 10.1(c).

Translated Exponential Weighting

The exponential weighting assigns positive weights to all solutions and varies only the amount of the weight, cf. (10.12). Introducing negative factors for inferior points will result in factoring in the inverse direction of these points instead. Therefore, the following translated exponential function can be used,

$$f_w(\eta) = \exp(\eta - 1) - 1. \tag{10.13}$$

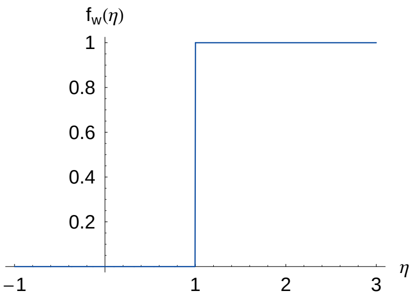
Superior points are assigned a factor $f_w > 0$ with the weights increasing with their relative fitness. Inferior points are weighted with a factor $-1 < f_w < 0$, see Figure 10.1(d).

Discontinuous Exponential Weighting

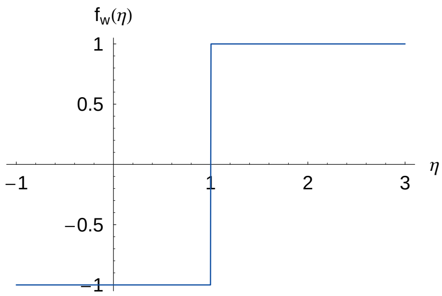
We can argue that with the above concept most solutions will be regarded not at all, since most of them will have fitness comparable to that of the distribution's mean and thus are assigned a factor about 0. To overcome this we can introduce a discontinuity into the exponential weighting, leading to the discontinuous exponential weighting,

$$f_w(\eta) = \operatorname{sign}(\eta - 1)\exp(\eta - 1). \tag{10.14}$$

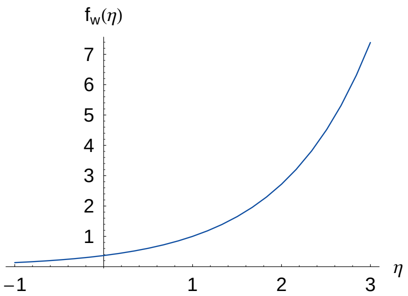
Then, superior points are assigned a factor $f_w > 1$, increasing dependent on their relative fitness and inferior points are weighted with a factor $-1 < f_w < 0$, where their consideration fades away for decreasing fitness, see Figure 10.1(e).



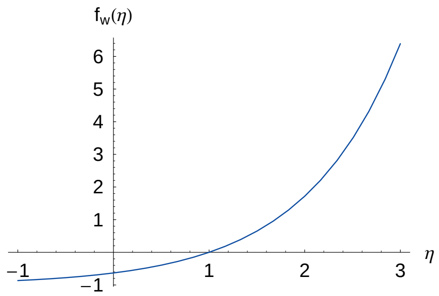
(a) Heaviside step function



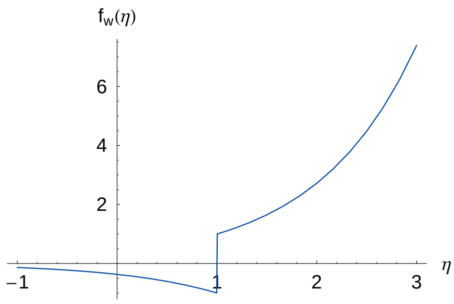
(b) Signum function



(c) Exponential function



(d) Translated exponential function



(e) Discontinous exponential function

Figure 10.1: Weighting functions for the intragenerational update. Plotted are the weights to be assigned versus the fitness ratio η

10.2.3 Dimensional Scaling

A last modification concerning the shape vector is the dimensional scaling. It turned out that changes of the update vector have to be rather drastic to yield an effect. It has to be kept in mind that processing of the CMA-ES is somewhat different to that of conventional ESs. Whereas the latter proceed relatively slowly via small variations and there remains some time to adapt parameters, the CMA-ES really hastens through the search space. Changes have to be done ad-hoc and vehement, otherwise the local topology might have changed already. Therefore, the length of the (unit) shape vector is multiplied with a factor n, equaling the number of dimensions.

10.2.4 Simulations with Different Update Strategies

Several intragenerational shape update strategies as combinations of the above methods have been investigated on a number of different test functions, see Figure 10.3 and Table 10.1. Investigated is only the performance gain that can be reached during the generation sequence of the offspring within single generations. Hence, the success rate of the generated samples is measured, i.e. the percentage of samples being superior compared with the distribution's mean. Therefore, an initial sample is generated at random in $[-1, 1]^n$ with n denoting the dimension. Then the shape vector is updated accordingly and subsequently further samples are generated continuing updating the shape vector. Experiments are done with {2, 5, 10, 20, 40, 80}-dimensional functions and with {1, 2, 3, 4, 5, 7, 10, 15, 20, 30, 40, 60, 80} generated samples in turn per experiment. The recommended number of samples by Hansen [2005] for the CMA-ES is $\lambda = 4 + \lfloor 3 \log n \rfloor$ which is illustrated in Figure 10.2. For less than 40 dimensions 2000 runs are performed for each combination, otherwise 1000 runs. Note that intergenerational update is of no importance in this scenario.

Table 10.1: Test functions

Name	Function
Sphere	$f_{\text{sphere}}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$
Schwefel	$f_{\text{Schwefel}}(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$
Cigar	$f_{\text{cigar}}(\boldsymbol{x}) = x_1^2 + \sum_{i=2}^{n} (1000 x_i)^2$
Tablet	$f_{\text{tablet}}(\boldsymbol{x}) = (1000 x_1)^2 + \sum_{i=2}^{n} x_i^2$
Ellipsoid	$f_{\text{elli}}(\boldsymbol{x}) = \sum_{i=1}^{n} \left(1000^{\frac{i-1}{n-1}} x_i \right)^2$
Parabolic ridge	$f_{\text{parabR}}(\boldsymbol{x}) = -x_1 + 100 \sum_{i=2}^{n} x_i^2$
Sharp ridge	$f_{\text{sharpR}}(\boldsymbol{x}) = -x_1 + 100 \sqrt{\sum_{i=2}^{n} x_i^2}$
Rosenbrock	$f_{\text{Rosen}}(\boldsymbol{x}) = \sum_{i=2}^{n-1} \left(100 (x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right)$

Dim.	Offspring
2	6
5	8
10	10
20	12
40	15
80	17

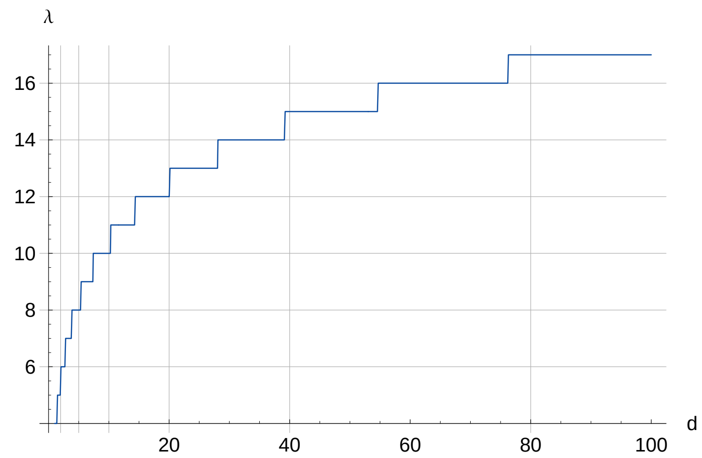


Figure 10.2: Recommended offspring per dimension for the CMA-ES

Simulation Results

A legend for the simulation results shown in the following Figures 10.4–10.19 is given in Figure 10.3. Tables with some statistics, i.e. means, standard deviations, and medians of the numerical values of these plots can be found in Appendix D.

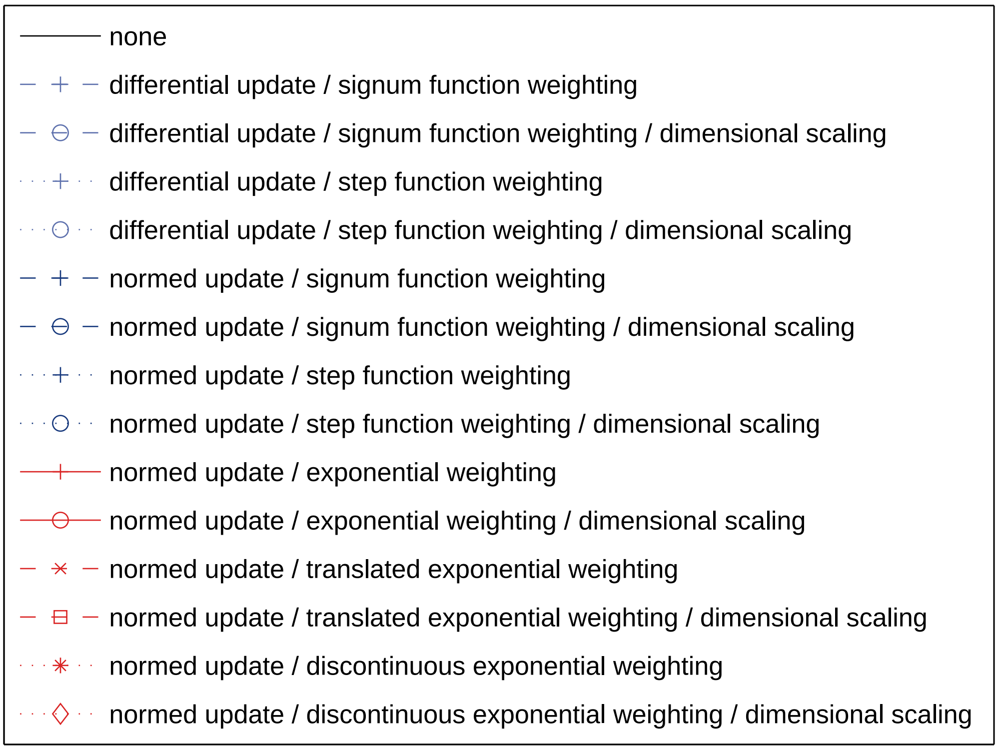
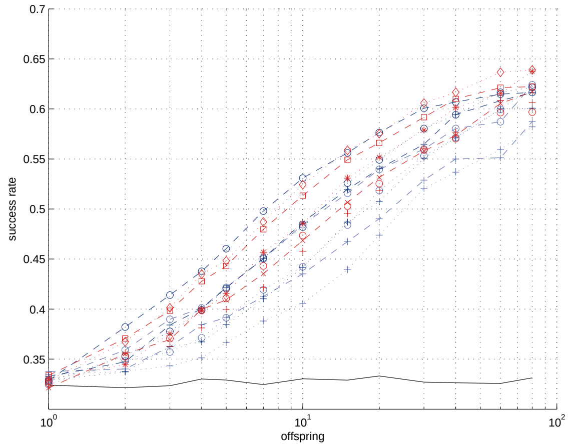
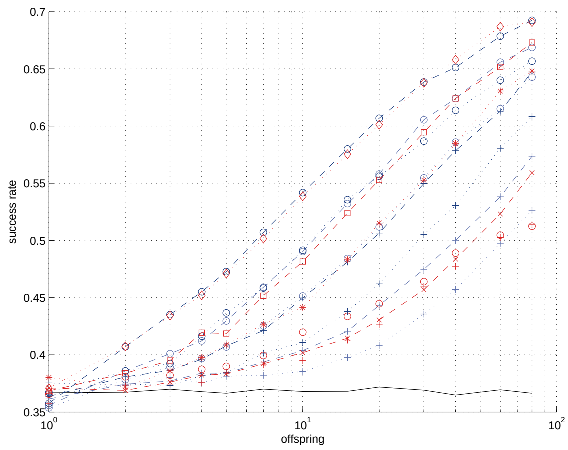


Figure 10.3: Legend for the simulation results of the shape vector's intragenerational update

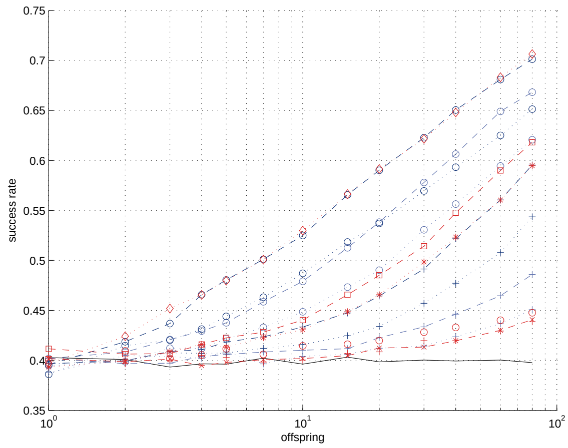
Sphere Function



(a) 2-dimensional

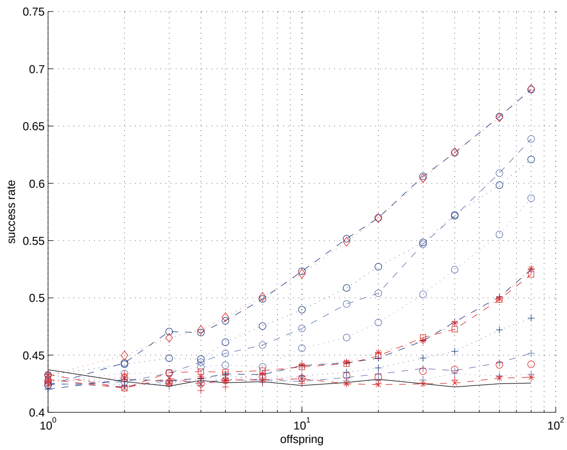


(b) 5-dimensional

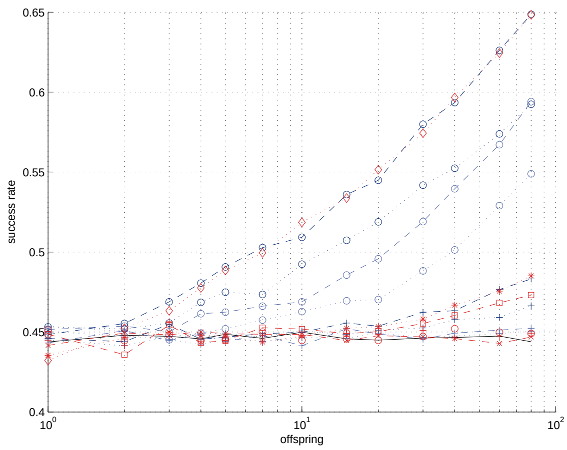


(c) 10-dimensional

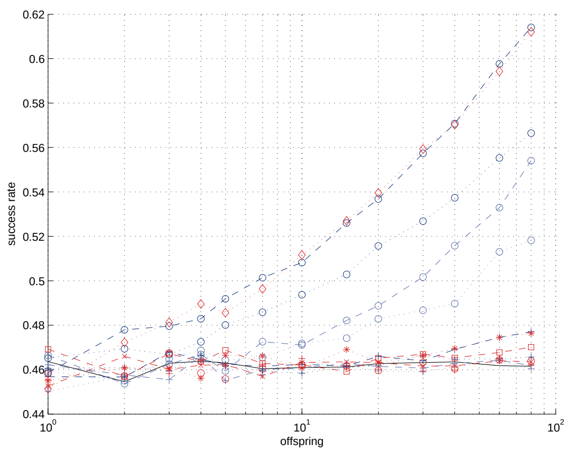
Figure 10.4: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional sphere functions



(a) 20-dimensional



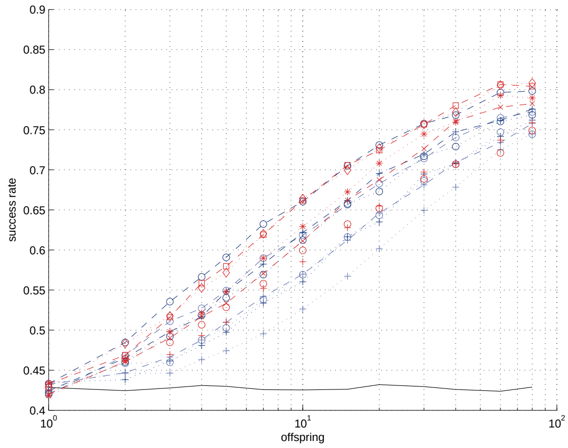
(b) 40-dimensional



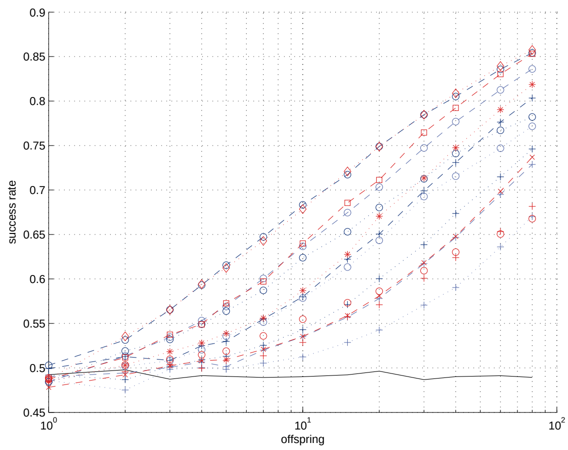
(c) 80-dimensional

Figure 10.5: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional sphere functions

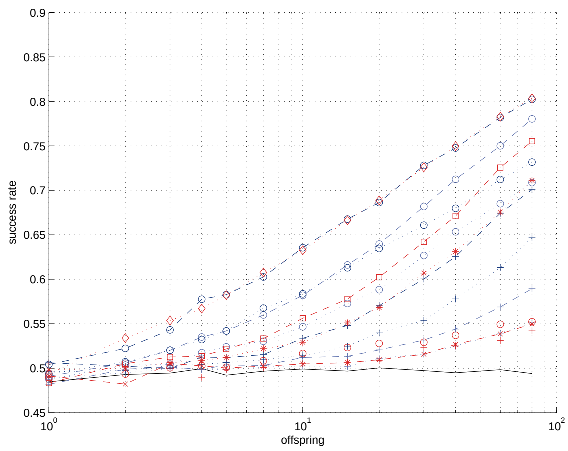
Schwefel's Function



(a) 2-dimensional

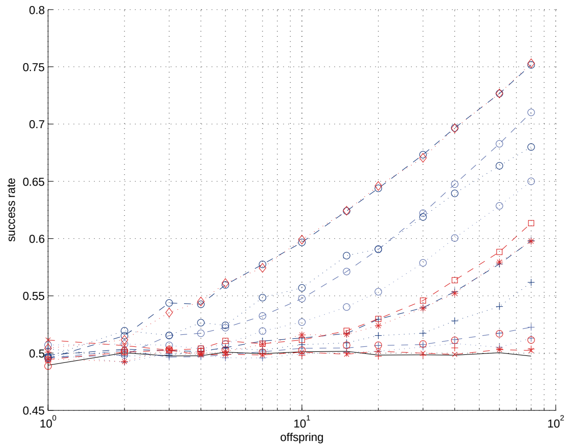


(b) 5-dimensional

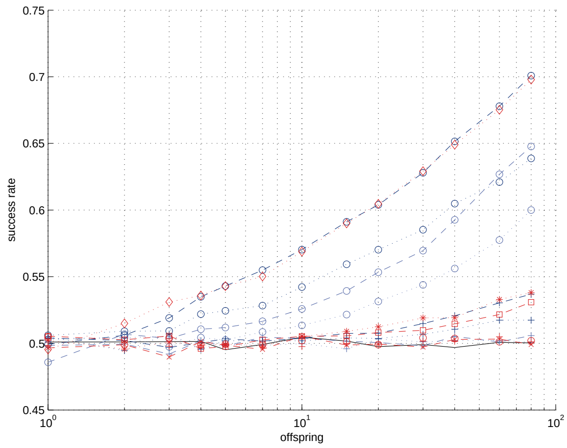


(c) 10-dimensional

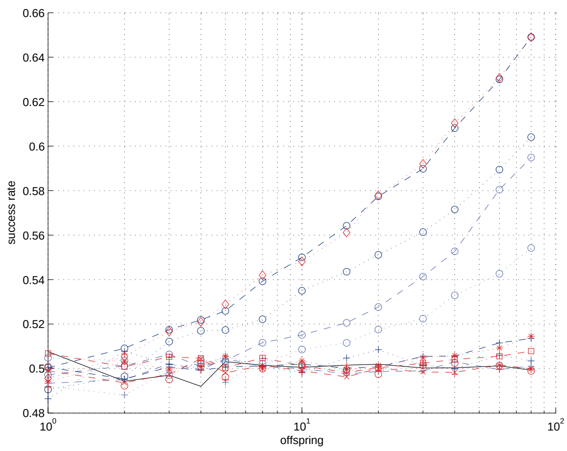
Figure 10.6: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional Schwefel functions



(a) 20-dimensional



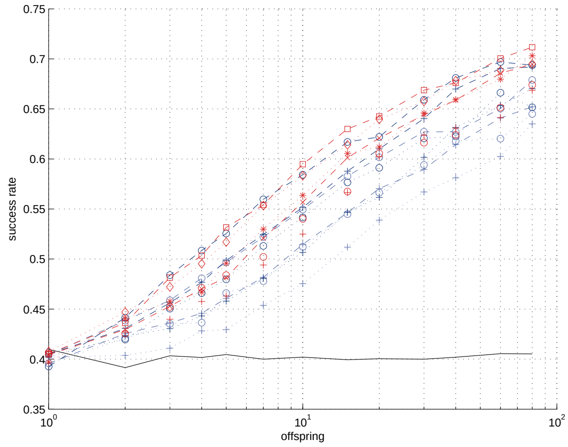
(b) 40-dimensional



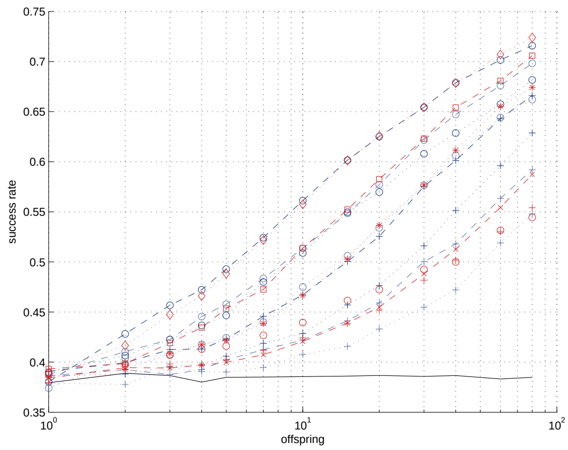
(c) 80-dimensional

Figure 10.7: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional Schwefel functions

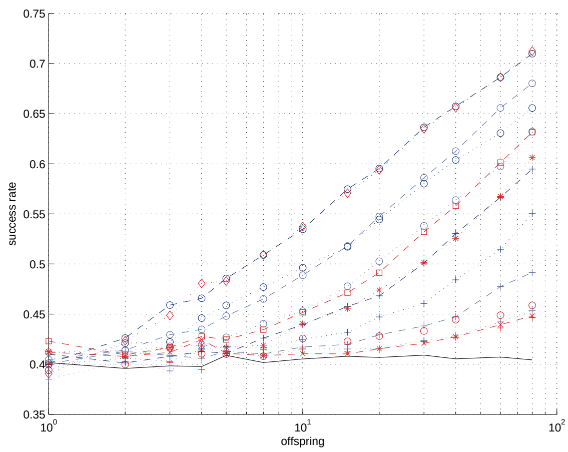
Cigar Function



(a) 2-dimensional

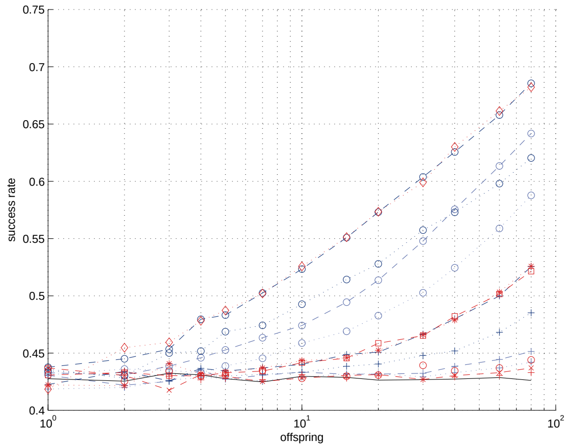


(b) 5-dimensional

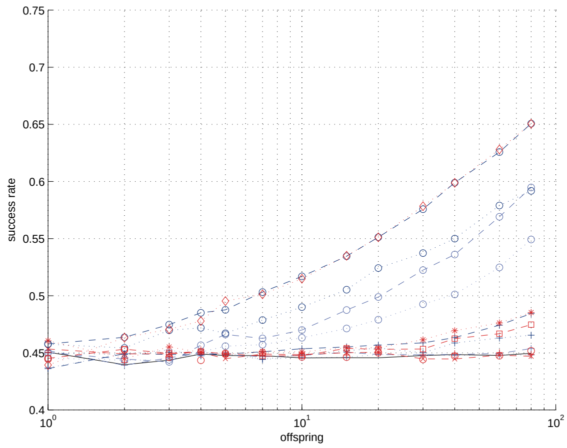


(c) 10-dimensional

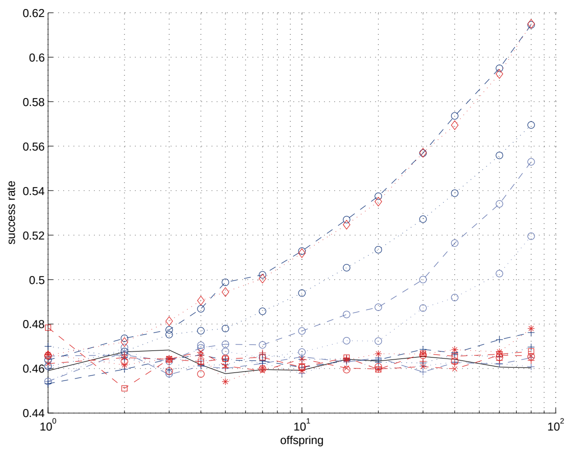
Figure 10.8: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional cigar functions



(a) 20-dimensional



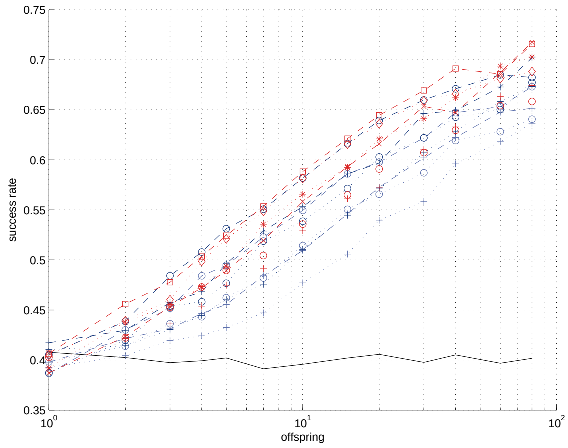
(b) 40-dimensional



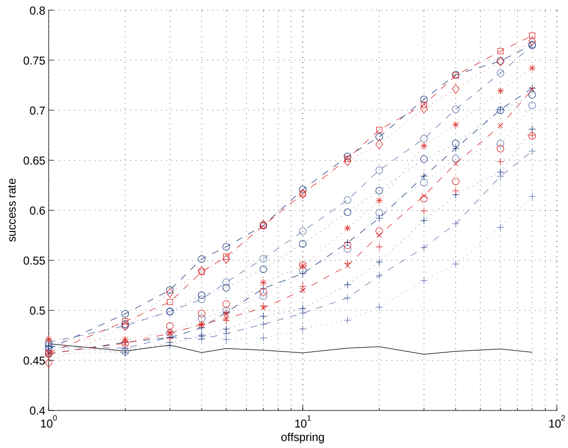
(c) 80-dimensional

Figure 10.9: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional cigar functions

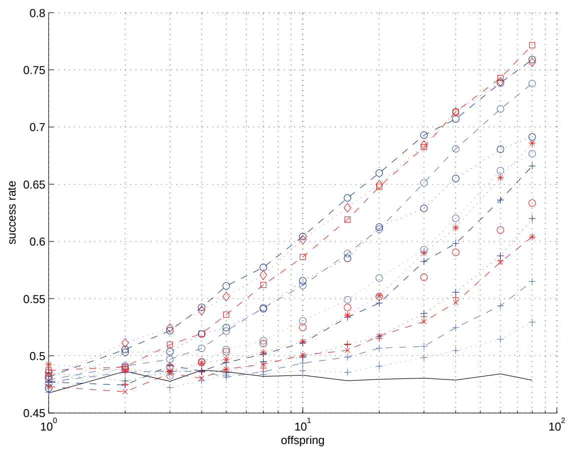
Tablet Function



(a) 2-dimensional

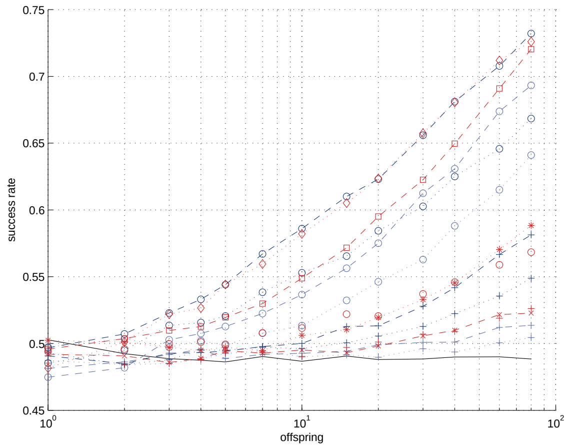


(b) 5-dimensional

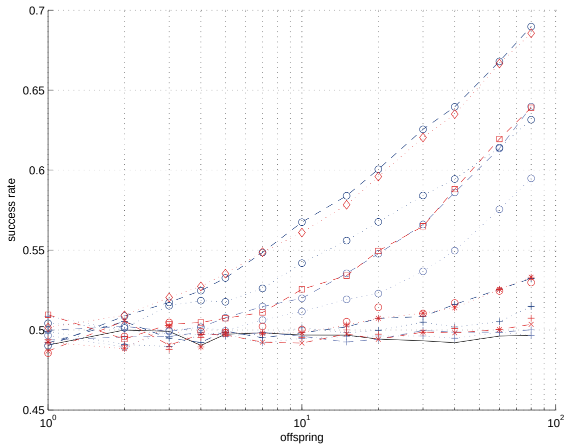


(c) 10-dimensional

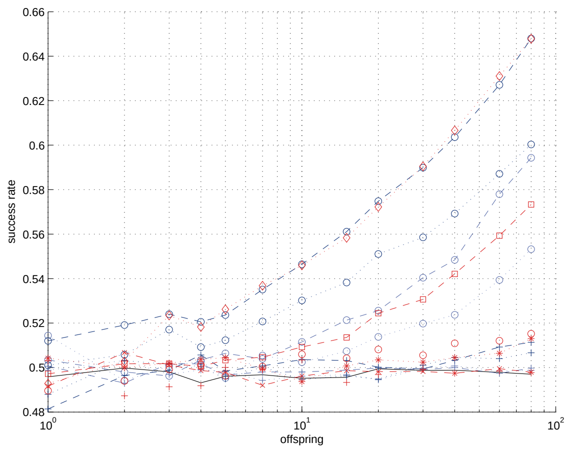
Figure 10.10: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional tablet functions



(a) 20-dimensional



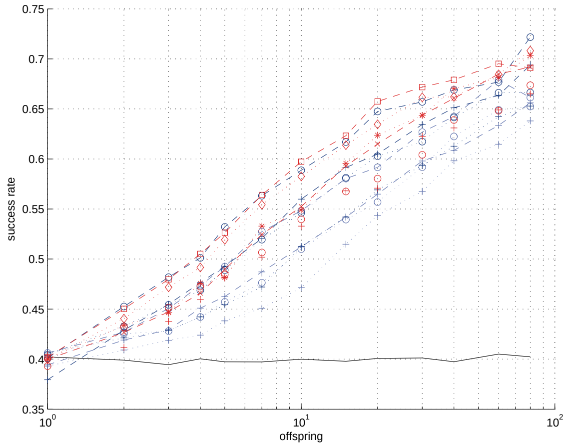
(b) 40-dimensional



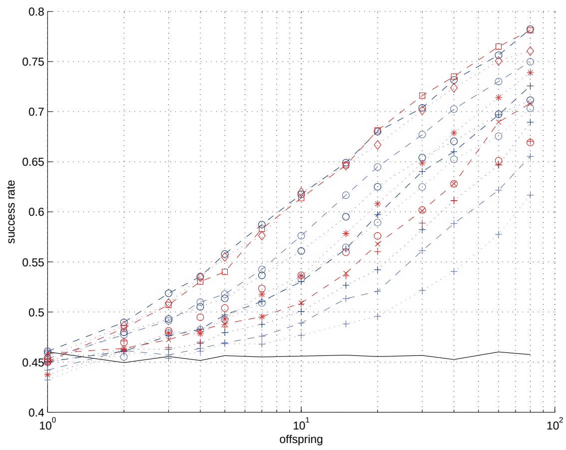
(c) 80-dimensional

Figure 10.11: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional tablet functions

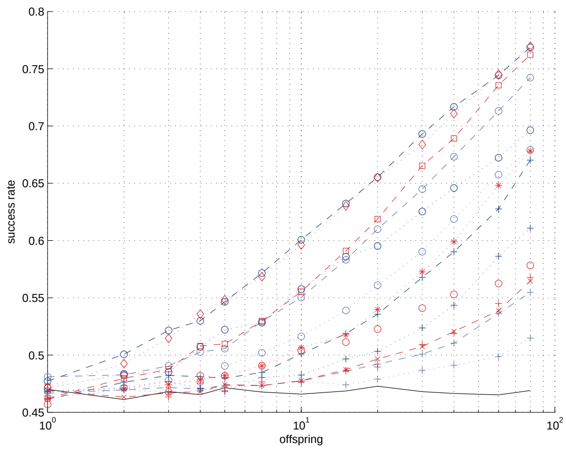
Ellipsoid Function



(a) 2-dimensional

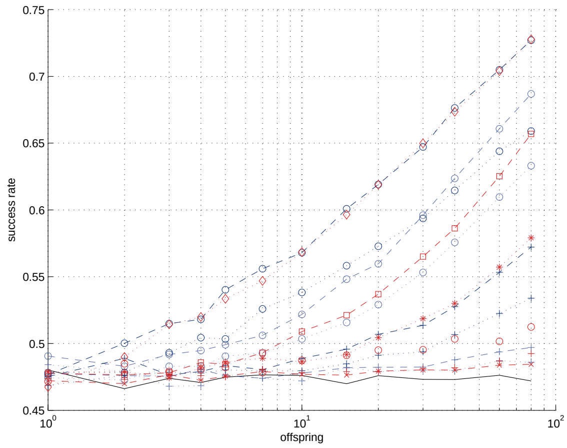


(b) 5-dimensional

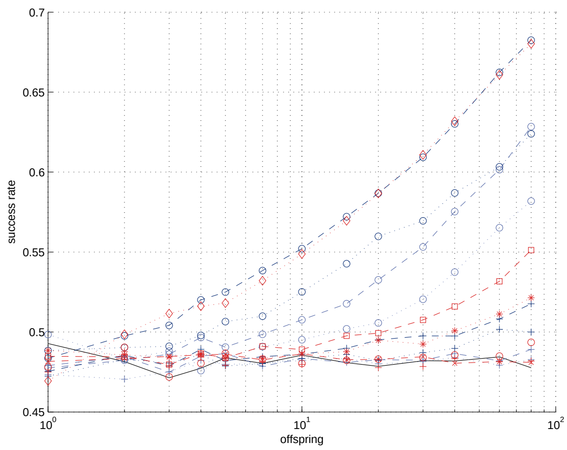


(c) 10-dimensional

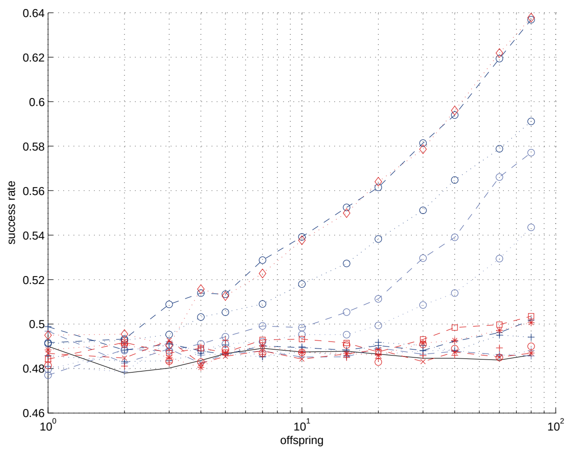
Figure 10.12: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional ellipsoid functions



(a) 20-dimensional



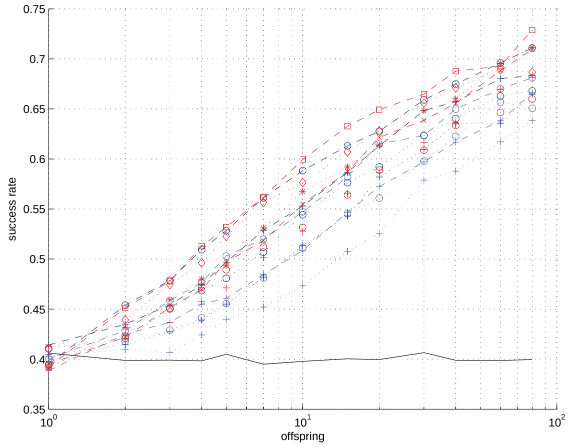
(b) 40-dimensional



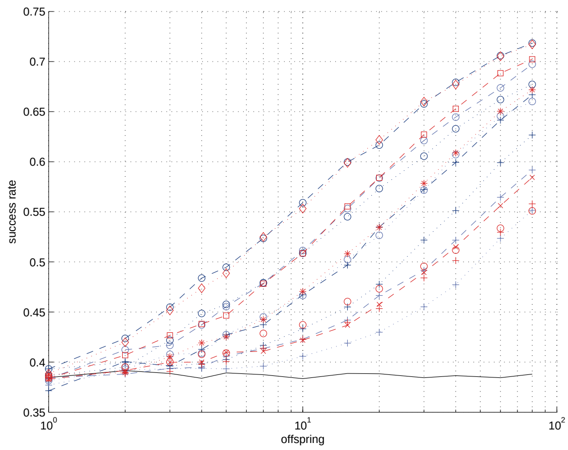
(c) 80-dimensional

Figure 10.13: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional ellipsoid functions

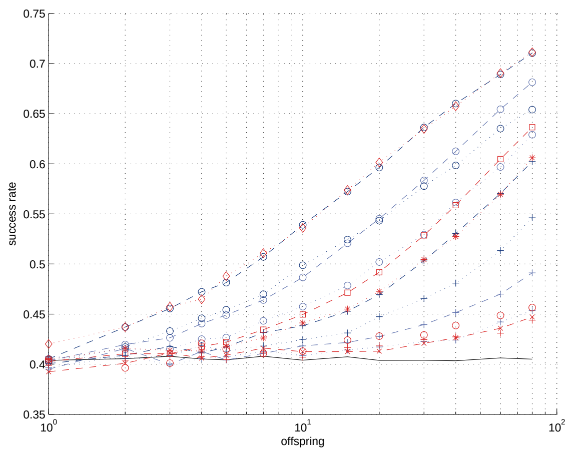
Parabolic Ridge Function



(a) 2-dimensional

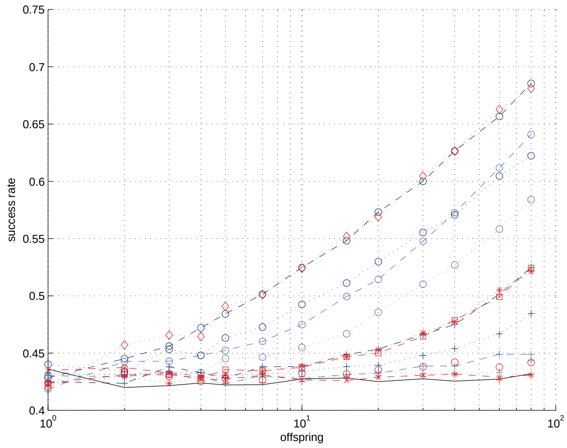


(b) 5-dimensional

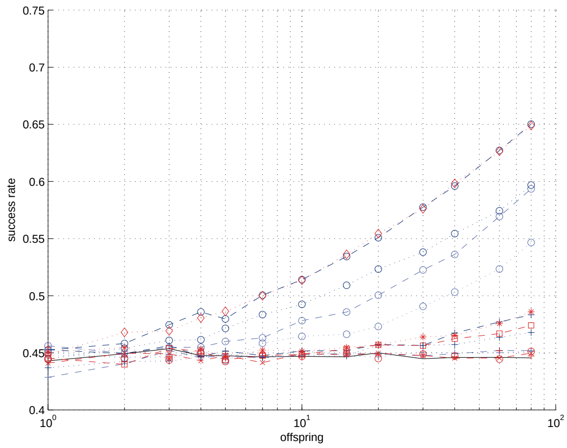


(c) 10-dimensional

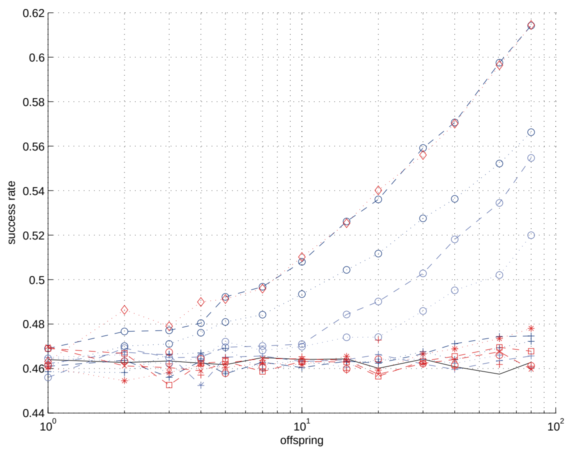
Figure 10.14: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional parabolic ridge functions



(a) 20-dimensional



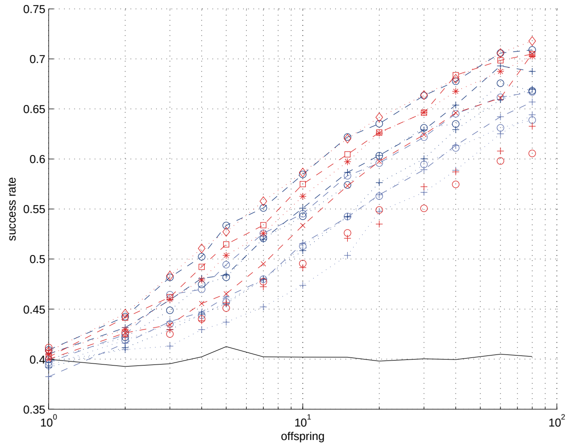
(b) 40-dimensional



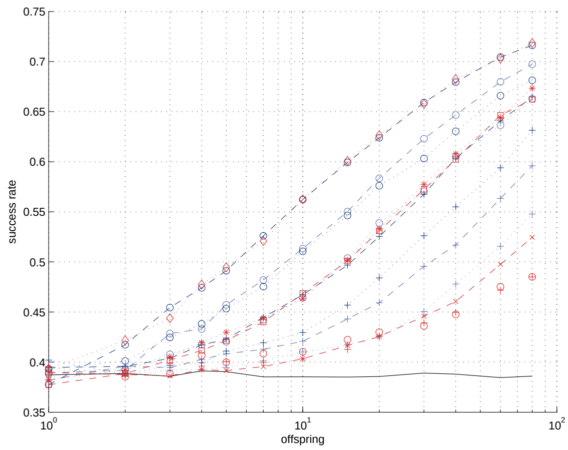
(c) 80-dimensional

Figure 10.15: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional parabolic ridge functions

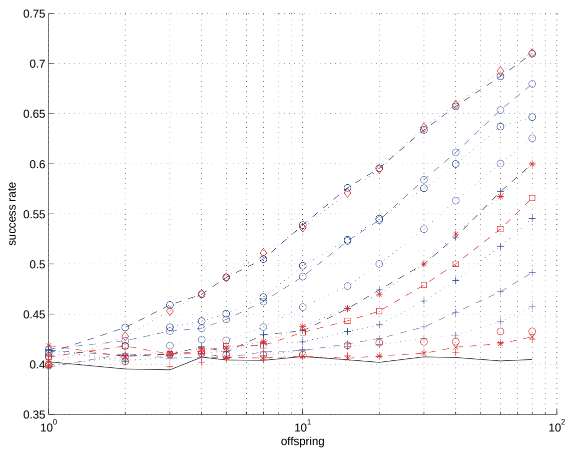
Sharp Ridge Function



(a) 2-dimensional

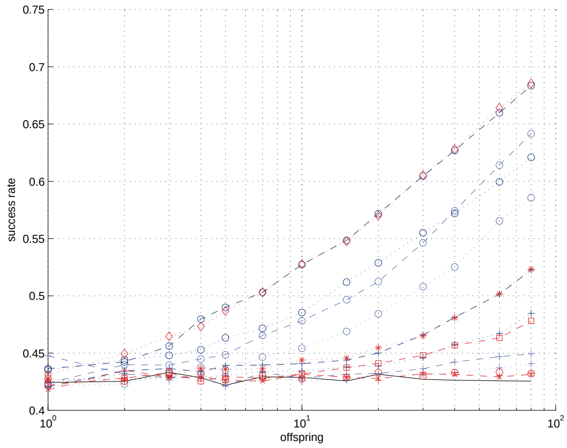


(b) 5-dimensional

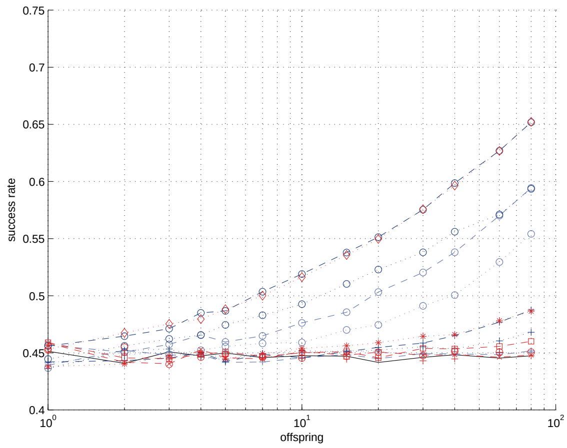


(c) 10-dimensional

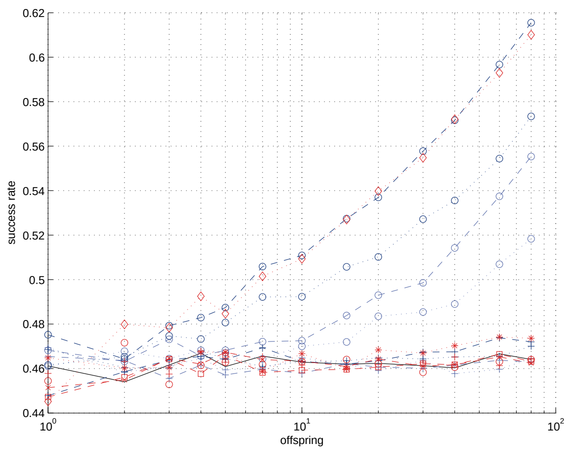
Figure 10.16: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional sharp ridge functions



(a) 20-dimensional



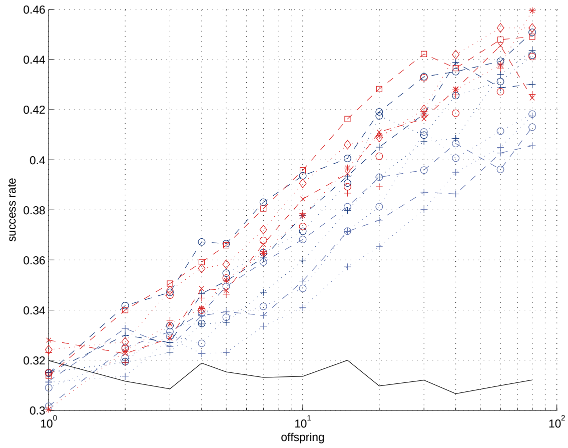
(b) 40-dimensional



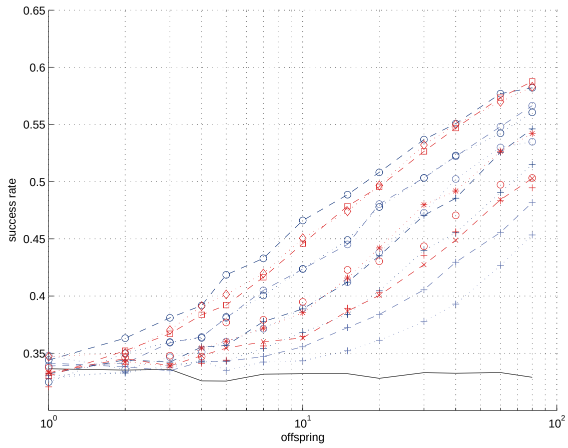
(c) 80-dimensional

Figure 10.17: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional sharp ridge functions

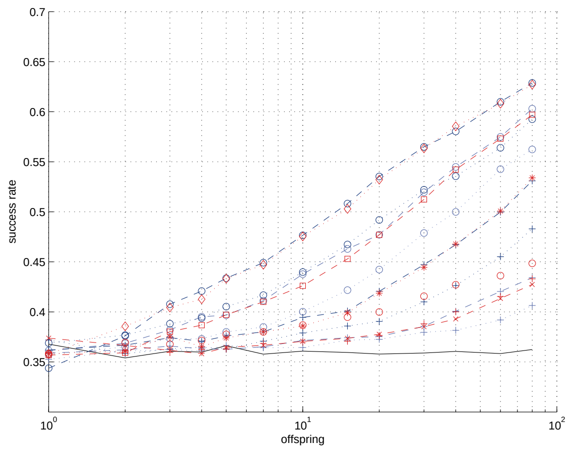
Rosenbrock's Function



(a) 2-dimensional

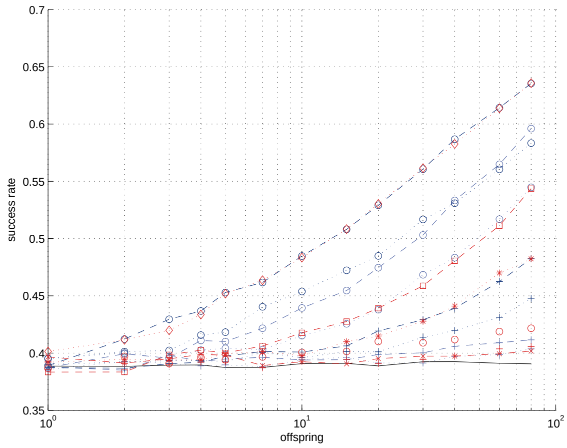


(b) 5-dimensional

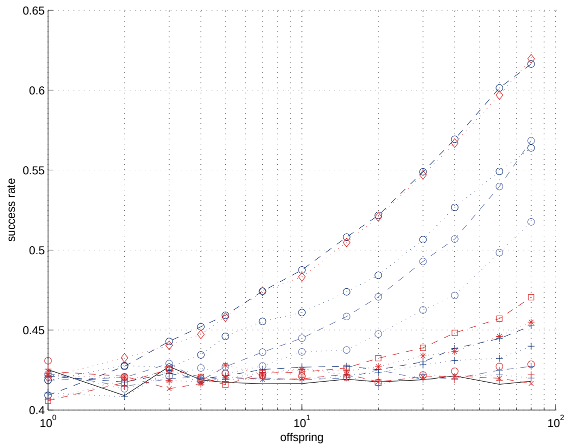


(c) 10-dimensional

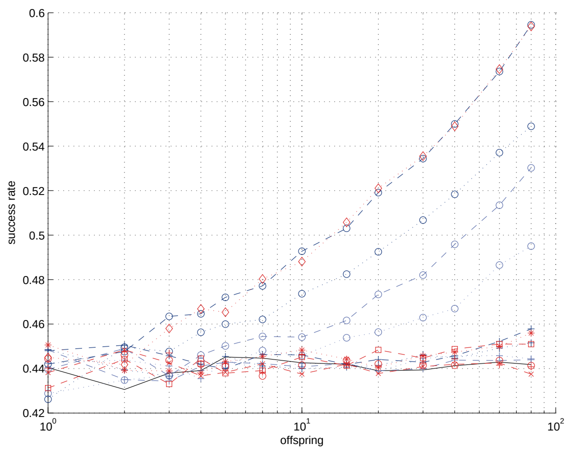
Figure 10.18: Success rate versus number of generated offspring λ on $\{2, 5, 10\}$-dimensional Rosenbrock functions



(a) 20-dimensional



(b) 40-dimensional



(c) 80-dimensional

Figure 10.19: Success rate versus number of generated offspring λ on $\{20, 40, 80\}$-dimensional Rosenbrock functions

10.2.5 Discussion of the Results

First, as expected the variant without intragenerational update shows a constant success rate, independent of the number of generated offspring. This holds for all dimensions on all functions. The variants with intragenerational update are in general more successful with increasing number of offspring.

On f_{sphere} in the 2-dimensional case the success rate approximately doubles from 0.33 for 1 offspring up to 0.64 for 80 offspring. In the case of 6 offspring as to be used by standard CMA-ES, there is roughly a 47% increase to 0.47. All versions with intragenerational update perform comparably. In the 80-dimensional case two versions perform clearly best, the normed update with discontinuous exponential weighting and dimensional scaling respectively the normed update with signum function weighting and dimensional scaling. The increase of the success rate from 1 to 80 offspring is about 33% from 0.46 to 0.61. There is a second group of update versions that perform second best: the normed update with step function weighting, the differential update with step function weighting and also with signum function weighting, all with additional dimensional scaling. The gain is about 17%. All other versions perform not significantly better than the variant without any intragenerational update. Regarding the more relevant case of 17 offspring, the standard of CMA-ES for this dimensionality, the two best strategies yield a 15% gain.

In general, this behavior can be observed on the other functions, too. In the 2-dimensional case the different variants perform comparably, the gain for 6 offspring varies from 19% on f_{Rosen} to 46% on f_{Schwefel}, with about 37% for the other functions. With increasing dimensionality, the differences in performance become more pronounced. Lastly, in the 80-dimensional case there emerge three efficiency groups: best performing methods are the normed update with discontinuous exponential weighting and dimensional scaling and the normed update with signum function weighting and dimensional scaling. Both perform almost equal, about 30% to 34% better than without intragenerational update for 80 offspring. Regarding the case of 17 offspring, the gain is roughly 14%. The group of second best performing methods comprises the normed update with step function weighting, the differential update with step function weighting as well as with signum function weighting, all with additional dimensional scaling. The gain is about the half of the two superior versions. Only on f_{tablet} the normed update with translated exponential weighting and dimensional scaling can reach this group, too. All other versions perform on all functions not significantly better than it is the case without intragenerational update.

The last observation: all well performing variants apply dimensional scaling, which supports the assertion that changes of the shape vector have to be rather drastic.

10.3 Conclusions

The DCMA-ES algorithm has been sketched and the utilized multivariate skew-normal distribution been introduced. Then the intragenerational shape vector update has been discussed whereof two variants have been identified to be clearly outperforming, the

- normed update with signum function weighting and dimensional scaling, and the
- normed update with discontinuous exponential weighting and dimensional scaling.

A comparison of both shows, that they are relatively equal, cf. Figures 10.1(b) and 10.1(e). Characteristically for both is the discontinuity in weighting between superior and inferior solutions. Around this discontinuity they also resemble in the weight they assign. Compared to the Heaviside step function they also factor in inferior solutions, adjusted with a negative weight.

The presented results clearly reveal the potential of the DCMA-ES. Although the enhancement of directed mutation causes only very small overhead, using the intragenerational shape update there is a gain in the success rate of created samples of about 14%. Compared to the CMA-ES no additional function evaluations are performed and in essence, only a further scalar product has to be calculated.

Regarding the presented results, it has to be kept in mind that intergenerational adaptation of the shape vector is mainly left open and the corresponding learning rate has to be investigated, too. Thus, much work is left to be done to tune the DCMA-ES. However, first results of entire simulations with the DCMA-ES on several test functions left us optimistic about the potential of this approach (see Section 11.2).

Part III

Empirical Studies

11 Simulation Results on Single-Objective Problems

To get an idea about of the capability of the Directed Evolution Strategies discussed just before, results of several simulations are presented next. This chapter is dedicated to single-objective problems solely; multiobjective optimization is postponed to the following one. The two main classes of DESs are treated separately, the directed variants of conventional ESs in Section 11.1 and the Directed Covariance Matrix Adaptation-ES in Section 11.2.

11.1 Results of the Directed Evolution Strategy

The following simulations use a (30,200)-ES with self-adaptive standard deviations. For each test function, an initial population is generated with the object variables $\boldsymbol{x}^{(0)}$ set by random in the intervals given in Table 11.1. The mutation strengths $\sigma^{(0)}$ are set to 0.1, the skewness and correlation parameters have initial values of 0. All experiments are carried out 50 times. In contrast to [Berlik and Reusch, 2004], reported here are the results until at least one optimization variant reaches fitness values better than 10^{-10}. The simulations are carried out using the EO-framework[1], supplemented with the missing classes and templates for directed mutation.

In total, seven mutation operators have been analyzed, whereof four realize directed mutation:

- *Naïve skew-normal mutation*
  This operator was discussed in Section 7.4, is defined in sections, and has convergent expectation and variance. Coupled recombination (cf. Section 9.4) and a damping factor $d = 0.05$ are used (cf. Section 9.2).

- *Skew-normal mutation*
  This most relevant operator was presented in Section 8.4 and is used with coupled recombination. Again, a damping factor $d = 0.05$ is used to mutate the shape parameters.

- *Standardized skew-normal mutation*
  The operator, being a standardized version of the previously mentioned, was presented in Section 8.5. Here it is used with coupled recombination and a damping factor $d = 0.05$.

[1] EO (abbr. for *Evolving Objects*) is a templates-based, ANSI-C++ compliant evolutionary computation library. http://eodev.sourceforge.net/

- *Asymmetric mutation*
  This was the first operator introducing true directionality into mutation, recapitulated in Section 7.2. Like the naïve skew-normal mutation its density is defined in sections but has divergent expectation and variance. The operator is used as proposed by Hildebrand [2001], i.e. with conventional recombination and shape parameters being mutated the same way the object variables are.

The other three operators are the conventional ones, as described in Section 4.7, *Mutation Operator*:

- *Isotropic mutation with one mutation strength*
  Only one mutation strength is used that is applied in turn to every object variable.

- *Scaled mutation with n mutation strengths*
  n mutation strengths are used, a separate one for each object variable.

- *Correlated mutation*
  Besides the control of the n mutation strengths, correlated mutation allows to rotate the coordinate pane arbitrarily by supporting a full covariance matrix.

The optimization algorithms have been applied to seven well-known test functions, see Table 11.1. The functions can be defined for arbitrary dimensions and have been treated here with 30 dimensions. All have the minimal value 0 which is for all but $f_{\text{Rosenbrock}}$ located at the origin and for Rosenbrock's function at $\mathbf{1}$. Except for f_{Ackley} they all are unimodal. The reason for this selection is the fact that the local behavior of the mutation operator should be studied. The number of local minima in f_{Ackley} increases exponentially with the function dimension. All but f_{step} are continuous functions.

For each function the results are depicted in a log scale plot in the following Figures 11.1–11.7. Since in a minimization a single outlier can completely disturb the mean, also the more robust median is reported.

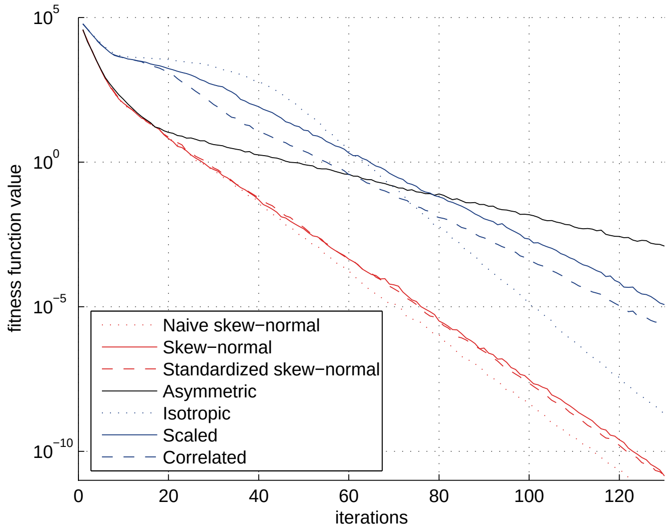
As the overall results are not competitive with the results of the Directed Covariance Matrix Adaptation-ES (which we will see later), we will not get involved in the details here. Hence, concrete figures of the results as well as significance analyses are omitted.

Table 11.1: Test functions

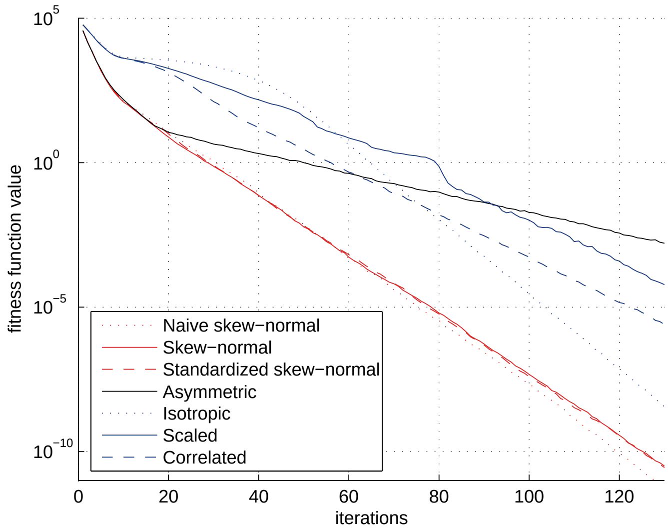
Name	Function	$\boldsymbol{x}^{(0)}$
Sphere	$f_{\text{sphere}}(\boldsymbol{x}) = \sum_{i=1}^{30} x_i^2$	$[-100, 100]^{30}$
Schwefel 2.22	$f_{\text{Schwefel}_{2.22}}(\boldsymbol{x}) = \sum_{i=1}^{30} \lvert x_i \rvert + \prod_{i=1}^{30} \lvert x_i \rvert$	$[-10, 10]^{30}$
Schwefel 1.2	$f_{\text{Schwefel}_{1.2}}(\boldsymbol{x}) = \sum_{i=1}^{30} \left(\sum_{j=1}^{i} x_j \right)^2$	$[-100, 100]^{30}$
Schwefel 2.21	$f_{\text{Schwefel}_{2.21}}(\boldsymbol{x}) = \max\{\lvert x_i \rvert, 1 \leq i \leq 30\}$	$[-100, 100]^{30}$
Rosenbrock	$f_{\text{Rosen}}(\boldsymbol{x}) = \sum_{i=2}^{29} \left(100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right)$	$[-30, 30]^{30}$
Step	$f_{\text{step}}(\boldsymbol{x}) = \sum_{i=1}^{30} \left(\lfloor x_i + 0.5 \rfloor \right)^2$	$[-100, 100]^{30}$
Ackley	$f_{\text{Ackley}}(\boldsymbol{x}) = -20 \exp\left(-0.2 \sqrt{\frac{1}{30} \sum_{i=1}^{30} x_i^2} \right) - \exp\left(\frac{1}{30} \sum_{i=1}^{30} \cos(2\pi x_i) \right) + 20 + e$	$[-32, 32]^{30}$

11.1.1 Simulation Results

Sphere Function



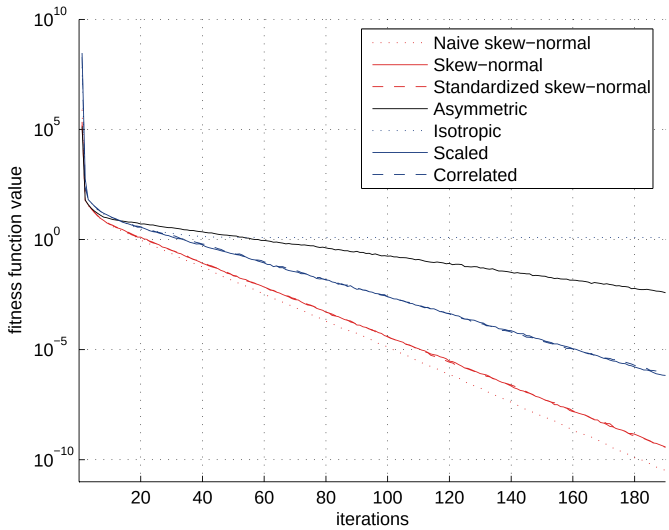
(a) Median



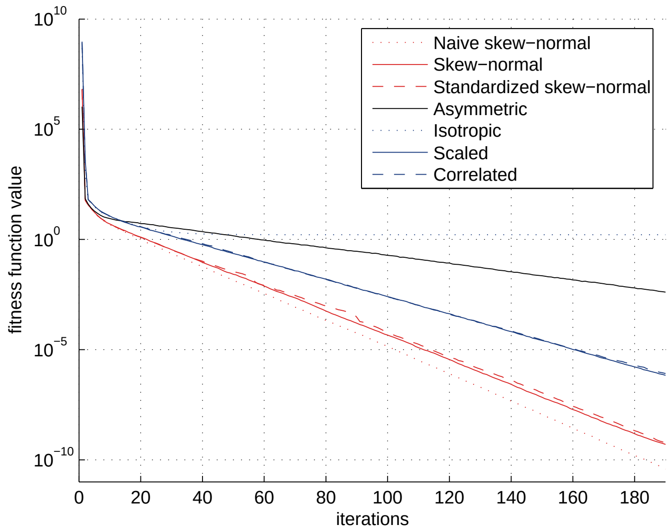
(b) Mean

Figure 11.1: Simulation results on the sphere function. Median respectively mean of fitness function values versus the number of iterations

Schwefel's Function 2.22



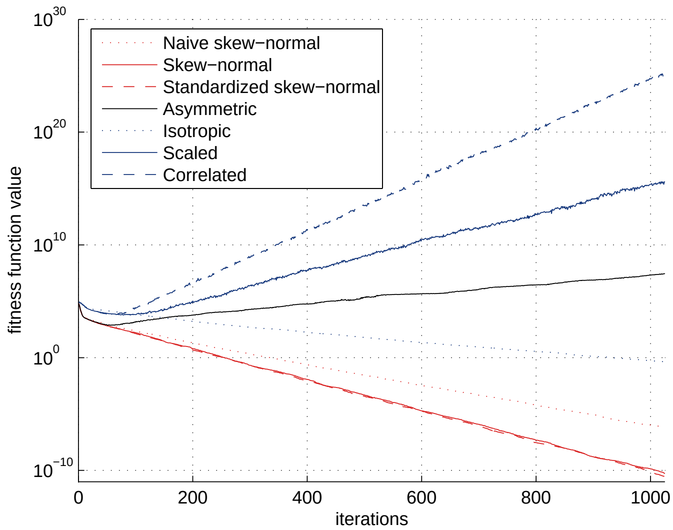
(a) Median



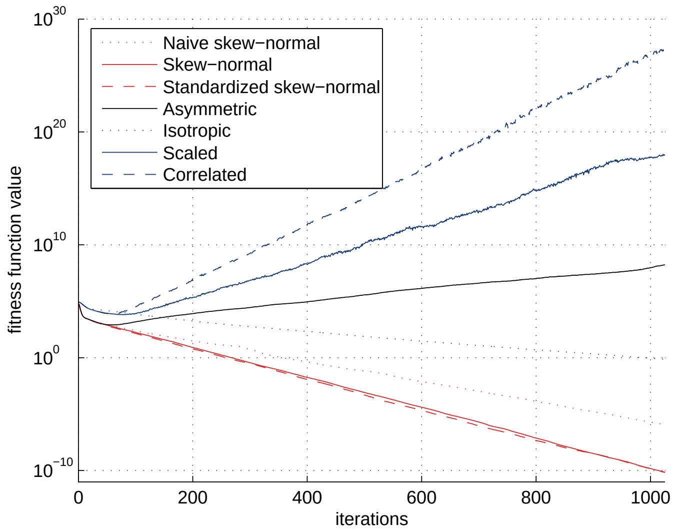
(b) Mean

Figure 11.2: Simulation results on Schwefel's function 2.22. Median respectively mean of fitness function values versus the number of iterations

Schwefel's Function 1.2



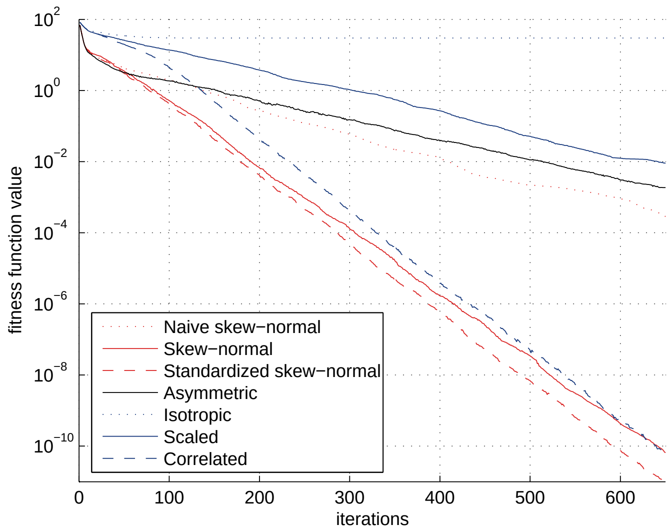
(a) Median



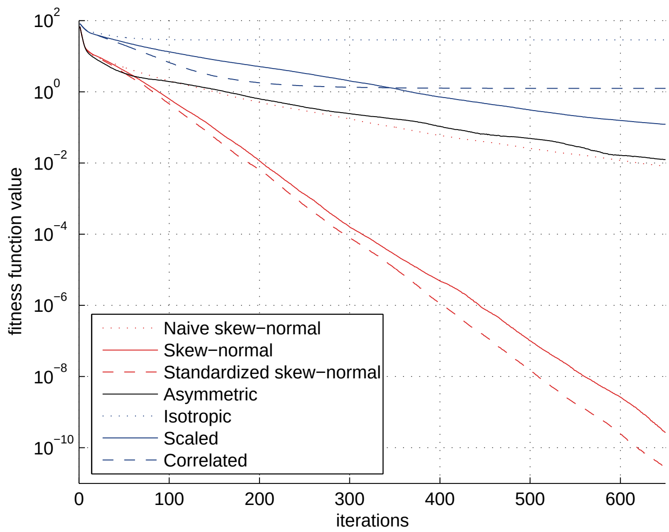
(b) Mean

Figure 11.3: Simulation results on Schwefel's function 1.2. Median respectively mean of fitness function values versus the number of iterations

Schwefel's Function 2.21



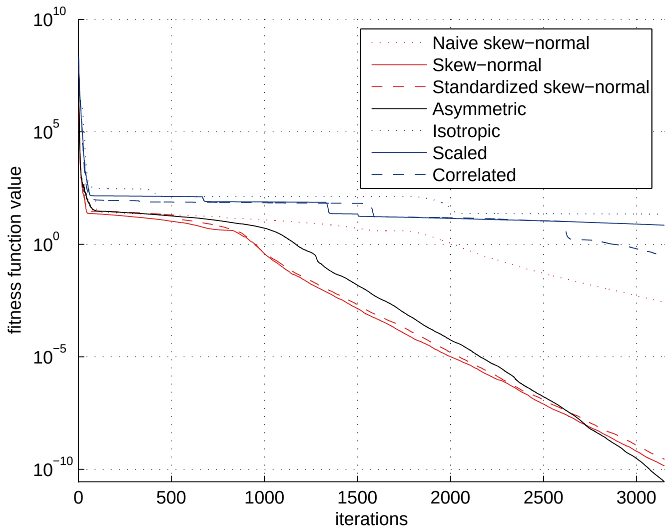
(a) Median



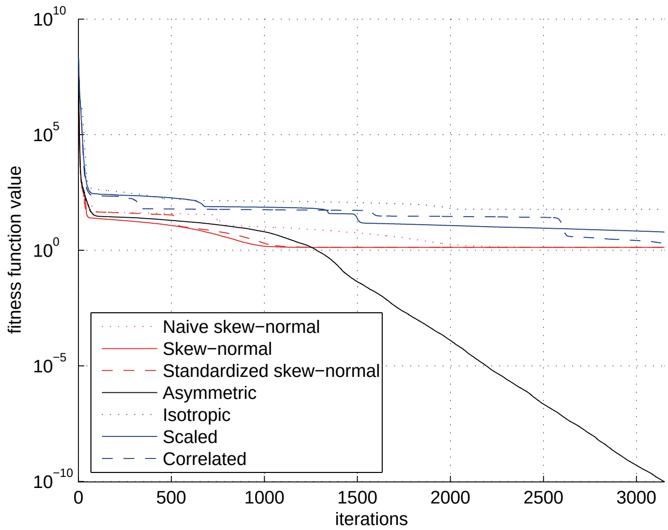
(b) Mean

Figure 11.4: Simulation results on Schwefel's function 2.21. Median respectively mean of fitness function values versus the number of iterations

Generalized Rosenbrock Function



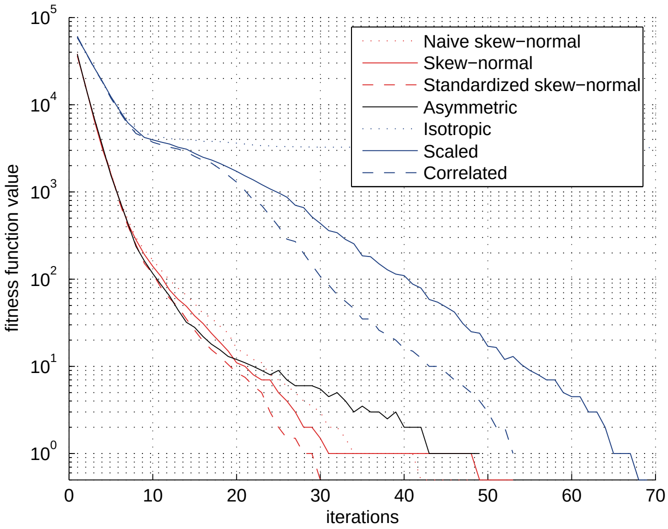
(a) Median



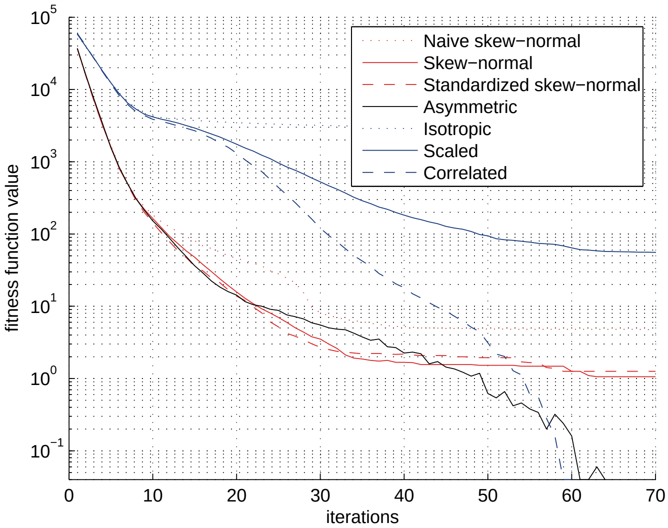
(b) Mean

Figure 11.5: Simulation results on Rosenbrock's generalized function. Median respectively mean of fitness function values versus the number of iterations

Step Function



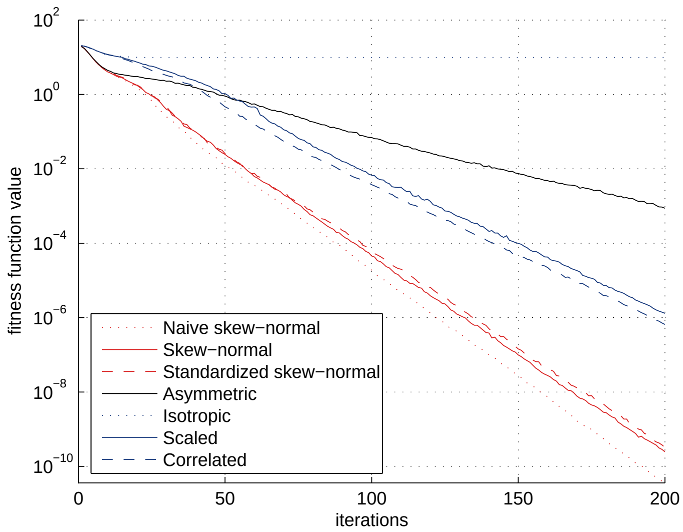
(a) Median



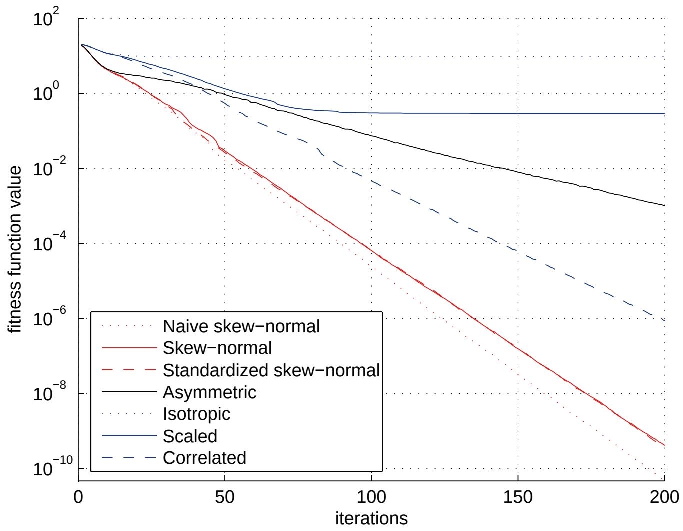
(b) Mean

Figure 11.6: Simulation results on the step function. Median respectively mean of fitness function values versus the number of iterations

Ackley's Function



(a) Median



(b) Mean

Figure 11.7: Simulation results on Ackley's function. Median respectively mean of fitness function values versus the number of iterations

11.1.2 Discussion of the Results

On the sphere function, the ES with NSN mutations performs slightly better than the almost equal SSN and SN. The asymmetric ES is clearly outperformed by the conventional ESs. Nearly the same happens on Schwefel's function 2.22. However, here the isotropic ES stagnates and performs worst; the asymmetric ES shows the slowest convergence of the other operators. More dramatically are the results on Schwefel's function 1.2. ESs with SN and SSN mutations perform almost equal and are best. Slightly worse is the ES with NSN mutation. The isotropic ES is the last converging algorithm. All other strategies diverge. On Schwefel's function 2.21 there are three efficiency groups. Best performing algorithms are ESs with SSN and SN mutations and the correlated ES. The group with slow convergence rate comprises the ES with NSN mutations, asymmetric and scaled ESs. The isotropic ES stagnates. On Rosenbrock's function ESs with SN and SSN mutation perform best together with the asymmetric ES. The other strategies converge relatively slow. Similar results occur for the step function. Here ESs with SSN, SN, and NSN mutation perform best together with the asymmetric ES. Last, on Ackley's function ESs with NSN, SN, and SSN mutations are the best performing algorithms. Correlated and scaled ESs are performing worse, better than the asymmetric ES and the stagnating isotropic ES.

According to the previous figures ESs with SN and SSN mutation outperform the other mutation operators. Very significantly this is the case for Schwefel's function 2.21 and Rosenbrock's function. We can also see that the naïve skew-normal mutation performs nearly as good, even slightly better for the sphere function, Schwefel's function 2.22, and Ackley's function. Overall, ESs with SN, SSN, and also NSN mutations form the group of best performing operators. The forth directed mutation operator, i.e. the asymmetric mutation clearly performs worse. Compared to the classical variants it yields no gain.

11.1.3 Conclusions

Directed mutation by means of the SN, SSN, and NSN distribution clearly outperform the other mutation operators, where the SN and SSN are a bit more efficient than the NSN. Between SN and SSN no significance difference can be ascertained.

However, the DES by means of the SN and SSN distribution is by far not competitive with the DCMA-ES. Three functions of the above are also investigated with the DCMA-ES: the sphere function, Rosenbrock's function, and Schwefel's function 1.2. Even if the results cannot be compared directly to them of the DCMA-ES reported in Section 11.2, since here 30-dimensional functions are treated and the initialization is different, compared to 40-dimensional simulations of the DCMA-ES the DES is inferior. The latter needs on the sphere function 124 iterations, i.e. 24.800 function evaluations where for the DCMA-ES 5.271 function evaluations are sufficient. The ratio on Rosenbrock's function is 614.400 versus 77.808 function evaluations and on Schwefel's function 1.2 203.000 versus 20.020 evaluations. Thus the DES needs up to ten times more function evaluations than the DCMA-ES on lower dimensional problems.

11.2 Results of the Directed Covariance Matrix Adaptation-ES

Two different CMA-ESs are experimentally investigated: the original variant as described by Hansen and Kern [2004], using $\mathrm{N}(0, \boldsymbol{C})$ distributed random vectors and the DCMA-ES, using instead $\mathcal{SN}_{\mathrm{n}}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ distributed random vectors with intragenerational shape update; to be precise: a normed update with signum function weighting and dimensional scaling, see subsection 10.2.1, *Update Vector*. As intergenerational shape update strategy the shape vector of a generation has been initialized with the vector of the previous generation normed to a length of one.

For the comparison of the two functions, a test suite consisting of the eight well known functions shown in Table 11.2 is used. Initial values are set to $\boldsymbol{x}^{(0)} \in [-1, 1]^n$, $\sigma^{(0)} = 1$, and $\alpha^{(0)} = \mathbf{0}$ for all functions except for Rosenbrock's case where $\boldsymbol{x}^{(0)} = \mathbf{0}$, $\sigma^{(0)} = 0.1$, and again $\alpha^{(0)} = \mathbf{0}$. As stopping criterion for all functions but f_{parabR} and f_{sharpR} fitness better than 10^{-10} is demanded, for the two others fitness has to be less than -10^{10}. Tests are carried out in $n = [2, 5, 10, 20, 40, 80]$ dimensions and for offspring numbers $\lambda = 4 + \lfloor 3 \log n \rfloor$ with parent numbers $\mu = \lceil \lambda/2 \rceil$. For each combination 25 runs are done. The simulations have been carried out using the technical computing system MATLAB[2].

For each function the results are depicted in log-log scale plot and the corresponding figures are given in a table. Reported are statistics of the number of necessary function evaluations: the mean $\bar{x}$, the standard deviation σ, and the median m. In the comparison $\Delta = \bar{x}_{\mathrm{CMA}} - \bar{x}_{\mathrm{DCMA}}$ gives the difference of the means, where positive values indicate better performance of the DCMA. $\eta = \bar{x}_{\mathrm{CMA}} / \bar{x}_{\mathrm{DCMA}}$ represents the ratio of the means, i.e. the factor the DCMA performs better. Further, for each constellation the significance of the results is analyzed. Therefore nonparametric Wilcoxon rank sum tests for equal medians are performed [Wilcoxon, 1945].

Tested is the hypothesis that two independent samples, here given in form of vectors of numbers of function evaluations for the CMA and DCMA algorithms, come from distributions with equal medians. The corresponding p-value from the test is reported where p is the probability of observing the given result. The hypothesis test is performed at the 0.05 significance level. h indicates if the null hypothesis, i.e. medians are equal, can be rejected at the 5% level. If $h = 0$, then the null hypothesis cannot be rejected at the 5% level. If $h = 1$, then the null hypothesis can be rejected at the 5% level. The Wilcoxon rank sum test is equivalent to the Mann-Whitney U test [Mann and Whitney, 1947]. For a detailed discussion of nonparametric statistical methods confer the book of Hollander and Wolfe [1999].

[2]MATLAB is a registered trademark of The MathWorks, Inc.

Table 11.2: Test functions

Name	Function	$\boldsymbol{x}^{(0)}$	$\sigma^{(0)}$	$\boldsymbol{\alpha}^{(0)}$	f^{stop}
Sphere	$f_{\text{sphere}}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$	$[-1,1]^n$	1	$\mathbf{0}$	10^{-10}
Schwefel	$f_{\text{Schwefel}}(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j\right)^2$	$[-1,1]^n$	1	$\mathbf{0}$	10^{-10}
Cigar	$f_{\text{cigar}}(\boldsymbol{x}) = x_1^2 + \sum_{i=2}^{n}(1000 x_i)^2$	$[-1,1]^n$	1	$\mathbf{0}$	10^{-10}
Tablet	$f_{\text{tablet}}(\boldsymbol{x}) = (1000 x_1)^2 + \sum_{i=2}^{n} x_i^2$	$[-1,1]^n$	1	$\mathbf{0}$	10^{-10}
Ellipsoid	$f_{\text{elli}}(\boldsymbol{x}) = \sum_{i=1}^{n} \left(1000^{\frac{i-1}{n-1}} x_i\right)^2$	$[-1,1]^n$	1	$\mathbf{0}$	10^{-10}
Parabolic ridge	$f_{\text{parabR}}(\boldsymbol{x}) = -x_1 + 100 \sum_{i=2}^{n} x_i^2$	$[-1,1]^n$	1	$\mathbf{0}$	-10^{10}
Sharp ridge	$f_{\text{sharpR}}(\boldsymbol{x}) = -x_1 + 100 \sqrt{\sum_{i=2}^{n} x_i^2}$	$[-1,1]^n$	1	$\mathbf{0}$	-10^{10}
Rosenbrock	$f_{\text{Rosen}}(\boldsymbol{x}) = \sum_{i=2}^{n-1} \Big(100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2\Big)$	$\mathbf{0}$	0.1	$\mathbf{0}$	10^{-10}

11.2.1 Simulation Results

Sphere Function

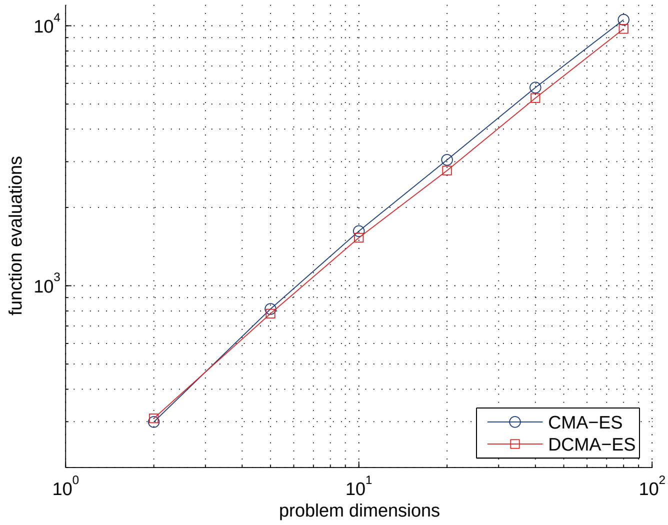


Figure 11.8: Number of function evaluations versus the problem dimensions of simulations on the sphere function

Table 11.3: Simulations on the sphere function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	309.12	27.06	312	299.28	32.88	294
5	780.80	49.96	776	814.40	53.91	800
10	1530.00	69.04	1510	1621.60	92.81	1610
20	2775.36	89.23	2748	3047.04	93.33	3060
40	5271.00	135.21	5265	5785.20	136.22	5805
80	9719.24	171.45	9724	10536.60	147.06	10540

Table 11.4: Analysis of the sphere function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	−9.84	0.968	0	$2.58 \cdot 10^{-1}$
5	33.60	1.043	1	$4.51 \cdot 10^{-2}$
10	91.60	1.060	1	$6.91 \cdot 10^{-4}$
20	271.68	1.098	1	$5.69 \cdot 10^{-9}$
40	514.20	1.098	1	$1.99 \cdot 10^{-9}$
80	817.36	1.084	1	$1.39 \cdot 10^{-9}$

Schwefel Function 1.2

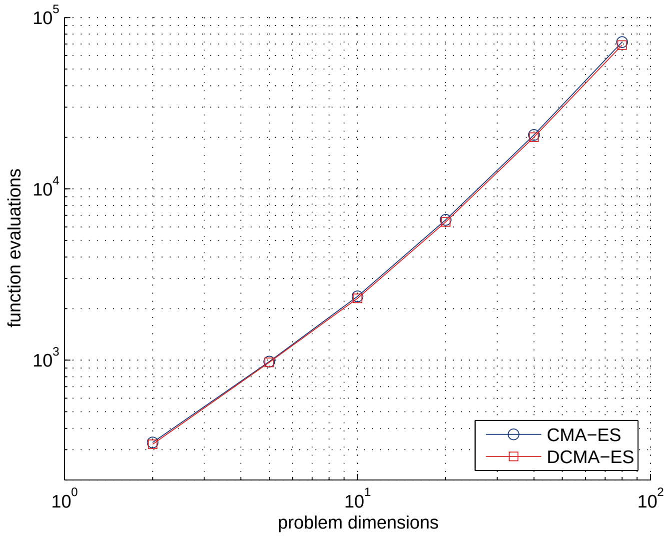


Figure 11.9: Number of function evaluations versus the problem dimensions of simulations on the Schwefel function 1.2

Table 11.5: Simulations on the Schwefel function 1.2

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	323.04	39.14	318	330.72	37.08	330
5	970.24	69.80	968	979.84	89.92	984
10	2297.20	104.10	2290	2358.40	92.77	2340
20	6409.92	210.70	6384	6596.64	255.65	6624
40	20020.80	355.77	20025	20630.40	404.77	20580
80	69013.20	1045.78	68816	71963.72	970.96	71944

Table 11.6: Analysis of the Schwefel function 1.2 results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	7.68	1.024	0	$4.14 \cdot 10^{-1}$
5	9.60	1.010	0	$6.13 \cdot 10^{-1}$
10	61.20	1.027	1	$4.65 \cdot 10^{-2}$
20	186.72	1.029	1	$9.82 \cdot 10^{-3}$
40	609.60	1.030	1	$5.11 \cdot 10^{-6}$
80	2950.52	1.043	1	$4.36 \cdot 10^{-9}$

Cigar Function

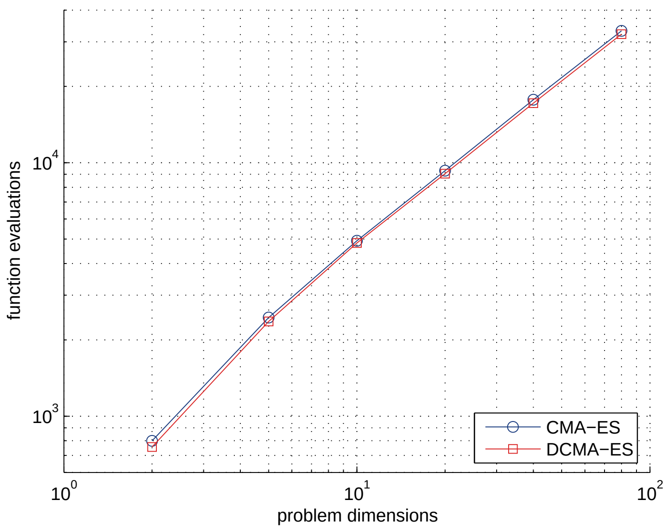


Figure 11.10: Number of function evaluations versus the problem dimensions of simulations on the cigar function

Table 11.7: Simulations on the cigar function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	756.00	72.37	738	799.20	67.42	798
5	2363.20	129.02	2368	2450.56	114.37	2416
10	4825.60	172.82	4800	4923.60	148.21	4930
20	9048.00	215.19	9036	9296.64	219.10	9300
40	17177.40	289.17	17100	17727.00	263.53	17790
80	32159.24	295.65	32079	33139.80	332.81	33184

Table 11.8: Analysis of the cigar function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	43.20	1.057	1	$2.81 \cdot 10^{-2}$
5	87.36	1.037	1	$2.61 \cdot 10^{-2}$
10	98.00	1.020	1	$3.27 \cdot 10^{-2}$
20	248.64	1.027	1	$2.94 \cdot 10^{-4}$
40	549.60	1.032	1	$7.05 \cdot 10^{-7}$
80	980.56	1.030	1	$3.86 \cdot 10^{-9}$

Tablet Function

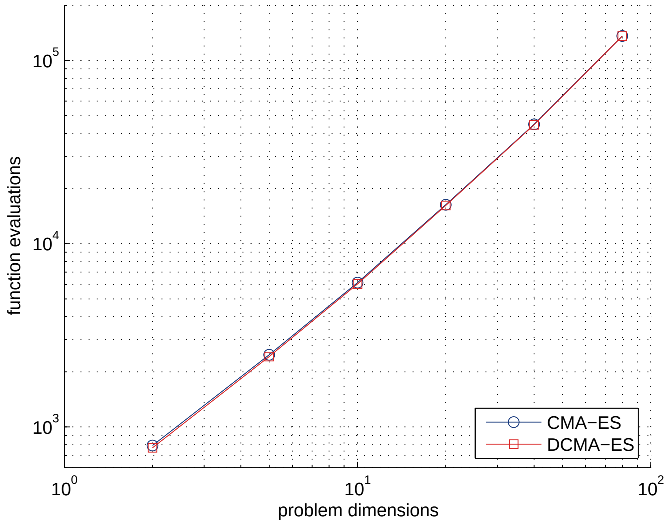


Figure 11.11: Number of function evaluations versus the problem dimensions of simulations on the tablet function

Table 11.9: Simulations on the tablet function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	770.16	56.79	768	792.24	85.99	768
5	2420.80	148.04	2408	2476.48	163.61	2448
10	6034.80	139.71	6040	6138.00	210.14	6160
20	16153.44	243.24	16212	16309.44	248.10	16248
40	44551.20	605.27	44580	44784.60	683.98	44610
80	135655.92	1324.06	135864	135951.04	1158.24	136102

Table 11.10: Analysis of the tablet function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	22.08	1.029	0	$3.82 \cdot 10^{-1}$
5	55.68	1.023	0	$2.48 \cdot 10^{-1}$
10	103.20	1.017	0	$7.74 \cdot 10^{-2}$
20	156.00	1.010	0	$6.23 \cdot 10^{-2}$
40	233.40	1.005	0	$3.47 \cdot 10^{-1}$
80	295.12	1.002	0	$4.49 \cdot 10^{-1}$

Ellipsoid Function

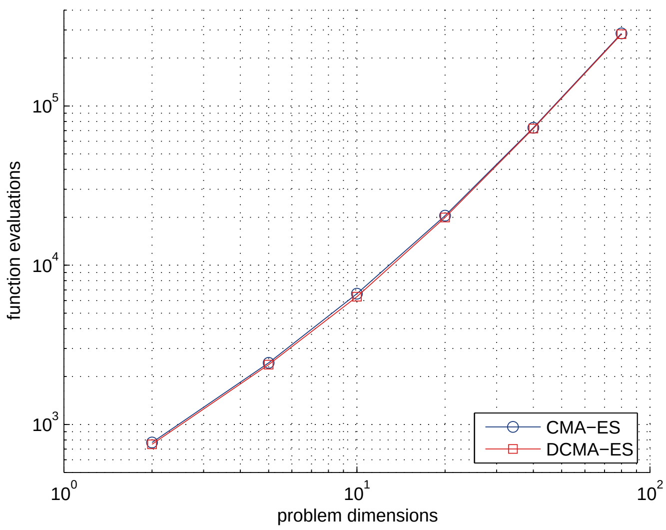


Figure 11.12: Number of function evaluations versus the problem dimensions of simulations on the ellipsoid function

Table 11.11: Simulations on the ellipsoid function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	750.24	58.54	756	771.12	56.95	786
5	2377.28	172.21	2392	2440.32	153.23	2440
10	6350.40	209.53	6330	6622.00	183.21	6670
20	19961.76	268.06	19920	20460.00	345.27	20508
40	72105.60	632.09	72195	72965.40	636.00	72825
80	282407.40	3617.43	281010	285470.12	3316.67	286365

Table 11.12: Analysis of the ellipsoid function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	20.88	1.028	0	$2.44 \cdot 10^{-1}$
5	63.04	1.027	0	$1.90 \cdot 10^{-1}$
10	271.60	1.043	1	$4.39 \cdot 10^{-5}$
20	498.24	1.025	1	$7.72 \cdot 10^{-6}$
40	859.80	1.012	1	$4.06 \cdot 10^{-5}$
80	3062.72	1.011	1	$2.24 \cdot 10^{-3}$

Parabolic Ridge Function

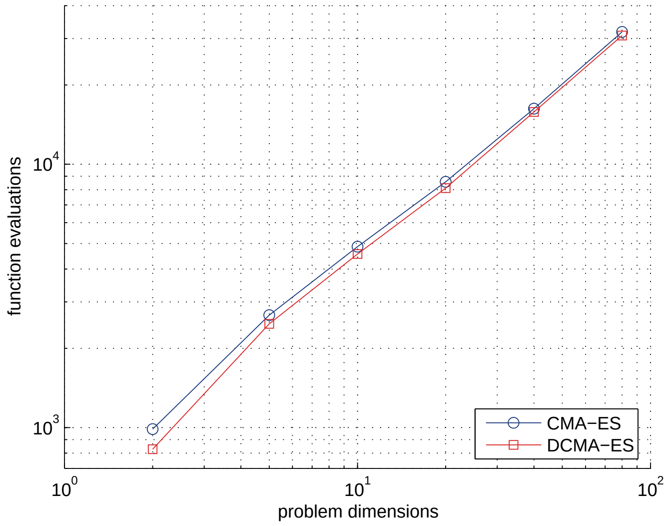


Figure 11.13: Number of function evaluations versus the problem dimensions of simulations on the parabolic ridge function

Table 11.13: Simulations on the parabolic ridge function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	827.52	58.48	828	986.64	104.64	984
5	2475.52	112.00	2472	2670.40	172.57	2672
10	4559.60	173.41	4540	4860.80	257.41	4910
20	8108.64	201.02	8052	8568.00	229.83	8544
40	15779.40	193.09	15765	16252.20	210.36	16305
80	30785.64	198.79	30770	31753.28	349.59	31722

Table 11.14: Analysis of the parabolic ridge function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	159.12	1.192	1	$2.42 \cdot 10^{-7}$
5	194.88	1.079	1	$6.64 \cdot 10^{-5}$
10	301.20	1.066	1	$8.45 \cdot 10^{-5}$
20	459.36	1.057	1	$8.47 \cdot 10^{-8}$
40	472.80	1.030	1	$7.60 \cdot 10^{-8}$
80	967.64	1.031	1	$1.40 \cdot 10^{-9}$

Sharp Ridge Function

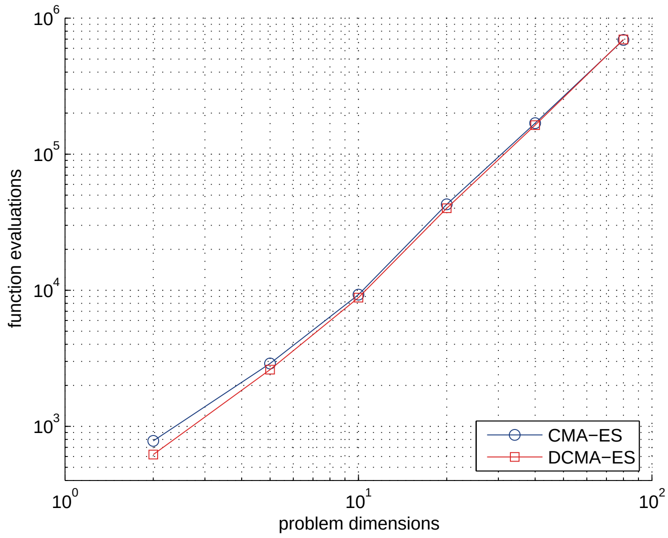


Figure 11.14: Number of function evaluations versus the problem dimensions of simulations on the sharp ridge function

Table 11.15: Simulations on the sharp ridge function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	619.20	94.82	606	781.20	68.98	780
5	2605.76	343.66	2504	2894.08	311.52	2880
10	8828.00	1350.98	8630	9272.80	1606.54	8990
20	40051.20	3364.80	40680	42744.96	3101.96	43044
40	163498.80	10048.32	160665	168999.60	7794.64	171120
80	695582.88	29410.41	700043	690688.92	22140.11	687854

Table 11.16: Analysis of the sharp ridge function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	162.00	1.262	1	$5.22 \cdot 10^{-7}$
5	288.32	1.111	1	$3.95 \cdot 10^{-3}$
10	444.80	1.050	0	$3.62 \cdot 10^{-1}$
20	2693.76	1.067	1	$8.81 \cdot 10^{-3}$
40	5500.80	1.034	1	$3.13 \cdot 10^{-2}$
80	−4893.96	0.993	0	$5.09 \cdot 10^{-1}$

Generalized Rosenbrock Function

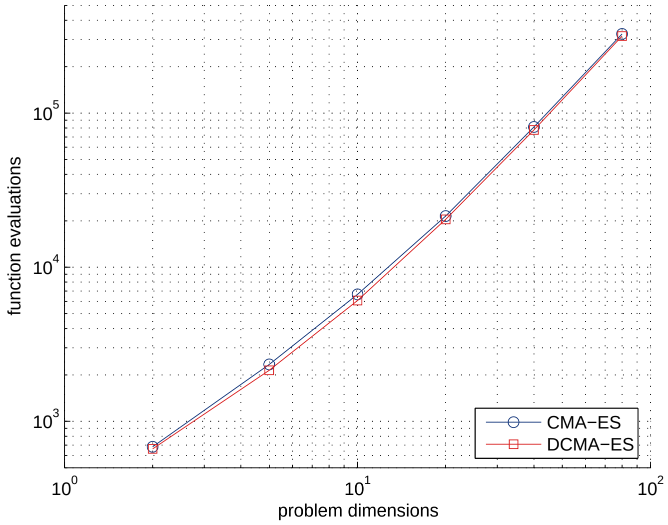


Figure 11.15: Number of function evaluations versus the problem dimensions of simulations on the generalized Rosenbrock function

Table 11.17: Simulations on the generalized Rosenbrock function

Dim	DCMA			CMA		
	$\bar{x}$	σ	m	$\bar{x}$	σ	m
2	661.92	51.87	666	683.76	74.53	672
5	2144.00	159.97	2168	2337.60	152.46	2360
10	6075.20	277.54	6060	6662.80	220.46	6690
20	20441.28	529.19	20448	21471.36	697.76	21300
40	77808.00	1003.06	77955	81338.40	1319.35	80985
80	316132.00	3740.01	315656	325992.00	3236.83	325754

Table 11.18: Analysis of the generalized Rosenbrock function results

Dim	Comparison DCMA / CMA			
	Δ	η	h	p
2	21.84	1.033	0	$4.26 \cdot 10^{-1}$
5	193.60	1.090	1	$2.26 \cdot 10^{-4}$
10	587.60	1.097	1	$6.81 \cdot 10^{-8}$
20	1030.08	1.050	1	$2.92 \cdot 10^{-6}$
40	3530.40	1.045	1	$2.42 \cdot 10^{-9}$
80	9860.00	1.031	1	$2.56 \cdot 10^{-8}$

11.2.2 Discussion of the Results

First, note that due to the log-log scale plot the differences between the algorithms appear quite small. The runs on f_{sphere} show a DCMA-ES that outperforms the CMA-ES from 4% up to nearly 10% for dimensions greater than 2. For $n = 2$ the CMA yields better but not significant better results. On f_{Rosen} approximately the same happened. For $n = 2$ the difference is also not significant, however the DCMA here is always performing better than the CMA. The gain decreases form almost 10% for low dimensionality to about 3% for 80 dimensions. On f_{cigar} and f_{parabR} the DCMA is always significantly better, on the former relative constantly about 3% and on the latter decreasing with increasing dimensions from 19(!)% to 3%. Although on f_{elli} and f_{Schwefel} the DCMA is again always better, there is no significance for $n = 2$ respectively $n = 5$ and the overall gain is less than for f_{cigar} and f_{parabR}. All results superior but none of them significant is the outcome on f_{tablet}. On the function f_{sharpR} the outcome is somewhat irregular. The results are not significant for $n = 10$ respectively $n = 80$. In the other cases the gain of the DCMA decreases with increasing dimensions from 26(!)% to 3%.

11.2.3 Conclusions

In general, the DCMA-ES performs better on all functions and on average, there is a gain of a few percentage points. This has to be seen against the background of the CMA-ES considered already as state-of-the-art in parameter optimization, the preliminary design of shape vector control, and the very small overhead caused by directed mutation. In fact, all that has to be done to create multivariate skew-normal instead of normal random vectors is to calculate one n-dimensional scalar product, generate one univariate random number, and do one comparison. Compared to a function evaluation in a real world application this can rather be neglected. Analog to uncorrelated EAs, where with directed mutation a promising new mutation principle had been presented, this new mutation principle has now been introduced in the CMA-ES context, too.

12 Multiobjective Optimization of Screw-Type Machines

This chapter presents simulation results of a directed Evolution Strategy applied to a real world scenario and is for the most part a consolidation of two papers [Berlik and Fathi, 2005b, Berlik et al., 2006]. Stating point is the well known NSGA-II optimization algorithm[1] [Deb et al., 2000], slightly modified and enhanced by the concept of directed mutation.

Directed mutation has already been shown to improve the efficiency of evolutionary algorithms significantly for a broad spectrum of test problems (see the last chapter or e.g. [Berlik, 2004b, Berlik and Reusch, 2004]). While the capability of directed mutation has thus been verified in the sandbox, in the case of real world applications this has not been done so far. The aim of this chapter is to make up for it, considering a high dimensional, multicriterial problem with several constraints. Optimizing a screw-type machine, we will be concerned with a problem from mechanical engineering.

The present results originate from a DFG[2] founded cooperation project of Prof. Dr. Reusch's Chair[3] and Prof. Dr.-Ing. Kauder's Chair[4]. Besides, a tool for the interactive design of screw-type machines arose from this cooperation that also will be presented. Apart from the usual features of computer aided design tools it has unique support for screw-type machine design, as for example automated calculation of the female rotor for an arbitrary given male rotor.

Therefore, first the screw-type machine will be explained in short. Then the optimization problem is outlined and the used algorithm is sketched. Simulation results for different mutation operators are presented and discussed. In the final section, conclusions are provided.

12.1 Screw-Type Machines

The most common form of screw-type machines are rotary compressors, especially the helical twin screw-type. Meshing male and female screw-rotors rotate inside a housing in opposite directions and thereby trap air, reducing the volume of the air along the rotors to the air discharge point. Rotary screw-type compressors have low initial cost, compact size, low weight, and are easy to maintain. Figure 12.1 shows such a screw-type compressor with its rotors torn out of the housing for demonstration purposes.

[1] abbr. for *Non-Dominated Sorting Genetic Algorithm-II*

[2] abbr. for *Deutsche Forschungsgemeinschaft*, German Research Foundation

[3] Automata- and Switching Theory and Computational Intelligence, Department of Computer Science, Universität Dortmund

[4] Fluid Energy Machines, Faculty of Mechanical Engineering, Universität Dortmund

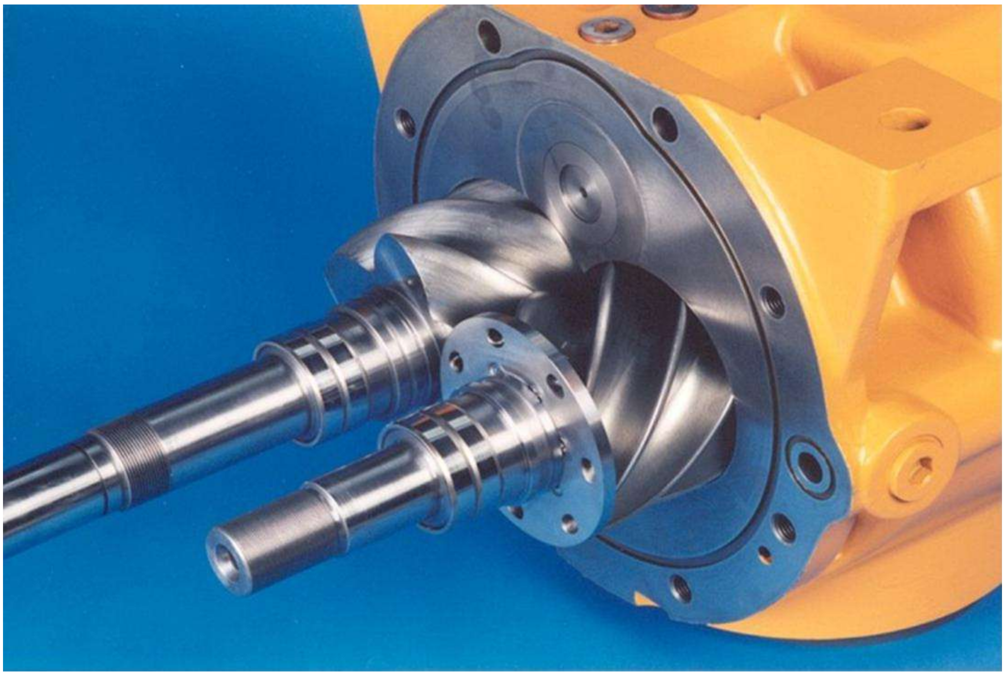


Figure 12.1: Screw-type compressor

A special topic during the construction of screw machines is to be seen in the design of the rotor geometry of an individual stage. One can differentiate here three-dimensional characteristics such as rotor length and wrap angle as well as two-dimensional characteristics such as rotor diameters, numbers of lobes and the lobe profile. At the beginning, there stands the draft of a suitable two-dimensional front section, since this already significantly exerts influence on the operational behavior, the thermodynamic and mechanical characteristics and the manufacturing. The meaning of the front sections results among other things from the operating behavior of the screw machine. A change of the lobe profile or the number of lobes affects on the one hand the contact line and by this the form and position of working chamber limiting clearances, on the other hand the size of the work space itself and the utilization of the construction volume. Both affect the quality of the thermodynamic processing.

The development of front sections was based in the past mainly on the combination of few geometrically simple curve sections. Frequently used curve types are e.g. straight lines, circular arcs, involutes, cycloids or equidistant ones to cycloids. Here, Bézier splines were chosen because of their great flexibility and thus having the convenience to operate with one single curve type only.

12.2 Computer Aided Design and Optimization Tool

In cooperation with a medium-sized German mechanical engineering company, a CAD tool for screw-type machine design, *ScrewView*, has been developed. A screen shot of it is given in Figure 12.2. Beside the usual traits of computer aided design tools its main features are:

- calculation of the female rotor for an arbitrary male rotor,
- calculation of the clearances,
- detection of flanks and edges,
- detection and automatic removal of loops,
- analysis of the normals,
- scaling of rotors (scale, curve normal, area normal),
- calculation of several key figures,
- ongoing check of the profiles' integrity.

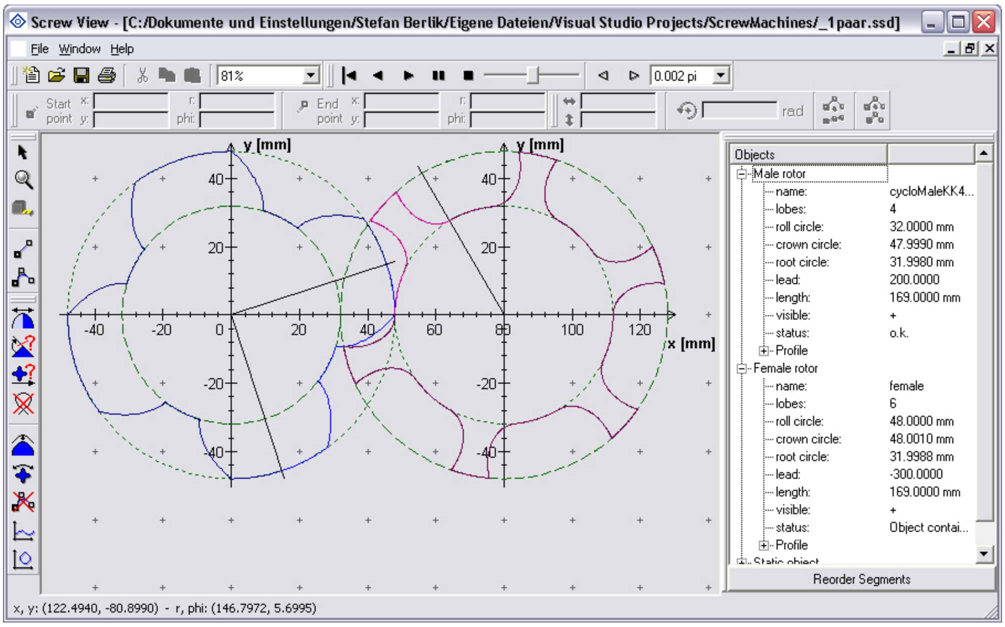


Figure 12.2: Screenshot of the *ScrewView* program

Further, an interface to optimization modules is provided. Thus, the implemented data structures and calculation routines can be used outside *ScrewView*, like it is here done for example with optimization algorithms build in the EO-framework, see subsection 12.3.3, *Algorithm*.

Another feature is its platform independence. Since *ScrewView* is written completely in C++ using STL[5] data structures and the graphical user interface

[5] The STL (abbr. for *Standard Template Library*) is a generic collection of class templates, containers, iterators, and algorithms included in the C++ standard library. http://www.stlport.org/

Qt[6], it can be used under all operating systems that are supported by the mentioned software, especially Microsoft Windows[7] and UNIX[8] derivatives.

12.3 Optimization

Designing any type of machine leads inevitably to the question of the optimized construction parameters. Mostly this can be estimated by the degree of performance with regard to the user's demands. Thereby it appears the difficulty to define a capable performance degree and to evaluate a possible solution with a maintainable expense. Usually there are many different criteria to be considered. Thus an appropriate optimization method has to be able to either represent different criteria within one utility function or to treat several criteria in parallel. This problem will be investigated below using a dry running twin-screw compressor as an example. The emphasis of the following subsections is to give an overview about the optimization, i.e. the necessary manipulation routines and the realized mutation operator. Note that the stated conditions for validity of the segments will directly meet some constraints in the multiobjective optimization problem.

12.3.1 Manipulation Routines

To be able to vary the front cut with its Bézier segments suitable manipulation routines have to be provided [Berlik, 2001]. A possibility to do this is given with one operator that shifts the initial or end base of a segment and another that modifies the gradients at these points. For points given in polar coordinates thus three operators that change

- the angle of the point,
- the radius of the point,
- the gradient at the point

are required. The modification of the segments then simply can be carried out by varying only the initial points. This applies, because the end point of every segment coincides per definition with the initial point of the following segment. Thus end points are modified in common with their corresponding initial points.

Change of the angle or radius

Bases can be shifted by changing their angle and / or radius. The effects on the two concerned Bézier segments are shown in Figure 12.3. Valid are only variations that change neither the gradient at the end point of the first segment nor the gradient at the initial point of the succeeding segment. The potential

[6] Qt is a cross-platform graphical widget toolkit for the development of graphical user interfaces produced by the Norwegian company Trolltech. Trolltech and Qt are trademarks of Trolltech AS.

[7] Microsoft and Windows are registered trademarks of Microsoft Corporation.

[8] UNIX is a registered trademark of The Open Group.

area of varied bases thus is limited by the tangents of the first initial and second end point and parallels to the tangent of the connecting point.

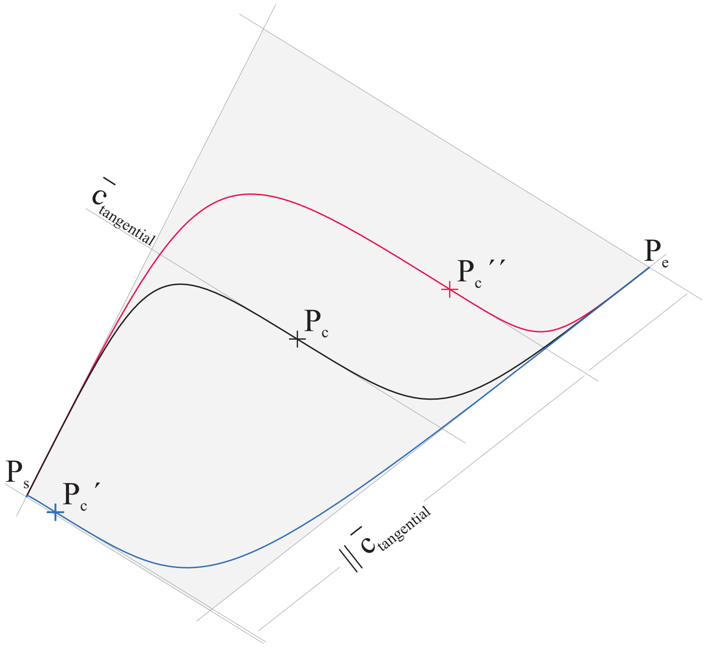


Figure 12.3: Shift of the conjoint point of two Bézier segments

$\mathrm{P_s}$:	initial point of the first segment
$\mathrm{P_c}$:	conjoint point of the segments
$\mathrm{P_e}$:	end point of the second segment
$\mathrm{P'_c}, \mathrm{P''_c}$:	shifted conjoint points
black line:	initial segments
red and blue line:	segments after shifting the conjoint points
gray:	valid area for conjoint points

Change of the gradient

The second operation for Bézier segment manipulation is changing the gradient at the initial respective end point. An example is depicted in Figure 12.4. Again, the operation is valid only within a certain range. One limit is reached if a Bézier segment degenerates into a line, the other if the tangents of the initial and end point of a segment become parallel.

12.3.2 Simulations

As stated before, the geometrical modeling of the front cut of the rotors is done using splines. In this example eight splines are used to describe a single lobe of the male profile, leading to a 32-dimensional optimization problem. The whole male rotor consists of four of these lobes; the female rotor with six lobes is calculated to fit to the male rotor. Several constraints have to be fulfilled by every generated profile pair to be valid. In this case ten constraints are

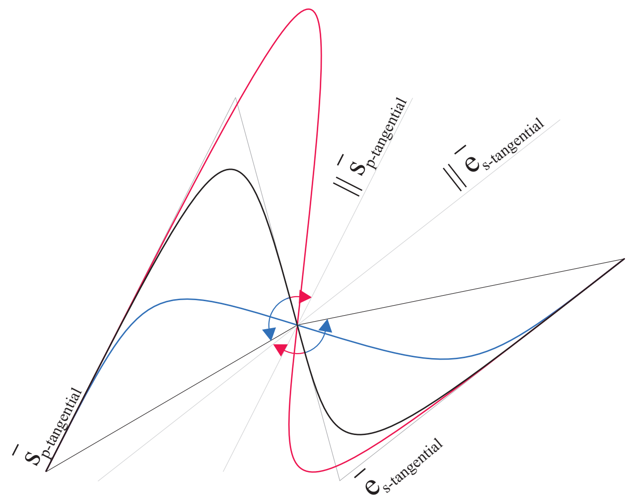


Figure 12.4: Varied gradients of the conjoint point of two Bézier segments

considered. While in a real-world optimization also several objectives are to be treated, for the sake of clarity in this example their number is limited. Just two very important objectives are used; the scoop area and the related polar moment of inertia (RPMI). The first is a measure of the volume flow through the machine and the second can be seen as a simplified measure of the stiffness of the rotors. These are conflicting goals that obviously both have to be maximized. For technical details on screw-type machines and general optimization of them see [Kauder, Reusch, Helpertz, and Berlik, 2001, 2002a, 2003]. Further details on multiobjective optimization of screw-type machines are provided in [Kauder, Reusch, Helpertz, and Berlik, 2002b].

For a more technical oriented treatise of the topic the reader is referred to [Helpertz, 2003]. There the calculation of fitting rotors, the objective functions and constraints are explained in great detail. Also, more realistic optimizations from the mechanical engineer's point of view are reported. However, as the primary interest was not the optimization algorithm itself, only conventional ESs have been applied.

12.3.3 Algorithm

The experiments here have been done using a NSGA-II like ES with self-adaptive standard deviations and shape parameters where necessary. Especially the constraint-domination principle and crowding distance calculation are taken over. An extensive presentation of these and multiobjective optimization in general is given by Deb [2001], van Veldhuizen and Lamont [1998], or Coello [1999]. In [van Veldhuizen and Lamont, 2000] the state-of-the-art in Multiobjective Evolutionary Algorithms is analyzed.

The implementation has been realized on basis of the EO-framework[9], supplemented with the missing classes and templates for constrained multiobjec-

[9] EO (abbr. for *Evolving Objects*) is a templates-based, ANSI-C++ compliant evolutionary computation library. http://eodev.sourceforge.net/

tive optimization. This optimization problem class also had an implication on the type of algorithm to be applied. Since the aim is to approximate the Pareto-optimal front, a spaciously Pareto-optimal set has to be maintained (cf. subsection 2.1.3). The CMA-ES is principally not applicable in this domain, as it consolidates the set of chosen solutions to the distribution's new expectation $\mathbf{m}^{(g+1)}$ in every iteration (4.36). Therefore, a conventional (30,100)-ES has been chosen for the optimizations.

To limit the runtime to acceptable values, the execution of the ES has been restricted to 5000 generations. Always the same initial profile has been used, mutation strengths have been set to 10^{-5}, the initial values of the skewness have been zero. All experiments have been carried out six times, where three mutation operators have been investigated:

- *isotropic mutation* with only one mutation strength for all dimensions together,
- *scaled mutation* with a separate mutation strength for every dimension,
- *directed mutation* by means of the skew-normal distribution.

12.3.4 Discussion of the Results

The results of the runs with isotropic mutation are given in Figure 12.5 and Table 12.1. Note that there are two runs with collapsing populations. At the end of the other runs relatively sharp Pareto fronts emerged. For all of them holds that they cover only a very small region of the search space. Figure 12.6 and Table 12.2 show the results using scaled mutation. Here the two criteria are not distributed equally. The standard deviations of the populations with respect to the second criterion are about 10 times larger. At last, Figure 12.7 and Table 12.3 show the results for the directed mutation. One can see that the populations form relatively sharp Pareto fronts and cover a good spectrum of the search space with respect to both criteria, see the average of the standard deviations of the single runs for the both criteria. Also, the populations form Pareto fronts sharper than it was the case with scaled mutation – which means that in the latter case there is a distinct fraction of individuals in the population that are not efficient.

To compare the results of the different mutation operators they have been compiled in Figure 12.8. It is apparent that directed mutation clearly outperforms the other two mutation strategies. All runs but one dominate all runs of the other strategies, i.e. are better in both criteria. The directed mutation also shows the greatest diversity under the different runs and within them.

12.4 Conclusions

With the directed mutation an operator is given that clearly outperforms the other mutation strategies for a high dimensional, constraint multiobjective real-world optimization problem. All results but one dominate all runs of the other strategies, i.e. are better in both criteria. Further, the resulting solutions form

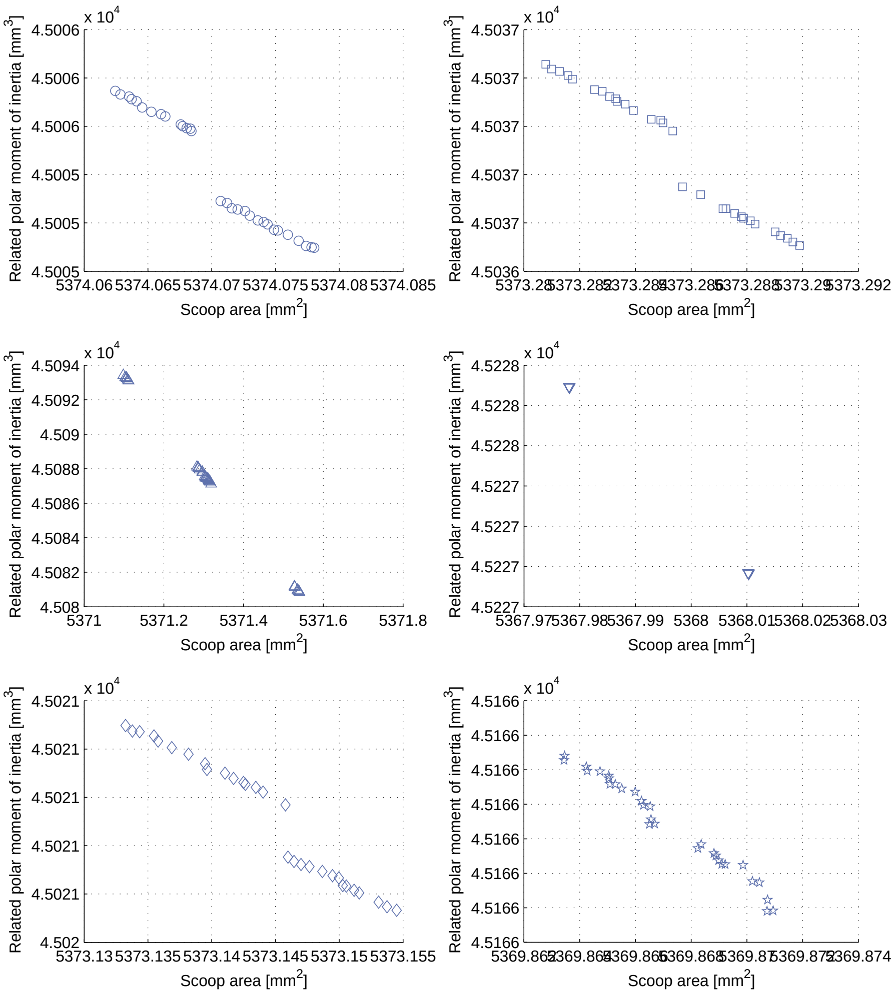


Figure 12.5: Final populations of the optimizations using isotropic mutation

Table 12.1: Results using isotropic mutation

Run	Scoop Area		RPMI	
	Avg.	Stdev.	Avg.	Stdev.
1	5374.07	0.005	45005.41	0.244
2	5373.29	0.003	45036.64	0.136
3	5371.30	0.143	45087.57	4.092
4	5367.99	0.016	45227.46	0.470
5	5373.14	0.006	45020.91	0.266
6	5369.87	0.002	45166.12	0.069
Avg.	5371.61	0.03	45090.69	0.88
Stdev.	2.342	0.056	88.846	1.580

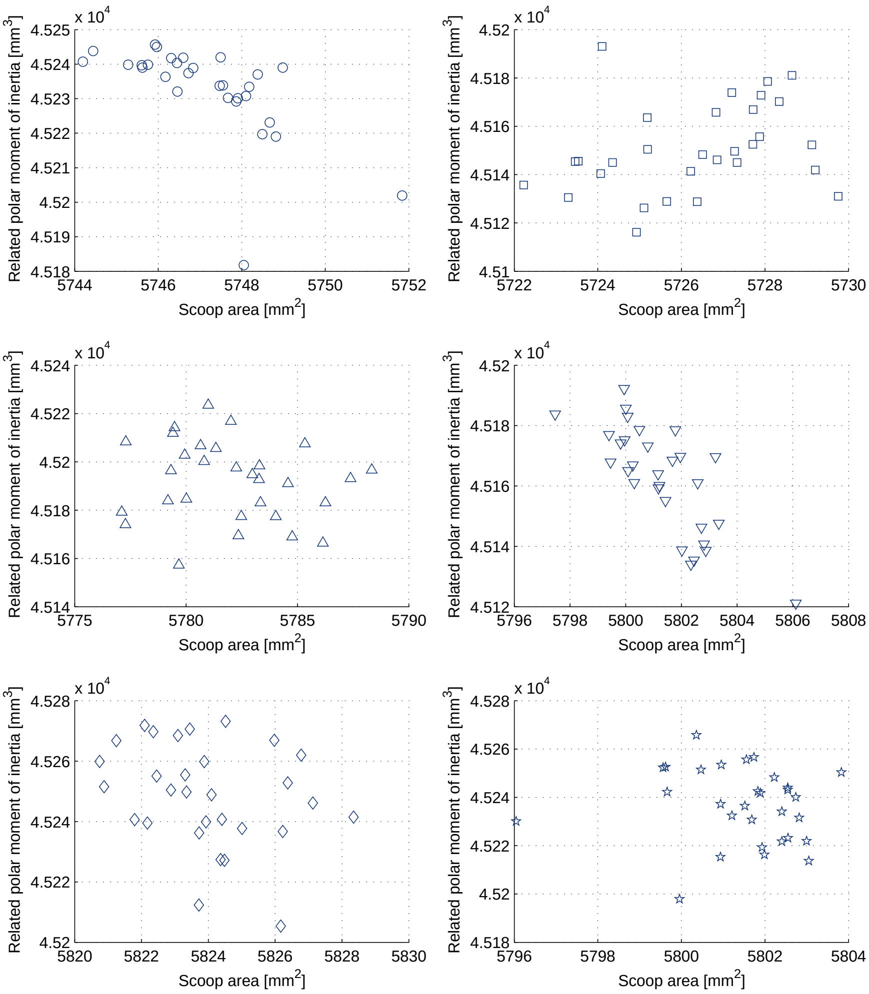


Figure 12.6: Final populations of the optimizations using scaled mutation

Table 12.2: Results using scaled mutation

Run	Scoop Area		RPMI	
	Avg.	Stdev.	Avg.	Stdev.
1	5747.13	1.560	45232.90	13.336
2	5726.33	1.989	45150.76	18.363
3	5782.04	2.966	45192.09	16.403
4	5801.30	1.644	45162.44	17.674
5	5823.97	1.909	45248.82	17.150
6	5801.47	1.506	45236.73	15.627
Avg.	5780.37	1.93	45203.96	16.43
Stdev.	36.909	0.543	41.516	1.790

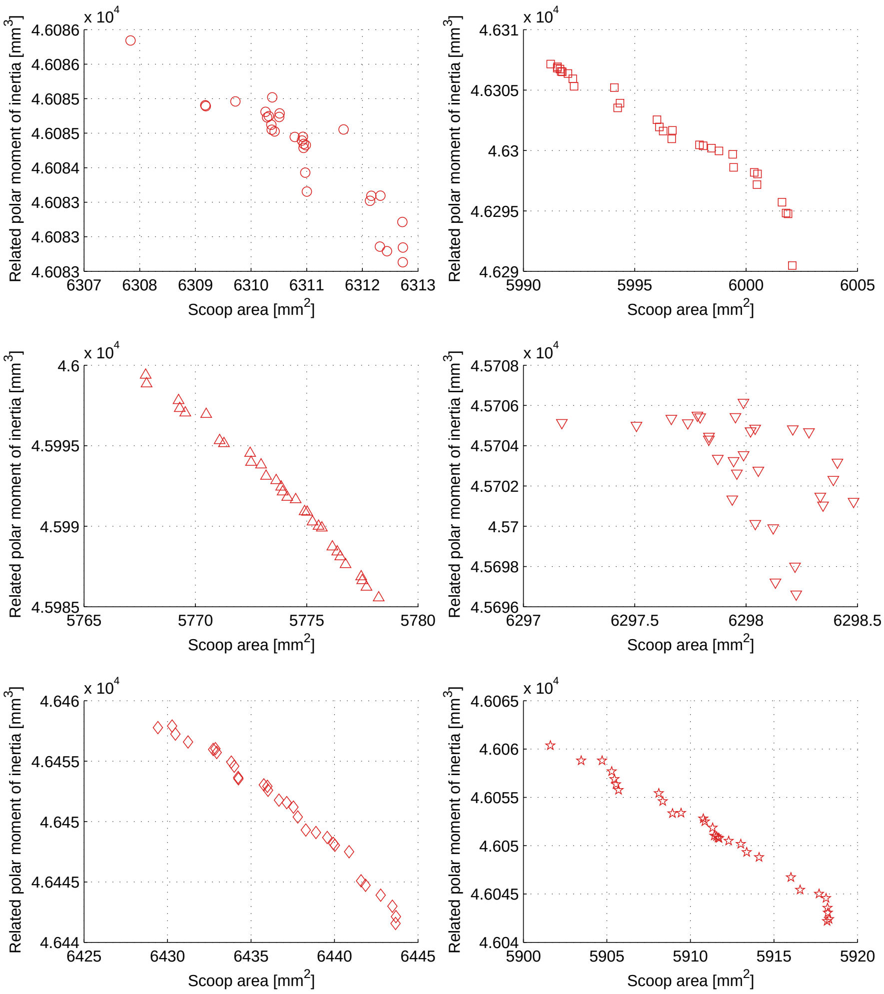


Figure 12.7: Final populations of the optimizations using directed mutation

Table 12.3: Results using directed mutation

Run	Scoop Area		RPMI	
	Avg.	Stdev.	Avg.	Stdev.
1	6310.94	1.164	46084.22	0.785
2	5996.37	3.760	46301.60	4.375
3	5773.69	3.007	45992.14	3.946
4	5911.34	4.940	46051.13	5.234
5	6298.01	0.283	45702.97	2.581
6	6436.93	4.243	46450.95	4.866
Avg.	6083.27	3.25	46099.76	4.20
Stdev.	263.564	1.821	259.100	1.670

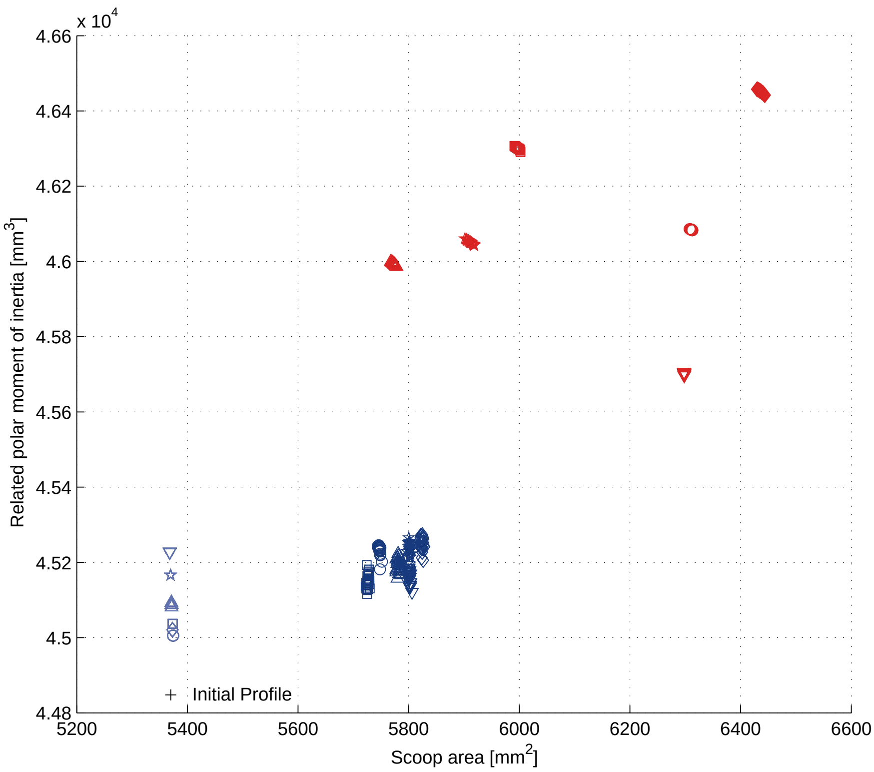


Figure 12.8: Comparison of the simulation results with isotropic mutation depicted in light blue, scaled mutation in dark blue, and directed mutation in red, respectively

relatively sharp Pareto fronts and cover a good spectrum of the search space with respect to both criteria. Last, the directed mutation shows the greatest diversity under and within the different runs. From the practical point of view it enjoys some more advantages. Its density function is simple and random number generation is easy and fast. Taking into account that the application of the mutation principle itself is fast (e.g. compared to the correlated mutation) the use of directed mutation might be quite beneficial for many other real-world optimization problems as well.

The by-product *ScrewView* allows customizing rotors with high efficiency very quick, which is an advantage in competition such that the industry considers the work a major success.

Part IV

Conclusions and Outlook

13 Conclusions

This dissertation aimed at revising the concept of directed mutation. The motivation behind this objective arose when the serious problems of the existing approach became evident.

Directed mutation was introduced as a mutation variant pursuing the target to adapt favorable directions over the generations and sustain further advance into it. To rephrase, for every problem dimension a tendency towards the positive or negative domain should be established by the mutation distribution. Thus, appropriate customizable distributions as the basis of any directed mutation operator were discussed, starting with several techniques recently found in the literature. Then with Hildebrand's *asymmetric mutation* [Hildebrand, 2001] the first serious approach was recapitulated. Denoting for this were the piecewise defined distributions, called Ξ-distributions. Dedicated functions were applied for the negative and positive semi-infinite support. As one part, the corresponding piece of the normal distribution was chosen; the other part was defined by a widened variant of the normal distribution with respect to the abscissa. Due to this construction principle the approach was categorized as constructive approach.

It was discussed that the dependence of the moments on the shape parameters is of crucially importance. Since the variance controls the mutation strength we concluded that it should not be modified implicitly by the shape parameters too much. In the ideal case it would be shape invariant but at least convergence was claimed to be necessary for shape parameters approaching infinity. This demand was shown to be violated by the Ξ-distributions. Therefore with the naïve skew-normal distribution an alternative using the same constructive approach was framed. Since there a normal distribution was applied that was compressed with respect to the abscissa, the moments converged and the influence of the shape was limited. Its moments were calculated and its distribution as well as the inverse of this distribution deduced. However, both methods suffered from their random number generators using the inversion method. Due to the relative complex inverse distribution functions random variate generation was comparatively expensive.

These problems were the motivation to quest for an alternative approach to directed mutation. As solution, the skewing function approach was proposed. The idea here was to multiply an arbitrary symmetric probability density function with a skewing function. This operation amplified one side of the density and attenuated the other at the same amount such that it on average remained the same up to a constant factor. Therefore, several skewing functions were discussed whereof some were cumulative distribution functions. A first approach using the logistic distribution as skewing function together with the normal density was discarded in favor of an approach using the normal distribution as skewing function. With the result, termed as skew-normal distribution, the most prominent member of the more substantive skewing function approach has been introduced. The arising skew distributions were given in closed form, mathematical well tractable, and equipped with a simple and fast random number generator using the density's stochastic representation. Stressed as their most important advantage was however the convergence of their moments. Also, a standardized version with shape invariant variance was derived.

With the appropriate skew distributions being available, next their integration into the Evolution Strategy framework was sketched. The self-adaptation of the skewness parameters as well as a customized recombination scheme were treated. Pseudo-code of a Directed Evolution Strategy was provided and the advantages and disadvantages of the different mutation operators were compared.

All the previously mentioned techniques dealt with usual Evolution Strategies as basis. But, as has already been hinted in the introduction, with the Covariance Matrix Adaptation-Evolution Strategy there exists another, more powerful method. How to equip this one with the concept of directed mutation was explained next. Especially the adaptation of the shape parameters was investigated since it was twofold: both, inter- and intragenerational mechanisms were possible and even combinations of these. Several different intragenerational update mechanisms were presented and compared.

After the investigations into the fundamental principles both, the Directed Evolution Strategy and the Directed Covariance Matrix Adaptation-Evolution Strategy were empirically validated. First single objective test functions were evaluated. It emerged that directed mutation by means of the skew-normal, standardized skew-normal, and naïve skew-normal distribution clearly outperformed the other mutation operators. In the case of correlated mutations the directed variant also performed better on all functions. As another result it turned out that the Directed Evolution Strategy needed up to ten times more function evaluations than the Directed Covariance Matrix Adaptation-Evolution Strategy on the same, but even lower dimensional problems.

Besides the usual single objective test functions a constrained, multiobjective design problem from mechanical engineering was studied: the rotor profile of a screw-type machine had to be optimized. Again, the Directed Evolution Strategy significantly outperformed the other mutation strategies; noticing that all results of it but one dominated all runs of the other strategies. Further, the resulting solutions formed relatively sharp Pareto fronts, covering a good spectrum of the search space at the same time.

14 Outlook

This thesis developed the fundamentals of two Directed Evolutionary Algorithms, namely the Directed Evolution Strategy and the Directed Covariance Matrix Adaptation-Evolution Strategy. Naturally there remain several open questions whereof some will be listed as topics for future work. Since the Directed Evolution Strategy is clearly outperformed by the Directed Covariance Matrix Adaptation-Evolution Strategy, attention will be focused on the latter one.

An aspect meriting future work might be to further investigate the potential offered by the intra- and intergenerational mutation shape adaptation. Especially the usual intergenerational adaptation is at the moment rather treated as an orphan. Thus, a next step to improve the procedure would be to optimize the ratio of the two mechanisms as well as the other control parameters like the learning rates in greater detail.

A notable remark was made by Prof. Dr. Reusch regarding skew distributions. He proposed to supersede the representation form of the distribution's shape information. Instead of a covariance matrix $\mathbf{C}$ the use of a simplified ellipsoid is proposed.

In the 2-dimensional case, an ellipse is defined as the set E of points P with equal sum of distances to the two given focal points F_1 and F_2,

$$E = \{P \mid |\overline{F_1 P}| + |\overline{F_2 P}| = d\}.$$

Hence, the two foci F_1 and F_2 together with the distance d define the ellipse. Equivalently, it can be defined via the center point $\mathbf{x}$, which represents an individual, the difference vector $F_1 - \mathbf{x}$, and the distance d. The difference vector $F_2 - \mathbf{x}$ then is given due to symmetry, see Figure 14.1.

Instead of using ellipsoids in higher dimensions, i.e. the higher dimensional analogues of an ellipse, Prof. Dr. Reusch proposed to retain the use of an ellipse as a solid of revolution.

The benefit becomes obvious by comparing the needed parameters to define the different objects. In the n-dimensional case, the covariance matrix comprises $n(n+1)/2$ free entries[1]. To define an ellipsoid $n^2 + 1$ parameters have to be provided, while the proposed alternative gets along with $2n + 1$ parameters. Hence, the afford of managing, adapting, and storing parameters is significantly reduced. Of course this reduced complexity in encoding concurs with also reduced possibilities in representation, i.e. shape variants. However, the presumption is that it will still meet the main demands. Note the affinity of

[1]neglecting the restriction of its positive definiteness

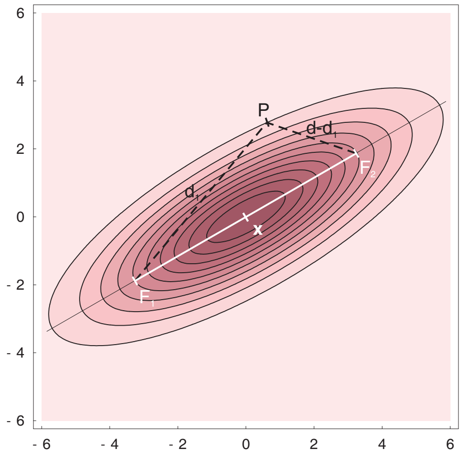


Figure 14.1: Ellipses with varied distances d

this approach with the one of Ghozeil and Fogel [1996], see subsection *Directed Mutations by Means of Direction Vectors* on page 68. In fact both rely on the directional information to be encoded in a direction vector. The main open task is to provide an efficient random number generator.

What remained undone are theoretical investigations into the convergence properties of Directed Evolutionary Algorithms. If such a theory can be formulated at all in the near future seems questionable noticing the lack of a general theory even for conventional Evolution Strategies. What makes things difficult in this context are on the on hand the consequences caused by recombination of individuals and on the other the potential diversity of the mutation distributions. The endogenous strategy parameters and their self-adaptation ought to be mentioned here.

However, a turn from local progress measures such as the progress rate, as proposed by Beyer [2001], towards global convergence results can be observed. Auger investigated for example a non-isotropic adaptive $(1, \lambda)$-Evolution Strategy by means of the supermartingale theory [Auger et al., 2003] and the theory of φ-irreducible Markov chains [Auger, 2005]. An interesting topic for ongoing research is generalizing this work to directed mutation by introducing shape parameters.

Appendix

A Publications and History

This appendix is devoted to work being published previously to this thesis. First, I would like to acknowledge the contributions of my co-authors of the several publications cited in this thesis. In addition, I want to relate this work and provide some background information about the historical development. Last, there are some further publications to be mentioned that have not been cited here.

I would like to begin this survey with a paper not referenced in this thesis, representing however the starting point of the work on directed mutation.

> *Stefan Berlik:* **Polymorphe Mutation.** 22. Workshop "Interdisziplinäre Methoden in der Informatik", Forschungsbericht Nr. 783, Universität Dortmund, 2003.

Key topics are a systematical analysis of the asymmetric mutation and a first proposal of an improved directed mutation, called *polymorphic mutation*. The immanent problem of the diverging moments of the former is revealed and fixed. However, the new proposed mutation operator still comprehends some unresolved problems. A revised version was later presented at the EUSFLAT conference [Berlik, 2003a].

The next two papers can be seen as milestones on the way to directed mutation. They originate from several valuable and fruitful discussions with B. Reusch and were written mainly by me. The following one is the first publication of a directed mutation operator by means of the skew-normal distribution.

> *Stefan Berlik and Bernd Reusch:* **Directed mutation by means of the skew-normal distribution.** In B. Reusch, editor, *Proc. of the Int. Conf. on Computational Intelligence, FUZZY DAYS, Sep. 29–Oct. 1, 2004, Dortmund, Germany*, Advances in Soft Computing, pages 35–50. Springer-Verlag Berlin Heidelberg, 2004.

A first systematical comparison of different directed mutation operators, including besides others also the standardized skew-normal mutation operator, is provided in the next paper.

> *Stefan Berlik and Bernd Reusch:* **Directed mutation operators – an overview.** In R. Khosla, R. J. Howlett, and L. C. Jain, editors, *Proc. of the 9th Int. Conf. on Knowledge-Based & Intelligent Information & Engineering Systems, KES 2005, Sept. 14–16, 2005, Melbourne, Australia*, volume 3683 of *Lecture Notes in Computer Science*, pages 1151–1159. Springer-Verlag Berlin Heidelberg, 2005.

There is another, although unpublished noteworthy presentation in the context of the advent of directed mutation. It was held at the PhD students seminar of chair I and listed all the many trials, setbacks and dilemmas happening over the time.

> *Stefan Berlik:* **Directed Mutation for Evolutionary Algorithms.** Presentation held at: *PhD students seminar of chair I, Department of computer science, Universität Dortmund*, Universität Dortmund, 2004 (unpublished).

Some parts of the early work on the Directed Evolution Strategy have also been published in a book chapter. There I wrote down the theoretical foundations of directed mutation in one single place, focused to be used in an engineering context.

> *Stefan Berlik and Bernd Reusch:* **Foundations of Directed Mutation.** In Xuan F. Zha and Robert J. Howlett, eds., *Integrated Intelligent Systems for Engineering Design*, IOS Press, 2006.

The following four journal publications resulted from our cooperation with Prof. Dr.-Ing. Kauder's chair and deal with the basics of screw-type machine optimization. The first three were published in succession in the *Schraubenmaschinen* journal. They were written in equal parts by M. Helpertz and me with M. Helpertz contributing the mechanical engineering parts and me the parts dealing with optimization.

> *Knut Kauder, Bernd Reusch, Markus Helpertz, and Stefan Berlik:* **Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 1.** *Schraubenmaschinen*, 9:27–46, 2001.

> *Knut Kauder, Bernd Reusch, Markus Helpertz, and Stefan Berlik:* **Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 2.** *Schraubenmaschinen*, 10:17–34, 2002.

> *Knut Kauder, Bernd Reusch, Markus Helpertz, and Stefan Berlik:* **Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 3.** *Schraubenmaschinen*, 11:15–29, 2003.

The article published in the *VDI Berichte* journal can be seen as a condensate of the first two articles mentioned above, supplemented with exemplary optimization results and was written in equal parts by M. Helpertz and me. The main new topic concerning the optimization was the presentation of the extended offspring generation scheme.

> *Knut Kauder, Bernd Reusch, Markus Helpertz, and Stefan Berlik:* **Optimisation methods for rotors of twin-screw compressors.** *VDI Berichte*, 1715:29–50, 2002b.

The next papers deal primary with the application of directed mutation in the mechanical engineering context. The following two were mainly written by me but profited essentially from constructive and inspiring discussions with M. Fathi. The first focuses on multiobjective optimization using directed mutation in general,

> *Stefan Berlik and Madjid Fathi:* **Multi-objective optimization using directed mutation.** In H. R. Arabnia and R. Joshua, editors, *Proc. of the Int. Conf. on Artificial Intelligence, ICAI 2005, June 27–30, 2005, Las Vegas, USA*, volume II, pages 870–875. CSREA Press, USA, 2005.

while the second one essentially presents *ScrewView*, a CAD tool to design and optimize screw-type machines.

> *Stefan Berlik and Madjid Fathi:* **A design and optimization tool for screw-type machines.** In *Proc. of the Int. Conf. on Systems, Man and Cybernetics, IEEE SMC 2005, Oct. 10–12, 2005, Hawaii, USA*. IEEE Press, 2005.

The paper I would like to comment last will be presented at the IEEE SMC 2006 conference in Taipei and is the result of discussions between M. Fathi, A. Holland, and me. I wrote down some first ideas of intragenerational mutation shape adaptation to be used to optimize screw-type machines.

> *Stefan Berlik, Madjid Fathi, and Alexander Holland:* **Advances in optimizing screw-type machines.** In *Proc. of the Int. Conf. on Systems, Man and Cybernetics, IEEE SMC 2006, Oct. 8–11, 2006, Taipei, Taiwan* . IEEE Press, 2006 [to appear].

B Error Function

The non-elementary error function $\operatorname{erf}(\cdot)$ is strongly related to the standard normal distribution as it is encountered in integrating the probability density function of the latter. It is defined by

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad \forall x \in \mathbb{R}. \tag{B.1}$$

The function is restricted in range and rapidly approaches the limits ± 1 as $x \to \pm\infty$. A plot of the error function is given in Figure B.1.

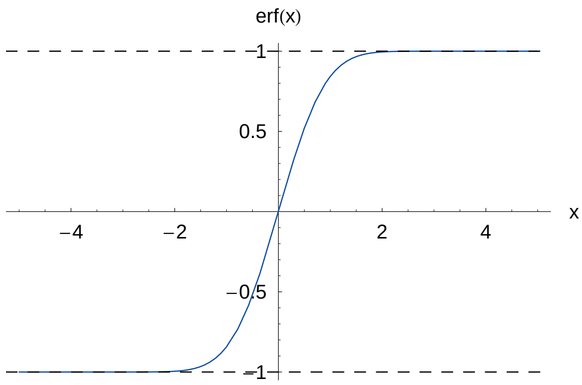


Figure B.1: Plot of the error function. The dashed lines indicate the limits.

As the error function is transcendental, i.e. the integral defining the error function (B.1) cannot be evaluated in closed form in terms of elementary functions, function values for arbitrary arguments cannot be calculated directly. To evaluate the function for example the following Maclaurin series expansion can be used

$$\operatorname{erf} x = \frac{x}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{\left(-x^2\right)^n}{n!\left(n+\frac{1}{2}\right)} = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \cdots \right). \tag{B.2}$$

Further, most computer algebra systems provide function implementations or a Fortran program by Cody [1969] or a C procedure [Press et al., 1993] can be used. For a more comprehensive discussion on the error function cf. [Spanier and Oldham, 1987, Ch. 40] or [Abramowitz and Stegun, 1972, Ch. 7].

B.1 Standard Normal Distribution Function

The error function essentially resembles the standard normal cumulative distribution function, as they differ only by scaling and translation. The standard normal cumulative distribution function, denoted by $\Phi(x)$, is defined as

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \tag{B.3}$$

illustrated in Figure B.2.

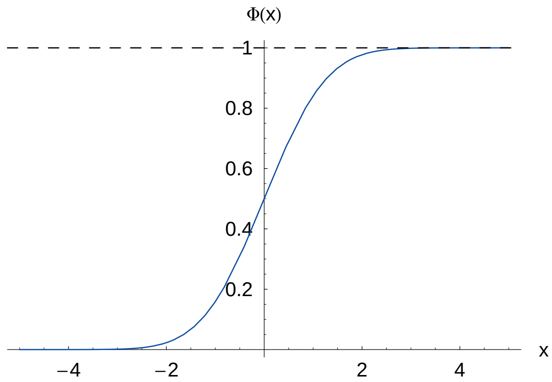


Figure B.2: Plot of the standard normal cumulative distribution function

B.2 Inverse Error Function

The inverse error function $\operatorname{erf}^{-1}(\cdot)$ is the inverse of the error function such that

$$\operatorname{erf}^{-1}(\operatorname{erf} x) = x, \tag{B.4}$$

$$\operatorname{erf}(\operatorname{erf}^{-1} x) = x. \tag{B.5}$$

It is defined for $-1 < x < 1$ with the special values $\operatorname{erf}^{-1}(-1) = -\infty$ and $\operatorname{erf}^{-1}(1) = \infty$, illustrated in Figure B.3. Again, most computer algebra systems provide function implementations of the inverse error function or for example the approximation method proposed by Hill and Davis [1973] can be used.

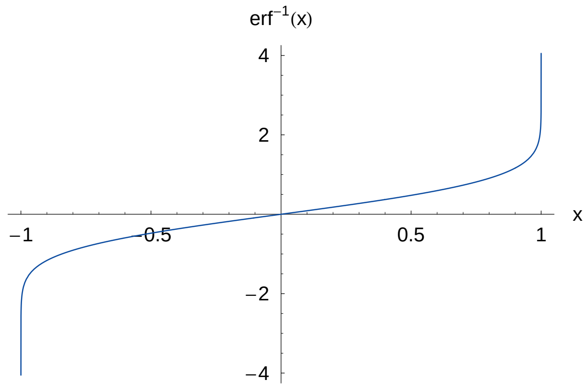


Figure B.3: Plot of the inverse error function

C Test Functions

C.1 Sphere Function

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

with

$$-1 \leq x_i \leq 1$$

Optimum

$$\min(f_{\text{sphere}}) = f_{\text{sphere}}(\mathbf{0}) = 0$$

Plot of the 2-dimensional variant

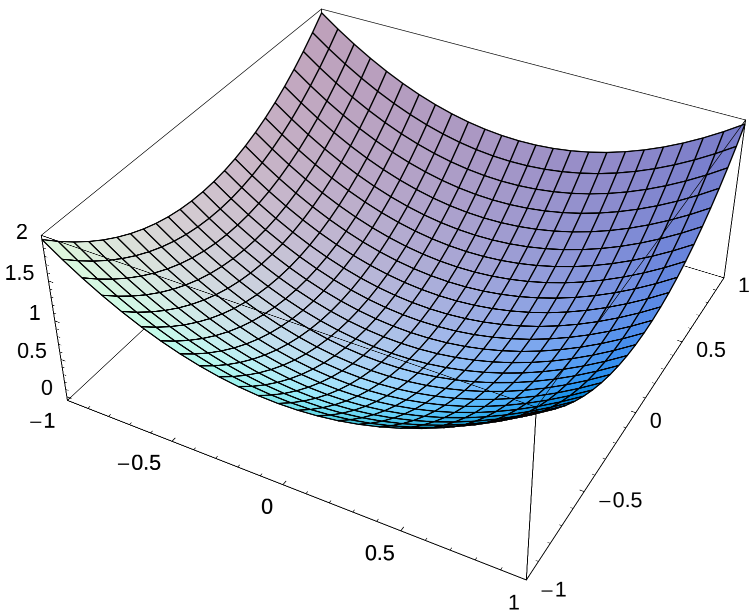


Figure C.1: Plot of the 2-dimensional sphere function

C.2 Schwefel's Function 1.2

$$f_{\text{Schwefel}}(\mathbf{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2 \tag{C.2}$$

with

$$-1 \leq x_i \leq 1$$

Optimum

$$\min(f_{\text{Schwefel}}) = f_{\text{Schwefel}}(\mathbf{0}) = 0$$

Plot of the 2-dimensional variant

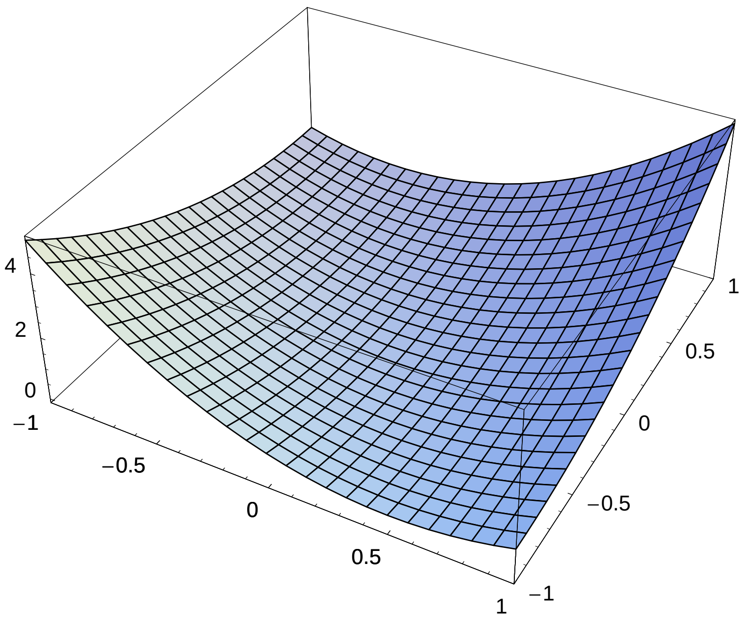


Figure C.2: Plot of the 2-dimensional Schwefel function 1.2

C.3 Cigar Function

$$f_{\text{cigar}}(\mathbf{x})\S = x_1^2 + \sum_{i=2}^{n}(1000x_i)^2 \quad \text{(C.3)}$$

with

$$-1 \leq x_i \leq 1$$

Optimum

$$\min(f_{\text{cigar}}) = f_{\text{cigar}}(\mathbf{0}) = 0$$

Plot of the 2-dimensional variant

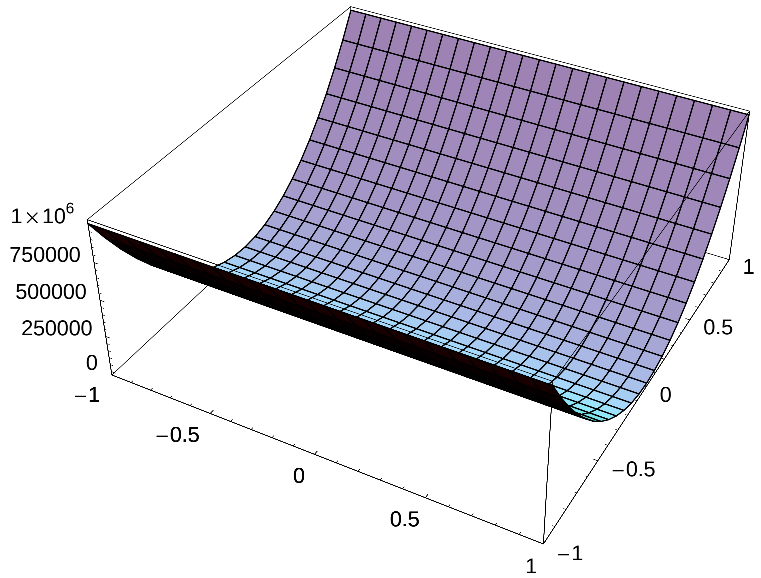


Figure C.3: Plot of the 2-dimensional cigar function

C.4 Tablet Function

$$f_{\text{tablet}}(\mathbf{x}) = (1000x_1)^2 + \sum_{i=2}^{n} x_i^2 \qquad \text{(C.4)}$$

with

$$-1 \leq x_i \leq 1$$

Optimum

$$\min(f_{\text{tablet}}) = f_{\text{tablet}}(\mathbf{0}) = 0$$

Plot of the 2-dimensional variant

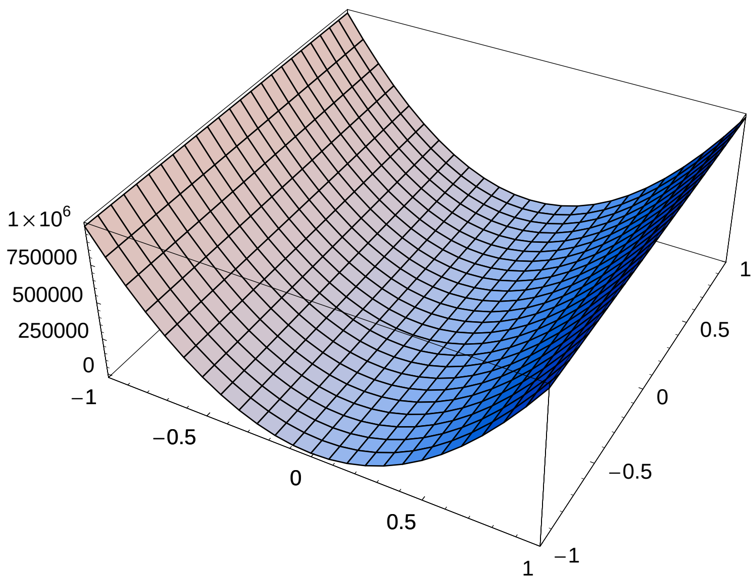


Figure C.4: Plot of the 2-dimensional tablet function

C.5 Ellipsoid Function

$$f_{\text{elli}}(\mathbf{x}) = \sum_{i=1}^{n} \left(1000^{\frac{i-1}{n-1}} x_i\right)^2 \tag{C.5}$$

with

$$-1 \leq x_i \leq 1$$

Optimum

$$\min(f_{\text{elli}}) = f_{\text{elli}}(\mathbf{0}) = 0$$

Plot of the 2-dimensional variant

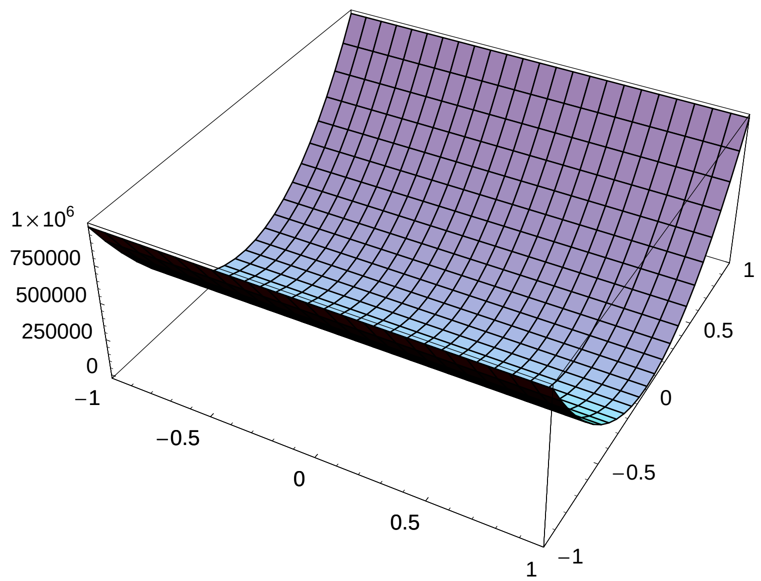


Figure C.5: Plot of the 2-dimensional ellipsoid function

C.6 Parabolic Ridge Function

$$f_{\text{parabR}}(\mathbf{x}) = -x_1 + 100 \sum_{i=2}^{n} x_i^2 \tag{C.6}$$

with

$$-1 \leq x_i \leq 1$$

Plot of the 2-dimensional variant

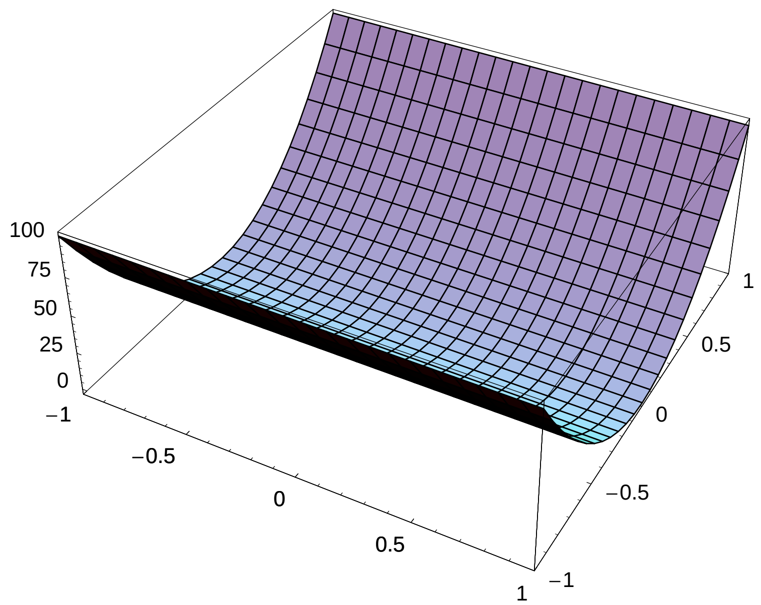


Figure C.6: Plot of the 2-dimensional parabolic ridge function

C.7 Sharp Ridge Function

$$f_{\mathrm{sharpR}}(\mathbf{x})\S = -x_1 + 100\sqrt{\sum_{i=2}^{n} x_i^2} \tag{C.7}$$

with

$$-1 \leq x_i \leq 1$$

Plot of the 2-dimensional variant

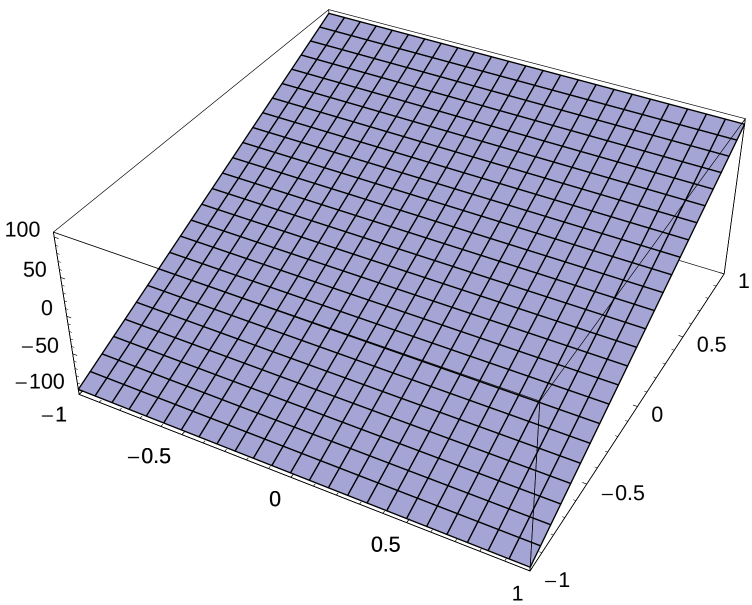


Figure C.7: Plot of the 2-dimensional sharp ridge function

C.8 Generalized Rosenbrock Function

$$f_{\text{Rosen}}(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100 \left(x_{i+1} - x_i^2 \right)^2 + (x_i - 1)^2 \right) \tag{C.8}$$

with

$$\mathbf{x} = \mathbf{0}$$

Optimum

$$\min(f_{\text{Rosen}}) = f_{\text{Rosen}}(\mathbf{1}) = 0$$

Plot of the 2-dimensional variant

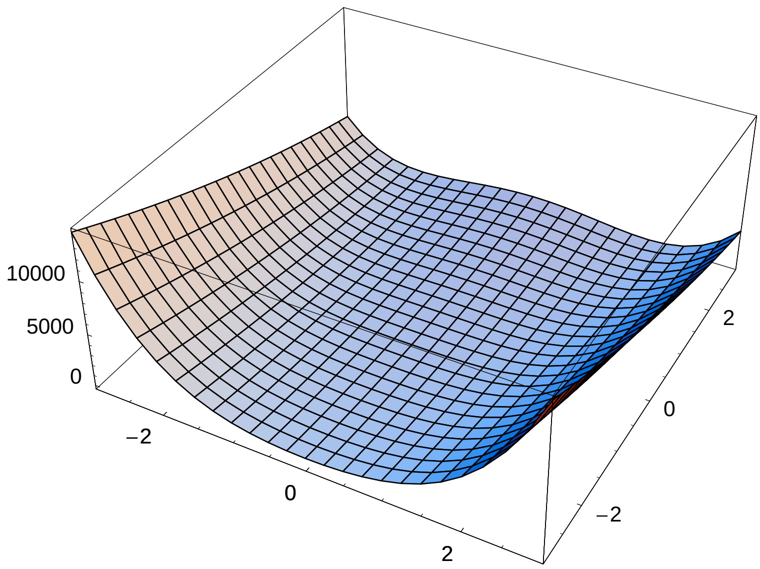


Figure C.8: Plot of the 2-dimensional Rosenbrock Function

D Tables

D.1 Sphere Function

Table D.1: Local update for 2-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.32	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.32	0.33
	σ	0.468	0.339	0.283	0.248	0.225	0.196	0.174	0.152	0.138	0.128	0.121	0.112	0.112
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.33	0.35	0.33	0.35	0.35	0.35
2	$\bar{x}$	0.33	0.35	0.36	0.37	0.39	0.41	0.44	0.48	0.49	0.53	0.55	0.57	0.59
	σ	0.469	0.348	0.296	0.269	0.252	0.23	0.221	0.211	0.205	0.204	0.21	0.214	0.211
	m	0	0.5	0.33	0.25	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.66
3	$\bar{x}$	0.33	0.35	0.38	0.41	0.42	0.45	0.48	0.51	0.53	0.56	0.57	0.59	0.6
	σ	0.472	0.35	0.308	0.281	0.271	0.247	0.235	0.226	0.221	0.218	0.22	0.219	0.217
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.63	0.67	0.68
4	$\bar{x}$	0.32	0.34	0.34	0.36	0.36	0.38	0.4	0.44	0.47	0.5	0.54	0.57	0.6
	σ	0.465	0.351	0.299	0.273	0.256	0.232	0.22	0.209	0.206	0.209	0.201	0.198	0.189
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.47	0.5	0.53	0.57	0.63	0.65
5	$\bar{x}$	0.32	0.35	0.36	0.37	0.39	0.42	0.45	0.49	0.51	0.56	0.57	0.59	0.61
	σ	0.468	0.361	0.312	0.288	0.271	0.252	0.236	0.223	0.221	0.208	0.201	0.208	0.201
	m	0	0.5	0.33	0.25	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.65	0.68
6	$\bar{x}$	0.34	0.36	0.38	0.4	0.42	0.45	0.48	0.53	0.54	0.57	0.58	0.6	0.61
	σ	0.472	0.351	0.304	0.276	0.262	0.243	0.226	0.217	0.218	0.208	0.212	0.205	0.207
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.64	0.67	0.68
7	$\bar{x}$	0.33	0.38	0.41	0.44	0.46	0.5	0.53	0.55	0.57	0.6	0.61	0.62	0.63
	σ	0.469	0.357	0.315	0.287	0.274	0.25	0.237	0.225	0.217	0.213	0.21	0.203	0.197
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.6	0.67	0.68	0.68	0.7
8	$\bar{x}$	0.32	0.35	0.35	0.37	0.39	0.41	0.45	0.49	0.52	0.55	0.57	0.6	0.61
	σ	0.468	0.358	0.305	0.288	0.268	0.248	0.23	0.218	0.212	0.201	0.199	0.2	0.197
	m	0	0.5	0.33	0.25	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.66	0.68
9	$\bar{x}$	0.33	0.35	0.37	0.4	0.42	0.45	0.49	0.53	0.55	0.58	0.59	0.61	0.64
	σ	0.468	0.367	0.323	0.303	0.286	0.267	0.244	0.227	0.215	0.208	0.201	0.202	0.181
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.65	0.67	0.69
10	$\bar{x}$	0.32	0.34	0.36	0.38	0.4	0.43	0.45	0.5	0.52	0.56	0.58	0.59	0.6
	σ	0.468	0.368	0.325	0.304	0.29	0.271	0.251	0.235	0.229	0.206	0.203	0.205	0.191
	m	0	0.5	0.33	0.25	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.65	0.66
11	$\bar{x}$	0.34	0.36	0.39	0.4	0.41	0.44	0.48	0.5	0.53	0.55	0.57	0.6	0.61
	σ	0.472	0.379	0.342	0.319	0.306	0.288	0.269	0.25	0.233	0.224	0.214	0.205	0.196
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.65	0.66
12	$\bar{x}$	0.34	0.36	0.38	0.4	0.41	0.44	0.47	0.51	0.53	0.56	0.58	0.6	0.61
	σ	0.472	0.354	0.299	0.273	0.258	0.237	0.22	0.21	0.209	0.208	0.202	0.199	0.199
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.67	0.68
13	$\bar{x}$	0.32	0.37	0.41	0.43	0.45	0.48	0.51	0.55	0.57	0.59	0.6	0.62	0.63
	σ	0.468	0.355	0.313	0.285	0.267	0.247	0.23	0.22	0.214	0.213	0.207	0.201	0.201
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.67	0.65	0.68	0.7
14	$\bar{x}$	0.33	0.36	0.38	0.4	0.42	0.45	0.49	0.53	0.55	0.58	0.61	0.61	0.62
	σ	0.471	0.365	0.321	0.296	0.282	0.262	0.243	0.226	0.224	0.211	0.205	0.203	0.205
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.6	0.63	0.68	0.67	0.69
15	$\bar{x}$	0.33	0.37	0.39	0.42	0.45	0.49	0.52	0.56	0.57	0.6	0.62	0.62	0.63
	σ	0.469	0.374	0.333	0.309	0.291	0.269	0.25	0.231	0.226	0.207	0.202	0.203	0.195
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.6	0.67	0.68	0.68	0.7

Table D.2: Local update for 5-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.36	0.37	0.37
	σ	0.482	0.346	0.278	0.24	0.218	0.187	0.155	0.128	0.115	0.096	0.083	0.071	0.063
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.33	0.35	0.37	0.38	0.37	0.38
2	$\bar{x}$	0.36	0.37	0.38	0.38	0.38	0.39	0.4	0.42	0.44	0.47	0.5	0.54	0.57
	σ	0.48	0.342	0.282	0.245	0.222	0.193	0.162	0.137	0.126	0.11	0.096	0.087	0.082
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.47	0.5	0.55	0.59
3	$\bar{x}$	0.36	0.39	0.4	0.41	0.43	0.46	0.49	0.53	0.56	0.61	0.62	0.66	0.67
	σ	0.479	0.347	0.288	0.26	0.229	0.201	0.174	0.148	0.129	0.113	0.101	0.09	0.088
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.67	0.69
4	$\bar{x}$	0.38	0.37	0.38	0.38	0.38	0.38	0.39	0.4	0.41	0.44	0.46	0.5	0.53
	σ	0.478	0.35	0.298	0.266	0.244	0.215	0.192	0.164	0.148	0.127	0.109	0.098	0.085
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.5	0.54
5	$\bar{x}$	0.35	0.38	0.39	0.4	0.41	0.43	0.45	0.48	0.51	0.55	0.59	0.62	0.64
	σ	0.484	0.344	0.285	0.245	0.223	0.194	0.165	0.141	0.126	0.11	0.105	0.098	0.086
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.57	0.6	0.62	0.65
6	$\bar{x}$	0.36	0.38	0.39	0.4	0.41	0.42	0.45	0.48	0.51	0.55	0.58	0.61	0.65
	σ	0.481	0.346	0.284	0.255	0.226	0.197	0.169	0.145	0.131	0.108	0.1	0.092	0.075
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.57	0.57	0.62	0.65
7	$\bar{x}$	0.36	0.41	0.44	0.46	0.47	0.51	0.54	0.58	0.61	0.64	0.65	0.68	0.69
	σ	0.479	0.351	0.295	0.265	0.237	0.206	0.173	0.143	0.123	0.104	0.1	0.092	0.086
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.68	0.7
8	$\bar{x}$	0.36	0.37	0.37	0.38	0.38	0.4	0.41	0.44	0.46	0.51	0.53	0.58	0.61
	σ	0.481	0.346	0.286	0.252	0.229	0.204	0.179	0.156	0.135	0.122	0.113	0.096	0.086
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.58	0.61
9	$\bar{x}$	0.37	0.39	0.39	0.42	0.44	0.46	0.49	0.54	0.56	0.59	0.61	0.64	0.66
	σ	0.482	0.362	0.313	0.281	0.258	0.227	0.197	0.167	0.147	0.123	0.104	0.092	0.086
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.65	0.66
10	$\bar{x}$	0.36	0.37	0.37	0.38	0.38	0.39	0.4	0.41	0.43	0.46	0.48	0.5	0.51
	σ	0.479	0.349	0.291	0.26	0.241	0.211	0.189	0.169	0.163	0.15	0.142	0.13	0.129
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.47	0.5	0.5	0.53
11	$\bar{x}$	0.37	0.38	0.38	0.39	0.39	0.4	0.42	0.43	0.44	0.46	0.49	0.5	0.51
	σ	0.482	0.37	0.322	0.292	0.273	0.25	0.229	0.208	0.195	0.18	0.164	0.154	0.147
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.47	0.45	0.47	0.5	0.52	0.53
12	$\bar{x}$	0.37	0.37	0.38	0.38	0.38	0.39	0.4	0.41	0.43	0.46	0.48	0.52	0.56
	σ	0.483	0.343	0.282	0.24	0.223	0.188	0.161	0.134	0.12	0.098	0.09	0.078	0.068
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.53	0.56
13	$\bar{x}$	0.37	0.38	0.39	0.42	0.42	0.45	0.48	0.52	0.55	0.59	0.62	0.65	0.67
	σ	0.482	0.342	0.286	0.25	0.229	0.199	0.17	0.143	0.124	0.107	0.096	0.086	0.078
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.66	0.68
14	$\bar{x}$	0.38	0.37	0.39	0.4	0.41	0.43	0.44	0.48	0.52	0.55	0.58	0.63	0.65
	σ	0.485	0.346	0.288	0.253	0.233	0.204	0.177	0.152	0.138	0.118	0.108	0.09	0.088
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.57	0.6	0.63	0.66
15	$\bar{x}$	0.37	0.41	0.43	0.45	0.47	0.5	0.54	0.58	0.6	0.64	0.66	0.69	0.69
	σ	0.483	0.354	0.304	0.272	0.247	0.215	0.182	0.15	0.134	0.11	0.1	0.091	0.086
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.68	0.7	0.7

Table D.3: Local update for 10-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.4	0.4	0.39	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
	σ	0.491	0.346	0.282	0.243	0.22	0.186	0.158	0.13	0.11	0.094	0.079	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
2	$\bar{x}$	0.4	0.4	0.4	0.4	0.41	0.41	0.41	0.41	0.42	0.43	0.45	0.46	0.49
	σ	0.49	0.342	0.281	0.247	0.219	0.184	0.156	0.129	0.111	0.095	0.082	0.071	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.47	0.49
3	$\bar{x}$	0.4	0.41	0.42	0.43	0.44	0.46	0.48	0.51	0.54	0.58	0.61	0.65	0.67
	σ	0.49	0.351	0.284	0.249	0.23	0.197	0.165	0.139	0.118	0.096	0.086	0.072	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.66
4	$\bar{x}$	0.39	0.4	0.4	0.4	0.41	0.4	0.4	0.41	0.41	0.41	0.42	0.44	0.45
	σ	0.487	0.345	0.283	0.246	0.223	0.188	0.158	0.129	0.114	0.092	0.085	0.073	0.066
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.43	0.45
5	$\bar{x}$	0.39	0.4	0.41	0.41	0.42	0.43	0.45	0.47	0.49	0.53	0.56	0.59	0.62
	σ	0.489	0.352	0.293	0.256	0.234	0.204	0.174	0.149	0.136	0.113	0.099	0.081	0.074
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.6	0.63
6	$\bar{x}$	0.4	0.4	0.41	0.41	0.42	0.42	0.43	0.45	0.46	0.49	0.52	0.56	0.6
	σ	0.489	0.344	0.285	0.245	0.223	0.189	0.158	0.131	0.116	0.098	0.086	0.072	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
7	$\bar{x}$	0.4	0.42	0.44	0.47	0.48	0.5	0.52	0.57	0.59	0.62	0.65	0.68	0.7
	σ	0.489	0.355	0.291	0.256	0.225	0.195	0.164	0.133	0.113	0.094	0.08	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.6	0.63	0.65	0.68	0.7
8	$\bar{x}$	0.4	0.4	0.41	0.41	0.41	0.41	0.42	0.42	0.43	0.46	0.48	0.51	0.54
	σ	0.489	0.346	0.29	0.25	0.226	0.185	0.165	0.139	0.12	0.106	0.092	0.082	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
9	$\bar{x}$	0.39	0.41	0.42	0.43	0.44	0.46	0.49	0.52	0.54	0.57	0.59	0.63	0.65
	σ	0.487	0.363	0.3	0.269	0.245	0.211	0.185	0.154	0.137	0.11	0.099	0.086	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.65
10	$\bar{x}$	0.4	0.4	0.4	0.41	0.4	0.4	0.4	0.41	0.41	0.42	0.42	0.43	0.44
	σ	0.491	0.356	0.286	0.253	0.226	0.198	0.168	0.144	0.13	0.116	0.108	0.106	0.101
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
11	$\bar{x}$	0.4	0.41	0.4	0.41	0.41	0.41	0.41	0.42	0.42	0.43	0.43	0.44	0.45
	σ	0.49	0.369	0.311	0.28	0.261	0.233	0.207	0.185	0.172	0.156	0.145	0.133	0.128
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.45	0.45
12	$\bar{x}$	0.4	0.4	0.4	0.39	0.4	0.4	0.4	0.4	0.41	0.41	0.42	0.43	0.44
	σ	0.49	0.345	0.284	0.242	0.223	0.184	0.158	0.124	0.112	0.088	0.081	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.43	0.44
13	$\bar{x}$	0.41	0.41	0.41	0.42	0.42	0.43	0.44	0.47	0.49	0.51	0.55	0.59	0.62
	σ	0.492	0.348	0.284	0.243	0.22	0.185	0.159	0.132	0.119	0.096	0.086	0.069	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.55	0.58	0.62
14	$\bar{x}$	0.39	0.4	0.41	0.42	0.41	0.42	0.43	0.45	0.47	0.5	0.52	0.56	0.59
	σ	0.489	0.344	0.289	0.251	0.223	0.192	0.16	0.134	0.119	0.096	0.089	0.07	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
15	$\bar{x}$	0.4	0.42	0.45	0.47	0.48	0.5	0.53	0.57	0.59	0.62	0.65	0.68	0.71
	σ	0.489	0.357	0.297	0.257	0.232	0.197	0.168	0.134	0.117	0.096	0.084	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.6	0.63	0.65	0.68	0.71

Table D.4: Local update for 20-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.44	0.43	0.42	0.43	0.43	0.43	0.42	0.43	0.43	0.42	0.42	0.42	0.43
	σ	0.496	0.35	0.286	0.249	0.225	0.19	0.157	0.129	0.111	0.09	0.079	0.067	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.42	0.42
2	$\bar{x}$	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.44	0.45
	σ	0.494	0.35	0.284	0.249	0.224	0.186	0.156	0.127	0.108	0.091	0.079	0.066	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.45	0.45
3	$\bar{x}$	0.42	0.43	0.43	0.44	0.45	0.46	0.47	0.49	0.5	0.55	0.57	0.61	0.64
	σ	0.494	0.352	0.283	0.248	0.227	0.191	0.161	0.132	0.114	0.092	0.083	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.57	0.62	0.64
4	$\bar{x}$	0.43	0.43	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.495	0.352	0.282	0.25	0.222	0.19	0.161	0.128	0.113	0.092	0.079	0.064	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.44
5	$\bar{x}$	0.43	0.43	0.43	0.44	0.44	0.44	0.46	0.47	0.48	0.5	0.52	0.56	0.59
	σ	0.495	0.353	0.293	0.251	0.23	0.195	0.168	0.141	0.123	0.105	0.093	0.08	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.59
6	$\bar{x}$	0.42	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.46	0.48	0.5	0.52
	σ	0.494	0.35	0.284	0.248	0.221	0.184	0.155	0.128	0.111	0.095	0.081	0.068	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
7	$\bar{x}$	0.42	0.44	0.47	0.47	0.48	0.5	0.52	0.55	0.57	0.61	0.63	0.66	0.68
	σ	0.494	0.353	0.291	0.256	0.226	0.191	0.161	0.131	0.115	0.091	0.079	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.67	0.69
8	$\bar{x}$	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.45	0.45	0.47	0.48
	σ	0.496	0.347	0.287	0.251	0.224	0.187	0.158	0.13	0.117	0.094	0.083	0.072	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.47	0.47
9	$\bar{x}$	0.43	0.44	0.45	0.45	0.46	0.48	0.49	0.51	0.53	0.55	0.57	0.6	0.62
	σ	0.495	0.361	0.295	0.261	0.238	0.206	0.175	0.144	0.131	0.108	0.094	0.078	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.57	0.6	0.63
10	$\bar{x}$	0.42	0.43	0.42	0.42	0.42	0.43	0.43	0.42	0.42	0.43	0.43	0.43	0.43
	σ	0.494	0.353	0.287	0.249	0.223	0.188	0.16	0.131	0.118	0.098	0.09	0.078	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.42
11	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.44	0.44
	σ	0.495	0.362	0.305	0.27	0.247	0.219	0.194	0.164	0.151	0.134	0.122	0.112	0.106
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.45	0.44
12	$\bar{x}$	0.43	0.42	0.43	0.42	0.43	0.43	0.43	0.42	0.42	0.42	0.43	0.43	0.43
	σ	0.496	0.349	0.286	0.248	0.218	0.187	0.155	0.127	0.111	0.091	0.078	0.062	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.42
13	$\bar{x}$	0.43	0.42	0.43	0.44	0.44	0.44	0.44	0.44	0.45	0.47	0.47	0.5	0.52
	σ	0.495	0.349	0.289	0.249	0.222	0.187	0.157	0.129	0.109	0.093	0.08	0.065	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
14	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.46	0.48	0.5	0.53
	σ	0.495	0.353	0.287	0.251	0.216	0.188	0.159	0.129	0.111	0.092	0.083	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
15	$\bar{x}$	0.43	0.45	0.47	0.47	0.48	0.5	0.52	0.55	0.57	0.6	0.63	0.66	0.68
	σ	0.495	0.354	0.292	0.251	0.228	0.194	0.161	0.132	0.112	0.091	0.078	0.061	0.051
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69

Table D.5: Local update for 40-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.44
	σ	0.497	0.357	0.287	0.247	0.223	0.187	0.157	0.13	0.112	0.093	0.078	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.44
2	$\bar{x}$	0.44	0.45	0.44	0.45	0.44	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.497	0.348	0.29	0.25	0.224	0.188	0.156	0.129	0.113	0.09	0.079	0.066	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.45	0.45	0.45	0.46	0.46	0.47	0.47	0.49	0.5	0.52	0.54	0.57	0.59
	σ	0.498	0.348	0.293	0.249	0.223	0.189	0.16	0.131	0.115	0.096	0.082	0.066	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.55	0.57	0.59
4	$\bar{x}$	0.45	0.45	0.45	0.44	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.351	0.289	0.249	0.226	0.188	0.158	0.128	0.113	0.089	0.077	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
5	$\bar{x}$	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.47	0.47	0.49	0.5	0.53	0.55
	σ	0.498	0.351	0.291	0.25	0.223	0.194	0.162	0.133	0.119	0.103	0.089	0.07	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.5	0.5	0.53	0.55
6	$\bar{x}$	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.46	0.45	0.46	0.46	0.48	0.48
	σ	0.497	0.347	0.287	0.247	0.224	0.184	0.157	0.128	0.11	0.091	0.079	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.47
7	$\bar{x}$	0.45	0.46	0.47	0.48	0.49	0.5	0.51	0.54	0.54	0.58	0.59	0.63	0.65
	σ	0.497	0.35	0.29	0.249	0.22	0.19	0.161	0.131	0.113	0.089	0.081	0.062	0.054
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65
8	$\bar{x}$	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.47
	σ	0.497	0.356	0.286	0.246	0.222	0.187	0.16	0.131	0.11	0.093	0.079	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.45	0.46
9	$\bar{x}$	0.45	0.45	0.46	0.47	0.47	0.47	0.49	0.51	0.52	0.54	0.55	0.57	0.59
	σ	0.498	0.36	0.294	0.258	0.234	0.2	0.167	0.141	0.124	0.104	0.086	0.073	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.57	0.59
10	$\bar{x}$	0.44	0.44	0.46	0.44	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.497	0.351	0.293	0.246	0.223	0.186	0.159	0.13	0.113	0.092	0.081	0.067	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
11	$\bar{x}$	0.45	0.44	0.45	0.44	0.45	0.45	0.44	0.45	0.44	0.45	0.45	0.45	0.45
	σ	0.498	0.359	0.304	0.27	0.24	0.207	0.183	0.156	0.139	0.119	0.109	0.097	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
12	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.44	0.45
	σ	0.497	0.347	0.287	0.25	0.226	0.188	0.157	0.13	0.109	0.089	0.077	0.062	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.43	0.45
13	$\bar{x}$	0.45	0.44	0.45	0.44	0.44	0.45	0.45	0.45	0.45	0.46	0.46	0.47	0.47
	σ	0.497	0.35	0.287	0.249	0.224	0.188	0.158	0.128	0.113	0.094	0.08	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.47	0.47
14	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.46	0.47	0.48	0.49
	σ	0.496	0.358	0.285	0.247	0.222	0.189	0.156	0.128	0.11	0.093	0.078	0.067	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.49
15	$\bar{x}$	0.43	0.45	0.46	0.48	0.49	0.5	0.52	0.53	0.55	0.57	0.6	0.62	0.65
	σ	0.495	0.352	0.29	0.25	0.225	0.193	0.157	0.13	0.11	0.089	0.079	0.064	0.052
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.62	0.65

Table D.6: Local update for 80-dimensional sphere function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.46	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.347	0.29	0.248	0.226	0.189	0.158	0.129	0.112	0.092	0.08	0.066	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
2	$\bar{x}$	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.347	0.284	0.25	0.222	0.189	0.154	0.128	0.112	0.09	0.081	0.065	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
3	$\bar{x}$	0.47	0.45	0.46	0.47	0.46	0.47	0.47	0.48	0.49	0.5	0.52	0.53	0.55
	σ	0.499	0.349	0.295	0.247	0.224	0.191	0.158	0.129	0.114	0.091	0.081	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
4	$\bar{x}$	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.356	0.287	0.247	0.222	0.185	0.159	0.127	0.109	0.091	0.079	0.062	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
5	$\bar{x}$	0.46	0.46	0.47	0.47	0.46	0.47	0.47	0.47	0.48	0.49	0.49	0.51	0.52
	σ	0.498	0.353	0.286	0.253	0.226	0.19	0.162	0.132	0.117	0.094	0.084	0.065	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.51
6	$\bar{x}$	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.47	0.48
	σ	0.498	0.352	0.284	0.249	0.223	0.188	0.155	0.129	0.112	0.09	0.079	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
7	$\bar{x}$	0.46	0.48	0.48	0.48	0.49	0.5	0.51	0.53	0.54	0.56	0.57	0.6	0.61
	σ	0.498	0.354	0.292	0.25	0.224	0.191	0.159	0.129	0.109	0.091	0.077	0.061	0.057
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.61
8	$\bar{x}$	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.47
	σ	0.498	0.354	0.285	0.25	0.222	0.188	0.16	0.128	0.11	0.091	0.079	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
9	$\bar{x}$	0.47	0.47	0.47	0.47	0.48	0.49	0.49	0.5	0.52	0.53	0.54	0.56	0.57
	σ	0.499	0.355	0.292	0.255	0.228	0.193	0.164	0.138	0.122	0.098	0.086	0.075	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.53	0.55	0.56
10	$\bar{x}$	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.498	0.352	0.285	0.248	0.224	0.189	0.157	0.128	0.113	0.091	0.08	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
11	$\bar{x}$	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.357	0.298	0.263	0.237	0.202	0.175	0.146	0.129	0.113	0.098	0.081	0.077
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
12	$\bar{x}$	0.45	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.353	0.287	0.25	0.221	0.19	0.158	0.127	0.109	0.094	0.081	0.066	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
13	$\bar{x}$	0.47	0.46	0.47	0.46	0.47	0.46	0.46	0.46	0.47	0.47	0.47	0.47	0.47
	σ	0.499	0.355	0.29	0.251	0.224	0.189	0.158	0.126	0.109	0.091	0.078	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
14	$\bar{x}$	0.46	0.46	0.46	0.46	0.47	0.47	0.46	0.47	0.46	0.47	0.47	0.47	0.48
	σ	0.498	0.351	0.29	0.249	0.223	0.186	0.158	0.127	0.111	0.091	0.078	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
15	$\bar{x}$	0.45	0.47	0.48	0.49	0.49	0.5	0.51	0.53	0.54	0.56	0.57	0.59	0.61
	σ	0.498	0.351	0.286	0.248	0.221	0.19	0.157	0.13	0.111	0.09	0.078	0.061	0.052
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.61

D.2 Schwefel's Function 1.2

Table D.7: Local update for 2-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.42	0.43
	σ	0.495	0.354	0.298	0.26	0.236	0.207	0.183	0.159	0.141	0.129	0.125	0.116	0.108
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.43	0.45	0.47	0.49	0.51	0.54	0.57	0.61	0.65	0.68	0.71	0.73	0.76
	σ	0.495	0.361	0.304	0.276	0.255	0.232	0.22	0.212	0.207	0.211	0.212	0.221	0.214
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.82	0.85
3	$\bar{x}$	0.42	0.47	0.51	0.53	0.55	0.59	0.62	0.66	0.68	0.71	0.74	0.76	0.77
	σ	0.494	0.364	0.308	0.275	0.265	0.243	0.233	0.221	0.222	0.223	0.216	0.215	0.22
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.86
4	$\bar{x}$	0.43	0.45	0.45	0.46	0.47	0.5	0.53	0.57	0.6	0.65	0.68	0.73	0.74
	σ	0.496	0.37	0.315	0.284	0.267	0.243	0.225	0.213	0.208	0.199	0.2	0.185	0.183
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.65	0.7	0.72	0.78	0.8
5	$\bar{x}$	0.42	0.46	0.46	0.49	0.5	0.54	0.57	0.62	0.64	0.69	0.71	0.75	0.74
	σ	0.494	0.379	0.321	0.298	0.28	0.257	0.238	0.217	0.22	0.203	0.206	0.194	0.211
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.8	0.81
6	$\bar{x}$	0.42	0.47	0.5	0.52	0.55	0.58	0.62	0.66	0.7	0.72	0.75	0.76	0.78
	σ	0.494	0.362	0.307	0.276	0.257	0.237	0.221	0.216	0.205	0.206	0.21	0.223	0.219
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.86
7	$\bar{x}$	0.43	0.48	0.54	0.57	0.59	0.63	0.66	0.7	0.73	0.76	0.77	0.8	0.8
	σ	0.496	0.359	0.312	0.279	0.263	0.241	0.231	0.213	0.205	0.202	0.212	0.201	0.212
	m	0	0.5	0.67	0.5	0.6	0.71	0.7	0.73	0.8	0.83	0.85	0.87	0.88
8	$\bar{x}$	0.43	0.44	0.46	0.48	0.5	0.53	0.56	0.62	0.63	0.69	0.71	0.74	0.76
	σ	0.496	0.373	0.322	0.294	0.277	0.257	0.231	0.212	0.214	0.197	0.202	0.202	0.188
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.82	0.82
9	$\bar{x}$	0.42	0.46	0.49	0.52	0.54	0.57	0.61	0.66	0.67	0.72	0.73	0.76	0.77
	σ	0.494	0.379	0.336	0.307	0.291	0.264	0.241	0.225	0.218	0.205	0.207	0.195	0.207
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.73	0.7	0.77	0.8	0.83	0.85
10	$\bar{x}$	0.43	0.46	0.47	0.49	0.51	0.55	0.59	0.63	0.65	0.7	0.71	0.74	0.76
	σ	0.496	0.383	0.34	0.317	0.3	0.275	0.26	0.24	0.227	0.21	0.212	0.201	0.195
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.8	0.82
11	$\bar{x}$	0.43	0.46	0.48	0.51	0.53	0.56	0.6	0.63	0.65	0.69	0.71	0.72	0.75
	σ	0.495	0.395	0.357	0.333	0.315	0.291	0.271	0.252	0.244	0.222	0.217	0.215	0.206
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.78	0.82
12	$\bar{x}$	0.42	0.46	0.49	0.52	0.53	0.57	0.61	0.66	0.69	0.73	0.76	0.78	0.78
	σ	0.494	0.364	0.309	0.282	0.263	0.24	0.223	0.212	0.208	0.206	0.192	0.199	0.212
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.73	0.75	0.8	0.82	0.85	0.86
13	$\bar{x}$	0.43	0.47	0.52	0.56	0.58	0.62	0.66	0.71	0.72	0.76	0.78	0.81	0.8
	σ	0.496	0.363	0.314	0.283	0.265	0.242	0.229	0.213	0.219	0.216	0.204	0.192	0.209
	m	0	0.5	0.67	0.5	0.6	0.71	0.7	0.73	0.8	0.83	0.85	0.88	0.89
14	$\bar{x}$	0.42	0.46	0.5	0.52	0.55	0.59	0.63	0.67	0.71	0.74	0.76	0.79	0.79
	σ	0.493	0.378	0.331	0.301	0.284	0.26	0.242	0.225	0.215	0.207	0.204	0.191	0.215
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.86	0.88
15	$\bar{x}$	0.43	0.48	0.52	0.55	0.57	0.62	0.66	0.7	0.73	0.76	0.77	0.81	0.81
	σ	0.495	0.383	0.338	0.31	0.293	0.265	0.236	0.223	0.216	0.212	0.215	0.193	0.192
	m	0	0.5	0.67	0.5	0.6	0.71	0.7	0.73	0.8	0.83	0.85	0.87	0.89

Table D.8: Local update for 5-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.49	0.5	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.49	0.49	0.49	0.49
	σ	0.5	0.353	0.289	0.249	0.223	0.194	0.161	0.128	0.116	0.091	0.08	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.48	0.49
2	$\bar{x}$	0.49	0.49	0.5	0.51	0.5	0.52	0.54	0.56	0.58	0.62	0.65	0.69	0.73
	σ	0.5	0.357	0.287	0.25	0.227	0.191	0.16	0.127	0.109	0.093	0.079	0.06	0.051
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.6	0.63	0.65	0.7	0.72
3	$\bar{x}$	0.48	0.51	0.53	0.55	0.57	0.6	0.64	0.67	0.7	0.75	0.78	0.81	0.84
	σ	0.5	0.355	0.291	0.249	0.224	0.187	0.154	0.124	0.104	0.078	0.067	0.051	0.043
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.78	0.82	0.84
4	$\bar{x}$	0.49	0.48	0.5	0.5	0.5	0.51	0.51	0.53	0.54	0.57	0.59	0.64	0.67
	σ	0.5	0.358	0.292	0.258	0.228	0.196	0.166	0.139	0.121	0.106	0.095	0.074	0.07
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.68
5	$\bar{x}$	0.48	0.5	0.51	0.52	0.54	0.55	0.58	0.61	0.64	0.69	0.72	0.75	0.77
	σ	0.5	0.365	0.305	0.267	0.243	0.211	0.183	0.153	0.134	0.108	0.094	0.078	0.069
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.75	0.78
6	$\bar{x}$	0.5	0.51	0.51	0.53	0.53	0.56	0.58	0.62	0.65	0.7	0.73	0.78	0.8
	σ	0.5	0.355	0.289	0.249	0.221	0.19	0.16	0.126	0.111	0.084	0.072	0.057	0.045
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.78	0.8
7	$\bar{x}$	0.5	0.53	0.57	0.59	0.62	0.65	0.68	0.72	0.75	0.78	0.81	0.84	0.85
	σ	0.5	0.353	0.28	0.244	0.214	0.178	0.145	0.116	0.094	0.075	0.06	0.046	0.039
	m	1	0.5	0.67	0.5	0.6	0.71	0.7	0.73	0.75	0.8	0.8	0.83	0.85
8	$\bar{x}$	0.49	0.49	0.5	0.51	0.51	0.53	0.54	0.57	0.6	0.64	0.67	0.71	0.75
	σ	0.5	0.355	0.293	0.257	0.235	0.202	0.172	0.147	0.132	0.11	0.095	0.08	0.065
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.6	0.6	0.63	0.68	0.72	0.75
9	$\bar{x}$	0.48	0.52	0.53	0.55	0.56	0.59	0.62	0.65	0.68	0.71	0.74	0.77	0.78
	σ	0.5	0.365	0.308	0.268	0.246	0.215	0.18	0.15	0.127	0.104	0.093	0.078	0.067
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.75	0.77	0.79
10	$\bar{x}$	0.49	0.5	0.5	0.5	0.51	0.51	0.53	0.56	0.57	0.6	0.62	0.65	0.68
	σ	0.5	0.361	0.3	0.266	0.247	0.22	0.194	0.173	0.165	0.148	0.142	0.134	0.125
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.6	0.6	0.65	0.67	0.7
11	$\bar{x}$	0.49	0.5	0.51	0.51	0.52	0.54	0.55	0.57	0.59	0.61	0.63	0.65	0.67
	σ	0.5	0.385	0.33	0.3	0.28	0.259	0.237	0.209	0.193	0.182	0.17	0.151	0.148
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.67	0.69
12	$\bar{x}$	0.48	0.49	0.5	0.51	0.51	0.52	0.54	0.56	0.58	0.62	0.65	0.7	0.74
	σ	0.5	0.352	0.291	0.252	0.226	0.191	0.161	0.129	0.116	0.095	0.085	0.072	0.062
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.6	0.63	0.65	0.7	0.74
13	$\bar{x}$	0.49	0.51	0.54	0.55	0.57	0.6	0.64	0.69	0.71	0.76	0.79	0.83	0.85
	σ	0.5	0.358	0.292	0.252	0.23	0.193	0.161	0.13	0.111	0.084	0.074	0.056	0.045
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
14	$\bar{x}$	0.48	0.5	0.52	0.53	0.54	0.56	0.59	0.63	0.67	0.71	0.75	0.79	0.82
	σ	0.5	0.354	0.292	0.26	0.227	0.197	0.168	0.138	0.117	0.091	0.076	0.06	0.047
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.68	0.73	0.75	0.8	0.82
15	$\bar{x}$	0.49	0.54	0.57	0.59	0.61	0.64	0.68	0.72	0.75	0.78	0.81	0.84	0.86
	σ	0.5	0.362	0.295	0.254	0.227	0.191	0.154	0.122	0.1	0.079	0.068	0.052	0.043
	m	0	0.5	0.67	0.5	0.6	0.71	0.7	0.73	0.75	0.8	0.82	0.85	0.86

Table D.9: Local update for 10-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.48	0.49	0.49	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.49	0.5	0.49
	σ	0.5	0.353	0.293	0.249	0.224	0.189	0.16	0.128	0.115	0.089	0.078	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.48	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.57	0.59
	σ	0.5	0.355	0.288	0.252	0.222	0.187	0.158	0.13	0.112	0.091	0.079	0.064	0.058
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.53	0.55	0.57	0.59
3	$\bar{x}$	0.49	0.51	0.52	0.54	0.54	0.56	0.58	0.62	0.64	0.68	0.71	0.75	0.78
	σ	0.5	0.351	0.288	0.249	0.223	0.189	0.16	0.124	0.107	0.084	0.071	0.057	0.044
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.72	0.75	0.78
4	$\bar{x}$	0.49	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.52	0.52	0.54	0.55
	σ	0.5	0.351	0.289	0.253	0.222	0.192	0.16	0.131	0.115	0.095	0.083	0.068	0.063
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.53	0.53	0.55
5	$\bar{x}$	0.49	0.5	0.5	0.52	0.52	0.53	0.55	0.57	0.59	0.63	0.65	0.69	0.71
	σ	0.5	0.357	0.295	0.256	0.238	0.201	0.169	0.145	0.13	0.107	0.094	0.079	0.066
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.6	0.6	0.63	0.65	0.68	0.71
6	$\bar{x}$	0.51	0.5	0.5	0.51	0.51	0.52	0.53	0.55	0.57	0.6	0.63	0.67	0.7
	σ	0.5	0.354	0.286	0.253	0.224	0.189	0.157	0.13	0.112	0.087	0.077	0.062	0.053
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.63	0.68	0.7
7	$\bar{x}$	0.5	0.52	0.54	0.58	0.58	0.6	0.64	0.67	0.69	0.73	0.75	0.78	0.8
	σ	0.5	0.358	0.29	0.244	0.222	0.183	0.153	0.119	0.099	0.079	0.067	0.052	0.045
	m	1	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.75	0.78	0.8
8	$\bar{x}$	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.54	0.55	0.58	0.61	0.65
	σ	0.5	0.355	0.287	0.254	0.226	0.192	0.164	0.136	0.12	0.102	0.089	0.075	0.069
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.57	0.62	0.65
9	$\bar{x}$	0.49	0.51	0.52	0.53	0.54	0.57	0.58	0.61	0.63	0.66	0.68	0.71	0.73
	σ	0.5	0.363	0.303	0.264	0.239	0.208	0.178	0.142	0.129	0.104	0.091	0.078	0.066
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.68	0.72	0.74
10	$\bar{x}$	0.51	0.5	0.5	0.49	0.5	0.5	0.51	0.51	0.51	0.52	0.53	0.53	0.54
	σ	0.5	0.352	0.296	0.256	0.232	0.202	0.172	0.146	0.13	0.115	0.11	0.102	0.104
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.53	0.5	0.53	0.53	0.53	0.54
11	$\bar{x}$	0.49	0.49	0.5	0.5	0.5	0.51	0.52	0.52	0.53	0.53	0.54	0.55	0.55
	σ	0.5	0.374	0.323	0.288	0.266	0.238	0.212	0.188	0.171	0.154	0.148	0.133	0.13
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.55	0.55
12	$\bar{x}$	0.49	0.48	0.51	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.55
	σ	0.5	0.358	0.287	0.252	0.224	0.189	0.159	0.129	0.11	0.092	0.079	0.067	0.057
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.53	0.53	0.53	0.55
13	$\bar{x}$	0.48	0.51	0.51	0.51	0.52	0.53	0.56	0.58	0.6	0.64	0.67	0.73	0.76
	σ	0.5	0.356	0.285	0.251	0.226	0.189	0.162	0.132	0.116	0.094	0.082	0.065	0.056
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.68	0.73	0.76
14	$\bar{x}$	0.5	0.5	0.51	0.51	0.51	0.52	0.53	0.55	0.57	0.61	0.63	0.68	0.71
	σ	0.5	0.353	0.286	0.251	0.225	0.192	0.162	0.129	0.113	0.093	0.081	0.064	0.054
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.63	0.68	0.71
15	$\bar{x}$	0.49	0.53	0.55	0.57	0.58	0.61	0.63	0.67	0.69	0.73	0.75	0.78	0.8
	σ	0.5	0.353	0.29	0.25	0.221	0.188	0.154	0.12	0.104	0.082	0.069	0.053	0.045
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.75	0.78	0.8

Table D.10: Local update for 20-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.353	0.29	0.249	0.225	0.192	0.159	0.129	0.109	0.088	0.079	0.063	0.056
	m	0	0.5	0.33	0.5	0.6	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.52
	σ	0.5	0.355	0.29	0.252	0.226	0.19	0.156	0.129	0.113	0.092	0.08	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.5	0.5	0.5	0.52	0.53
3	$\bar{x}$	0.5	0.5	0.52	0.52	0.52	0.53	0.55	0.57	0.59	0.62	0.65	0.68	0.71
	σ	0.5	0.353	0.29	0.25	0.222	0.187	0.157	0.128	0.111	0.091	0.075	0.064	0.052
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.68	0.71
4	$\bar{x}$	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.51
	σ	0.5	0.355	0.288	0.254	0.223	0.19	0.157	0.128	0.114	0.092	0.08	0.067	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.51
5	$\bar{x}$	0.51	0.51	0.51	0.5	0.51	0.52	0.53	0.54	0.55	0.58	0.6	0.63	0.65
	σ	0.5	0.357	0.297	0.252	0.228	0.195	0.166	0.137	0.122	0.101	0.092	0.076	0.065
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65
6	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.55	0.58	0.6
	σ	0.5	0.353	0.288	0.249	0.225	0.19	0.159	0.129	0.114	0.093	0.077	0.065	0.055
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
7	$\bar{x}$	0.5	0.52	0.54	0.54	0.56	0.58	0.6	0.62	0.64	0.67	0.7	0.73	0.75
	σ	0.5	0.35	0.288	0.25	0.219	0.185	0.152	0.124	0.108	0.083	0.071	0.057	0.048
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.73	0.75
8	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.52	0.53	0.54	0.56
	σ	0.5	0.358	0.29	0.25	0.224	0.192	0.161	0.131	0.116	0.094	0.083	0.069	0.062
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.53	0.54	0.56
9	$\bar{x}$	0.5	0.52	0.52	0.53	0.52	0.55	0.56	0.59	0.59	0.62	0.64	0.66	0.68
	σ	0.5	0.357	0.301	0.26	0.235	0.2	0.171	0.142	0.127	0.103	0.089	0.074	0.065
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.67	0.69
10	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.356	0.286	0.257	0.228	0.191	0.163	0.136	0.117	0.099	0.088	0.079	0.073
	m	1	0.5	0.67	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
11	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.51	0.52	0.51
	σ	0.5	0.369	0.308	0.274	0.251	0.22	0.193	0.169	0.154	0.137	0.123	0.11	0.107
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.52	0.51
12	$\bar{x}$	0.51	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.351	0.286	0.25	0.224	0.188	0.158	0.13	0.111	0.089	0.081	0.065	0.059
	m	1	0.5	0.67	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.5
13	$\bar{x}$	0.49	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.53	0.55	0.56	0.59	0.61
	σ	0.5	0.352	0.288	0.254	0.225	0.19	0.158	0.131	0.113	0.093	0.08	0.068	0.059
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.57	0.58	0.61
14	$\bar{x}$	0.49	0.49	0.5	0.5	0.5	0.51	0.52	0.52	0.52	0.54	0.55	0.58	0.6
	σ	0.5	0.349	0.293	0.252	0.227	0.188	0.16	0.13	0.111	0.091	0.078	0.065	0.055
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
15	$\bar{x}$	0.5	0.51	0.54	0.55	0.56	0.57	0.6	0.62	0.65	0.67	0.7	0.73	0.75
	σ	0.5	0.351	0.29	0.252	0.224	0.185	0.154	0.124	0.107	0.083	0.072	0.057	0.047
	m	1	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.73	0.75

Table D.11: Local update for 40-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.352	0.293	0.252	0.222	0.19	0.159	0.13	0.111	0.091	0.079	0.065	0.056
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.53	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.5	0.51
	σ	0.5	0.354	0.293	0.248	0.223	0.191	0.158	0.128	0.113	0.092	0.079	0.064	0.054
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.5	0.5	0.5	0.5	0.51
3	$\bar{x}$	0.49	0.51	0.5	0.51	0.51	0.52	0.53	0.54	0.55	0.57	0.59	0.63	0.65
	σ	0.5	0.355	0.287	0.249	0.223	0.19	0.159	0.13	0.11	0.09	0.081	0.063	0.052
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65
4	$\bar{x}$	0.5	0.5	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.353	0.29	0.254	0.227	0.189	0.157	0.131	0.113	0.092	0.08	0.066	0.058
	m	1	0.5	0.67	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
5	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.56	0.58	0.6
	σ	0.5	0.352	0.292	0.254	0.225	0.195	0.159	0.135	0.12	0.098	0.088	0.071	0.066
	m	1	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
6	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.53	0.54
	σ	0.5	0.355	0.287	0.254	0.226	0.192	0.159	0.13	0.11	0.094	0.079	0.063	0.054
	m	1	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.53	0.53	0.54
7	$\bar{x}$	0.5	0.51	0.52	0.53	0.54	0.55	0.57	0.59	0.6	0.63	0.65	0.68	0.7
	σ	0.5	0.356	0.293	0.253	0.222	0.192	0.154	0.127	0.107	0.086	0.073	0.056	0.049
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.52	0.51
8	$\bar{x}$	0.5	0.49	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.52
	σ	0.5	0.359	0.29	0.254	0.222	0.188	0.157	0.128	0.114	0.093	0.082	0.068	0.06
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.62	0.64
9	$\bar{x}$	0.51	0.51	0.51	0.52	0.52	0.53	0.54	0.56	0.57	0.59	0.6	0.62	0.64
	σ	0.5	0.358	0.295	0.263	0.231	0.196	0.166	0.139	0.117	0.103	0.088	0.07	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
10	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.354	0.29	0.25	0.225	0.195	0.161	0.133	0.116	0.096	0.082	0.07	0.06
	m	1	0.5	0.67	0.5	0.6	0.5	0.5	0.53	0.5	0.5	0.5	0.5	0.5
11	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.365	0.302	0.27	0.242	0.211	0.184	0.151	0.14	0.119	0.107	0.095	0.09
	m	0	0.5	0.33	0.5	0.6	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
12	$\bar{x}$	0.5	0.5	0.49	0.5	0.5	0.5	0.51	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.355	0.293	0.25	0.225	0.188	0.156	0.129	0.112	0.09	0.08	0.066	0.056
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.53	0.5	0.5	0.53	0.52	0.53
13	$\bar{x}$	0.51	0.5	0.51	0.5	0.5	0.5	0.51	0.51	0.51	0.51	0.51	0.52	0.53
	σ	0.5	0.353	0.289	0.25	0.223	0.191	0.157	0.128	0.112	0.091	0.078	0.065	0.056
	m	1	0.5	0.67	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.53	0.53	0.54
14	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.52	0.53	0.54
	σ	0.5	0.358	0.289	0.247	0.222	0.187	0.157	0.128	0.114	0.094	0.078	0.064	0.057
	m													
15	$\bar{x}$	0.5	0.52	0.53	0.54	0.54	0.55	0.57	0.59	0.6	0.63	0.65	0.68	0.7
	σ	0.5	0.356	0.284	0.25	0.22	0.188	0.155	0.125	0.106	0.086	0.075	0.057	0.05
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.68	0.7

Table D.12: Local update for 80-dimensional Schwefel function 1.2

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.51	0.49	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.356	0.286	0.252	0.223	0.188	0.16	0.13	0.107	0.091	0.079	0.066	0.057
	m	1	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.353	0.289	0.251	0.223	0.192	0.157	0.127	0.112	0.092	0.08	0.065	0.056
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
3	$\bar{x}$	0.5	0.5	0.51	0.5	0.5	0.51	0.52	0.52	0.53	0.54	0.55	0.58	0.59
	σ	0.5	0.353	0.287	0.248	0.223	0.188	0.157	0.128	0.112	0.091	0.078	0.064	0.056
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
4	$\bar{x}$	0.49	0.49	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.351	0.281	0.253	0.22	0.191	0.157	0.128	0.112	0.091	0.08	0.066	0.054
	m	0	0.5	0.67	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
5	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.52	0.53	0.54	0.55
	σ	0.5	0.354	0.283	0.251	0.228	0.191	0.16	0.13	0.114	0.097	0.085	0.072	0.061
	m	1	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.5	0.53	0.53	0.55	0.55
6	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.51
	σ	0.5	0.354	0.287	0.249	0.224	0.189	0.159	0.131	0.112	0.092	0.08	0.065	0.058
	m	1	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.52	0.51
7	$\bar{x}$	0.5	0.51	0.52	0.52	0.53	0.54	0.55	0.56	0.58	0.59	0.61	0.63	0.65
	σ	0.5	0.348	0.291	0.248	0.223	0.19	0.158	0.127	0.109	0.088	0.077	0.061	0.051
	m	1	0.5	0.67	0.5	0.6	0.57	0.6	0.53	0.6	0.6	0.6	0.63	0.65
8	$\bar{x}$	0.49	0.51	0.5	0.5	0.51	0.5	0.5	0.5	0.51	0.5	0.5	0.51	0.5
	σ	0.5	0.353	0.293	0.25	0.225	0.187	0.158	0.133	0.113	0.091	0.079	0.066	0.056
	m	0	0.5	0.67	0.5	0.6	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.5
9	$\bar{x}$	0.49	0.5	0.51	0.52	0.52	0.52	0.53	0.54	0.55	0.56	0.57	0.59	0.6
	σ	0.5	0.357	0.297	0.258	0.23	0.199	0.165	0.135	0.12	0.099	0.085	0.07	0.061
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.57	0.58	0.6
10	$\bar{x}$	0.49	0.51	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.352	0.288	0.25	0.223	0.19	0.156	0.128	0.11	0.091	0.082	0.066	0.059
	m	0	0.5	0.67	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.5
11	$\bar{x}$	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.364	0.298	0.267	0.241	0.204	0.176	0.147	0.125	0.112	0.1	0.088	0.076
	m	1	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.5
12	$\bar{x}$	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.357	0.286	0.252	0.223	0.189	0.157	0.131	0.11	0.092	0.08	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
13	$\bar{x}$	0.51	0.5	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51
	σ	0.5	0.353	0.291	0.25	0.222	0.189	0.158	0.129	0.111	0.091	0.078	0.062	0.055
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.51
14	$\bar{x}$	0.49	0.5	0.5	0.5	0.51	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51
	σ	0.5	0.355	0.285	0.248	0.224	0.192	0.159	0.129	0.113	0.092	0.081	0.064	0.053
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.47	0.5	0.5	0.5	0.51	0.51
15	$\bar{x}$	0.5	0.51	0.52	0.52	0.53	0.54	0.55	0.56	0.58	0.59	0.61	0.63	0.65
	σ	0.5	0.35	0.29	0.247	0.223	0.188	0.157	0.128	0.11	0.091	0.077	0.06	0.053
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.53	0.6	0.6	0.63	0.63	0.65

D.3 Cigar Function

Table D.13: Local update for 2-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.41	0.39	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.41	0.41
	σ	0.492	0.357	0.307	0.277	0.25	0.227	0.204	0.188	0.175	0.163	0.154	0.15	0.141
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.4	0.43	0.44	0.45	0.46	0.48	0.51	0.55	0.57	0.59	0.61	0.64	0.65
	σ	0.489	0.365	0.317	0.29	0.278	0.261	0.257	0.26	0.265	0.273	0.282	0.285	0.296
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.75	0.78	0.8
3	$\bar{x}$	0.4	0.44	0.46	0.48	0.5	0.52	0.55	0.58	0.6	0.63	0.63	0.65	0.68
	σ	0.491	0.368	0.322	0.3	0.29	0.276	0.272	0.275	0.28	0.288	0.293	0.308	0.302
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.78	0.82	0.84
4	$\bar{x}$	0.4	0.4	0.41	0.43	0.43	0.45	0.48	0.51	0.54	0.57	0.58	0.6	0.64
	σ	0.49	0.367	0.318	0.296	0.278	0.261	0.249	0.246	0.246	0.264	0.267	0.271	0.273
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.67	0.68	0.7	0.75
5	$\bar{x}$	0.4	0.42	0.44	0.44	0.47	0.48	0.51	0.54	0.57	0.59	0.62	0.62	0.64
	σ	0.489	0.372	0.329	0.303	0.288	0.277	0.268	0.265	0.267	0.271	0.273	0.282	0.285
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.67	0.72	0.73	0.76
6	$\bar{x}$	0.41	0.43	0.46	0.48	0.5	0.52	0.55	0.59	0.61	0.64	0.67	0.69	0.69
	σ	0.491	0.363	0.321	0.295	0.282	0.272	0.268	0.27	0.276	0.282	0.278	0.286	0.294
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
7	$\bar{x}$	0.39	0.44	0.48	0.51	0.53	0.56	0.58	0.62	0.62	0.66	0.68	0.7	0.69
	σ	0.488	0.373	0.327	0.308	0.292	0.284	0.283	0.285	0.296	0.292	0.293	0.292	0.307
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.86
8	$\bar{x}$	0.4	0.42	0.43	0.44	0.46	0.48	0.51	0.55	0.56	0.6	0.63	0.65	0.67
	σ	0.49	0.376	0.33	0.305	0.291	0.278	0.264	0.26	0.261	0.269	0.266	0.282	0.274
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.6	0.7	0.72	0.78	0.79
9	$\bar{x}$	0.41	0.42	0.45	0.47	0.48	0.51	0.54	0.58	0.59	0.62	0.62	0.67	0.65
	σ	0.491	0.38	0.345	0.314	0.303	0.287	0.275	0.273	0.27	0.276	0.287	0.276	0.292
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.65	0.7	0.72	0.78	0.78
10	$\bar{x}$	0.41	0.43	0.44	0.46	0.46	0.49	0.52	0.57	0.6	0.62	0.63	0.64	0.67
	σ	0.491	0.388	0.344	0.32	0.308	0.298	0.283	0.276	0.267	0.274	0.281	0.283	0.285
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.73	0.79
11	$\bar{x}$	0.41	0.43	0.45	0.47	0.48	0.5	0.54	0.57	0.61	0.62	0.62	0.65	0.67
	σ	0.491	0.391	0.358	0.336	0.322	0.307	0.294	0.287	0.275	0.281	0.282	0.283	0.285
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.75	0.79
12	$\bar{x}$	0.4	0.43	0.45	0.47	0.48	0.52	0.56	0.6	0.62	0.64	0.66	0.69	0.69
	σ	0.491	0.369	0.324	0.298	0.289	0.269	0.269	0.267	0.273	0.282	0.288	0.291	0.301
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
13	$\bar{x}$	0.41	0.44	0.48	0.5	0.53	0.55	0.59	0.63	0.64	0.67	0.68	0.7	0.71
	σ	0.491	0.371	0.329	0.312	0.296	0.288	0.283	0.281	0.288	0.293	0.3	0.297	0.304
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.88
14	$\bar{x}$	0.4	0.44	0.46	0.47	0.5	0.53	0.56	0.61	0.61	0.65	0.66	0.68	0.7
	σ	0.489	0.383	0.343	0.312	0.303	0.288	0.282	0.275	0.285	0.285	0.298	0.298	0.294
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
15	$\bar{x}$	0.41	0.45	0.47	0.5	0.52	0.55	0.58	0.61	0.64	0.66	0.68	0.69	0.7
	σ	0.491	0.387	0.351	0.325	0.31	0.296	0.286	0.286	0.288	0.293	0.294	0.298	0.302
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.73	0.75	0.77	0.8	0.83	0.85

Table D.14: Local update for 5-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.38	0.39	0.39	0.38	0.38	0.39	0.39	0.39	0.39	0.39	0.39	0.38	0.38
	σ	0.485	0.344	0.281	0.243	0.222	0.186	0.156	0.131	0.116	0.096	0.085	0.071	0.065
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.38	0.39
2	$\bar{x}$	0.39	0.39	0.39	0.39	0.4	0.41	0.42	0.44	0.46	0.5	0.52	0.56	0.59
	σ	0.487	0.346	0.284	0.251	0.225	0.191	0.165	0.139	0.126	0.108	0.101	0.093	0.093
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
3	$\bar{x}$	0.39	0.41	0.42	0.45	0.46	0.48	0.51	0.55	0.58	0.62	0.65	0.68	0.7
	σ	0.487	0.349	0.29	0.258	0.234	0.204	0.176	0.149	0.136	0.112	0.101	0.096	0.081
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.65	0.68	0.71
4	$\bar{x}$	0.38	0.38	0.39	0.39	0.39	0.39	0.41	0.42	0.43	0.45	0.47	0.52	0.55
	σ	0.485	0.337	0.283	0.251	0.225	0.196	0.172	0.144	0.131	0.117	0.107	0.096	0.092
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
5	$\bar{x}$	0.37	0.4	0.41	0.42	0.42	0.44	0.47	0.51	0.53	0.58	0.61	0.64	0.66
	σ	0.484	0.357	0.3	0.262	0.245	0.214	0.19	0.165	0.146	0.124	0.111	0.094	0.087
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.61	0.65	0.66
6	$\bar{x}$	0.39	0.4	0.41	0.41	0.42	0.45	0.47	0.5	0.53	0.58	0.6	0.64	0.67
	σ	0.488	0.349	0.289	0.251	0.229	0.197	0.169	0.147	0.129	0.11	0.104	0.089	0.085
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.68
7	$\bar{x}$	0.38	0.43	0.46	0.47	0.49	0.52	0.56	0.6	0.63	0.65	0.68	0.7	0.72
	σ	0.485	0.356	0.298	0.261	0.236	0.204	0.175	0.147	0.131	0.11	0.098	0.094	0.09
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.7	0.72	0.74
8	$\bar{x}$	0.39	0.39	0.4	0.4	0.41	0.42	0.43	0.46	0.48	0.52	0.55	0.6	0.63
	σ	0.487	0.352	0.289	0.255	0.232	0.205	0.179	0.155	0.145	0.125	0.112	0.098	0.087
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.6	0.64
9	$\bar{x}$	0.39	0.41	0.42	0.44	0.45	0.48	0.51	0.55	0.57	0.61	0.63	0.66	0.68
	σ	0.488	0.366	0.314	0.277	0.256	0.227	0.197	0.168	0.149	0.125	0.111	0.098	0.085
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.6	0.63	0.67	0.69
10	$\bar{x}$	0.38	0.4	0.4	0.4	0.4	0.41	0.42	0.44	0.45	0.48	0.5	0.53	0.55
	σ	0.485	0.359	0.297	0.266	0.24	0.211	0.193	0.173	0.165	0.146	0.141	0.126	0.124
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.5	0.53	0.56
11	$\bar{x}$	0.39	0.4	0.41	0.41	0.42	0.43	0.44	0.46	0.47	0.49	0.5	0.53	0.54
	σ	0.487	0.369	0.323	0.293	0.277	0.251	0.229	0.205	0.201	0.181	0.173	0.15	0.142
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.53	0.55	0.56
12	$\bar{x}$	0.38	0.39	0.39	0.4	0.4	0.41	0.42	0.44	0.45	0.49	0.51	0.55	0.59
	σ	0.486	0.349	0.285	0.247	0.225	0.19	0.163	0.134	0.122	0.102	0.093	0.088	0.076
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.59
13	$\bar{x}$	0.39	0.4	0.42	0.43	0.45	0.47	0.51	0.55	0.58	0.62	0.65	0.68	0.71
	σ	0.489	0.352	0.286	0.253	0.23	0.2	0.175	0.146	0.131	0.109	0.099	0.089	0.08
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.65	0.68	0.71
14	$\bar{x}$	0.38	0.39	0.41	0.42	0.42	0.44	0.47	0.5	0.54	0.58	0.61	0.66	0.67
	σ	0.487	0.351	0.292	0.257	0.234	0.203	0.179	0.155	0.138	0.114	0.108	0.092	0.088
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.67	0.69
15	$\bar{x}$	0.39	0.42	0.45	0.47	0.49	0.52	0.56	0.6	0.63	0.65	0.68	0.71	0.72
	σ	0.488	0.363	0.304	0.272	0.245	0.217	0.184	0.15	0.134	0.118	0.1	0.094	0.082
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.68	0.72	0.74

Table D.15: Local update for 10-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.4	0.41	0.4	0.4	0.41	0.41
	σ	0.491	0.349	0.282	0.248	0.22	0.185	0.156	0.126	0.11	0.093	0.078	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
2	$\bar{x}$	0.41	0.4	0.4	0.41	0.41	0.41	0.42	0.42	0.43	0.44	0.45	0.47	0.49
	σ	0.491	0.349	0.286	0.248	0.218	0.186	0.157	0.131	0.111	0.092	0.082	0.07	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.47	0.49
3	$\bar{x}$	0.41	0.41	0.42	0.42	0.44	0.46	0.49	0.52	0.54	0.59	0.61	0.65	0.68
	σ	0.492	0.346	0.29	0.25	0.224	0.194	0.162	0.137	0.118	0.099	0.085	0.069	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.65	0.68
4	$\bar{x}$	0.4	0.4	0.41	0.42	0.41	0.41	0.41	0.41	0.42	0.43	0.43	0.44	0.45
	σ	0.49	0.346	0.282	0.248	0.219	0.186	0.157	0.13	0.113	0.095	0.085	0.075	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
5	$\bar{x}$	0.4	0.41	0.41	0.43	0.43	0.44	0.46	0.48	0.5	0.54	0.56	0.6	0.63
	σ	0.49	0.349	0.291	0.259	0.236	0.203	0.177	0.151	0.133	0.112	0.098	0.085	0.071
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.57	0.6	0.63
6	$\bar{x}$	0.41	0.42	0.41	0.41	0.42	0.43	0.44	0.45	0.48	0.5	0.53	0.56	0.6
	σ	0.492	0.347	0.287	0.248	0.224	0.187	0.158	0.132	0.12	0.093	0.085	0.072	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.53	0.57	0.6
7	$\bar{x}$	0.42	0.44	0.45	0.47	0.49	0.51	0.54	0.57	0.6	0.63	0.66	0.69	0.71
	σ	0.493	0.352	0.29	0.256	0.229	0.194	0.164	0.133	0.117	0.091	0.079	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.7	0.71
8	$\bar{x}$	0.41	0.4	0.41	0.41	0.41	0.42	0.43	0.43	0.44	0.46	0.48	0.51	0.55
	σ	0.491	0.351	0.284	0.246	0.226	0.193	0.165	0.136	0.12	0.103	0.091	0.08	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
9	$\bar{x}$	0.41	0.43	0.44	0.44	0.46	0.47	0.49	0.52	0.54	0.58	0.6	0.63	0.65
	σ	0.492	0.359	0.307	0.269	0.248	0.216	0.186	0.153	0.134	0.111	0.098	0.079	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.65
10	$\bar{x}$	0.41	0.42	0.4	0.41	0.41	0.41	0.41	0.41	0.42	0.43	0.43	0.43	0.45
	σ	0.492	0.354	0.285	0.255	0.225	0.196	0.169	0.143	0.132	0.119	0.108	0.102	0.097
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
11	$\bar{x}$	0.42	0.4	0.4	0.41	0.41	0.41	0.42	0.42	0.43	0.44	0.44	0.45	0.46
	σ	0.493	0.365	0.311	0.283	0.258	0.233	0.208	0.186	0.171	0.15	0.147	0.146	0.127
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.45	0.45
12	$\bar{x}$	0.42	0.41	0.41	0.4	0.4	0.41	0.41	0.42	0.42	0.42	0.42	0.44	0.45
	σ	0.494	0.347	0.285	0.243	0.223	0.186	0.156	0.13	0.112	0.091	0.077	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
13	$\bar{x}$	0.41	0.41	0.42	0.42	0.43	0.44	0.45	0.47	0.49	0.53	0.56	0.6	0.64
	σ	0.491	0.35	0.288	0.249	0.222	0.192	0.159	0.132	0.116	0.099	0.086	0.069	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.6	0.64
14	$\bar{x}$	0.42	0.41	0.41	0.42	0.42	0.43	0.44	0.46	0.47	0.5	0.53	0.57	0.6
	σ	0.493	0.346	0.285	0.247	0.223	0.187	0.165	0.133	0.121	0.1	0.087	0.075	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.58	0.6
15	$\bar{x}$	0.42	0.43	0.45	0.47	0.48	0.51	0.54	0.57	0.6	0.63	0.66	0.69	0.71
	σ	0.494	0.356	0.292	0.255	0.235	0.198	0.164	0.135	0.117	0.092	0.081	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.7	0.71

Table D.16: Local update for 20-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.495	0.351	0.288	0.25	0.221	0.185	0.158	0.124	0.108	0.091	0.078	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.43	0.43	0.42	0.43	0.42
2	$\bar{x}$	0.43	0.42	0.43	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45
	σ	0.495	0.35	0.282	0.246	0.221	0.188	0.158	0.127	0.111	0.092	0.08	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.43	0.43	0.44	0.45	0.45	0.46	0.47	0.49	0.51	0.55	0.58	0.61	0.64
	σ	0.496	0.348	0.286	0.254	0.222	0.189	0.158	0.133	0.116	0.096	0.084	0.068	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.57	0.57	0.62	0.64
4	$\bar{x}$	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44
	σ	0.493	0.353	0.288	0.247	0.22	0.187	0.155	0.129	0.111	0.091	0.082	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.45
5	$\bar{x}$	0.43	0.44	0.43	0.43	0.44	0.45	0.46	0.47	0.48	0.5	0.52	0.56	0.59
	σ	0.495	0.357	0.287	0.254	0.227	0.194	0.167	0.137	0.124	0.107	0.093	0.078	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.57	0.59
6	$\bar{x}$	0.42	0.43	0.43	0.44	0.43	0.44	0.44	0.45	0.45	0.47	0.48	0.5	0.53
	σ	0.494	0.353	0.289	0.245	0.223	0.188	0.159	0.13	0.112	0.092	0.079	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
7	$\bar{x}$	0.44	0.45	0.45	0.48	0.48	0.5	0.52	0.55	0.57	0.6	0.63	0.66	0.69
	σ	0.496	0.352	0.291	0.256	0.227	0.192	0.158	0.13	0.116	0.092	0.079	0.061	0.052
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.65	0.69
8	$\bar{x}$	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.45	0.47	0.49
	σ	0.496	0.35	0.288	0.249	0.22	0.191	0.159	0.133	0.117	0.095	0.085	0.07	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.47	0.49
9	$\bar{x}$	0.43	0.43	0.45	0.45	0.47	0.47	0.49	0.51	0.53	0.56	0.57	0.6	0.62
	σ	0.496	0.357	0.298	0.262	0.236	0.206	0.175	0.144	0.128	0.109	0.093	0.077	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.63
10	$\bar{x}$	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.494	0.35	0.289	0.25	0.224	0.192	0.161	0.134	0.117	0.101	0.088	0.078	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.44
11	$\bar{x}$	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.43	0.44	0.44
	σ	0.496	0.36	0.306	0.271	0.253	0.218	0.19	0.166	0.156	0.133	0.118	0.111	0.107
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
12	$\bar{x}$	0.44	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44
	σ	0.496	0.346	0.286	0.249	0.22	0.185	0.157	0.127	0.109	0.092	0.076	0.06	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.44
13	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.45	0.46	0.47	0.48	0.5	0.52
	σ	0.495	0.351	0.283	0.246	0.219	0.188	0.157	0.131	0.112	0.09	0.08	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
14	$\bar{x}$	0.42	0.42	0.44	0.43	0.43	0.44	0.44	0.45	0.45	0.47	0.48	0.5	0.53
	σ	0.494	0.348	0.293	0.249	0.222	0.189	0.158	0.129	0.114	0.092	0.083	0.068	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
15	$\bar{x}$	0.42	0.45	0.46	0.48	0.49	0.5	0.53	0.55	0.57	0.6	0.63	0.66	0.68
	σ	0.493	0.349	0.297	0.25	0.23	0.195	0.159	0.131	0.113	0.09	0.077	0.061	0.052
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.57	0.6	0.63	0.67	0.69

Table D.17: Local update for 40-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.45	0.44	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.348	0.285	0.251	0.222	0.187	0.157	0.128	0.112	0.092	0.078	0.064	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.354	0.288	0.249	0.223	0.189	0.157	0.129	0.112	0.09	0.077	0.062	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.45	0.44	0.44	0.46	0.47	0.46	0.47	0.49	0.5	0.52	0.54	0.57	0.59
	σ	0.498	0.351	0.283	0.25	0.223	0.187	0.161	0.13	0.111	0.096	0.083	0.062	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.53	0.57	0.6
4	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.496	0.351	0.284	0.25	0.22	0.188	0.154	0.129	0.108	0.092	0.078	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
5	$\bar{x}$	0.45	0.44	0.44	0.45	0.46	0.46	0.46	0.47	0.48	0.49	0.5	0.52	0.55
	σ	0.497	0.348	0.289	0.25	0.227	0.192	0.161	0.135	0.12	0.099	0.089	0.075	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.53	0.55
6	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.46	0.47	0.48
	σ	0.496	0.352	0.29	0.246	0.218	0.185	0.158	0.128	0.113	0.094	0.079	0.062	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.49
7	$\bar{x}$	0.46	0.46	0.47	0.49	0.49	0.5	0.52	0.53	0.55	0.58	0.6	0.63	0.65
	σ	0.498	0.357	0.292	0.251	0.223	0.19	0.162	0.131	0.114	0.089	0.079	0.061	0.055
	m	0	0.5	0.33	0.5	0.4	0.5	0.5	0.53	0.55	0.57	0.6	0.63	0.65
8	$\bar{x}$	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.46	0.46	0.47
	σ	0.498	0.356	0.287	0.25	0.222	0.189	0.159	0.13	0.113	0.092	0.081	0.066	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.47	0.46
9	$\bar{x}$	0.46	0.45	0.47	0.47	0.47	0.48	0.49	0.51	0.52	0.54	0.55	0.58	0.59
	σ	0.498	0.358	0.297	0.261	0.234	0.203	0.169	0.14	0.122	0.1	0.088	0.075	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.55	0.58	0.59
10	$\bar{x}$	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.347	0.29	0.248	0.227	0.192	0.159	0.133	0.117	0.094	0.082	0.068	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
11	$\bar{x}$	0.44	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45
	σ	0.497	0.357	0.301	0.266	0.243	0.211	0.18	0.155	0.14	0.121	0.109	0.099	0.089
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
12	$\bar{x}$	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.44	0.44	0.45	0.45
	σ	0.498	0.351	0.285	0.248	0.22	0.189	0.157	0.128	0.11	0.092	0.078	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
13	$\bar{x}$	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.46	0.47	0.47
	σ	0.497	0.35	0.287	0.251	0.219	0.187	0.157	0.129	0.111	0.09	0.079	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.47	0.47
14	$\bar{x}$	0.46	0.45	0.46	0.45	0.45	0.45	0.45	0.45	0.45	0.46	0.47	0.48	0.49
	σ	0.498	0.352	0.287	0.251	0.218	0.188	0.156	0.128	0.112	0.092	0.08	0.066	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.49
15	$\bar{x}$	0.44	0.46	0.47	0.48	0.5	0.5	0.51	0.54	0.55	0.58	0.6	0.63	0.65
	σ	0.496	0.354	0.289	0.252	0.226	0.191	0.16	0.131	0.11	0.089	0.075	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65

Table D.18: Local update for 80-dimensional cigar function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.46	0.47	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.46
	σ	0.498	0.35	0.284	0.249	0.222	0.185	0.157	0.13	0.112	0.09	0.08	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
2	$\bar{x}$	0.45	0.47	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.357	0.289	0.246	0.222	0.186	0.156	0.13	0.112	0.089	0.08	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
3	$\bar{x}$	0.47	0.47	0.46	0.47	0.47	0.47	0.48	0.48	0.49	0.5	0.52	0.53	0.55
	σ	0.499	0.351	0.286	0.248	0.226	0.189	0.162	0.13	0.116	0.092	0.078	0.067	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
4	$\bar{x}$	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.352	0.287	0.249	0.226	0.189	0.155	0.129	0.109	0.092	0.077	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
5	$\bar{x}$	0.45	0.47	0.46	0.47	0.47	0.46	0.47	0.47	0.47	0.49	0.49	0.5	0.52
	σ	0.498	0.354	0.294	0.255	0.227	0.19	0.161	0.131	0.114	0.093	0.086	0.07	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.5	0.5	0.5	0.53
6	$\bar{x}$	0.45	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.47	0.47	0.48
	σ	0.498	0.35	0.286	0.252	0.221	0.188	0.159	0.128	0.109	0.091	0.077	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
7	$\bar{x}$	0.46	0.47	0.48	0.49	0.5	0.5	0.51	0.53	0.54	0.56	0.57	0.6	0.61
	σ	0.499	0.354	0.294	0.255	0.222	0.188	0.157	0.128	0.112	0.093	0.076	0.066	0.056
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.61
8	$\bar{x}$	0.47	0.47	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.47	0.47	0.47
	σ	0.499	0.354	0.285	0.25	0.224	0.189	0.156	0.129	0.11	0.091	0.08	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
9	$\bar{x}$	0.46	0.47	0.48	0.48	0.48	0.49	0.49	0.51	0.51	0.53	0.54	0.56	0.57
	σ	0.499	0.358	0.295	0.254	0.229	0.196	0.165	0.138	0.12	0.1	0.087	0.072	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.53	0.55	0.57
10	$\bar{x}$	0.47	0.47	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.351	0.287	0.25	0.22	0.189	0.16	0.128	0.112	0.092	0.081	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
11	$\bar{x}$	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.47
	σ	0.499	0.358	0.294	0.259	0.234	0.205	0.175	0.147	0.132	0.111	0.102	0.088	0.076
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
12	$\bar{x}$	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.47
	σ	0.499	0.353	0.285	0.248	0.219	0.186	0.158	0.129	0.11	0.09	0.079	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
13	$\bar{x}$	0.48	0.45	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.47	0.47	0.47	0.47
	σ	0.5	0.352	0.289	0.248	0.223	0.185	0.155	0.129	0.114	0.092	0.076	0.067	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
14	$\bar{x}$	0.47	0.46	0.46	0.47	0.45	0.46	0.46	0.46	0.47	0.47	0.47	0.47	0.48
	σ	0.499	0.35	0.29	0.248	0.22	0.19	0.159	0.129	0.111	0.089	0.078	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
15	$\bar{x}$	0.47	0.47	0.48	0.49	0.49	0.5	0.51	0.52	0.54	0.56	0.57	0.59	0.62
	σ	0.499	0.355	0.289	0.251	0.225	0.188	0.156	0.13	0.113	0.089	0.077	0.063	0.054
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.61

D.4 Tablet Function

Table D.19: Local update for 2-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.41	0.4	0.4	0.4	0.4	0.39	0.4	0.4	0.41	0.4	0.41	0.4	0.4
	σ	0.491	0.359	0.304	0.273	0.252	0.226	0.204	0.185	0.17	0.163	0.157	0.157	0.149
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.43	0.45
2	$\bar{x}$	0.4	0.42	0.43	0.45	0.46	0.48	0.51	0.55	0.57	0.6	0.62	0.65	0.65
	σ	0.489	0.362	0.313	0.29	0.277	0.26	0.257	0.256	0.264	0.272	0.28	0.289	0.303
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.75	0.78	0.81
3	$\bar{x}$	0.39	0.43	0.45	0.48	0.49	0.52	0.55	0.59	0.6	0.62	0.65	0.65	0.67
	σ	0.487	0.367	0.324	0.303	0.291	0.275	0.273	0.275	0.28	0.29	0.292	0.301	0.306
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.8	0.82	0.84
4	$\bar{x}$	0.39	0.4	0.42	0.42	0.43	0.45	0.48	0.51	0.54	0.56	0.6	0.62	0.64
	σ	0.488	0.369	0.319	0.294	0.275	0.264	0.251	0.252	0.25	0.258	0.263	0.274	0.276
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.68	0.73	0.76
5	$\bar{x}$	0.4	0.41	0.44	0.44	0.46	0.48	0.51	0.55	0.57	0.59	0.62	0.63	0.64
	σ	0.49	0.371	0.329	0.305	0.292	0.274	0.266	0.264	0.267	0.271	0.272	0.287	0.288
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.67	0.72	0.75	0.78
6	$\bar{x}$	0.42	0.43	0.46	0.47	0.5	0.53	0.55	0.59	0.6	0.65	0.65	0.67	0.7
	σ	0.493	0.366	0.317	0.296	0.286	0.269	0.269	0.268	0.28	0.279	0.288	0.294	0.289
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.78	0.82	0.85
7	$\bar{x}$	0.41	0.44	0.48	0.51	0.53	0.55	0.58	0.62	0.64	0.66	0.67	0.68	0.68
	σ	0.491	0.369	0.324	0.306	0.293	0.288	0.284	0.286	0.283	0.291	0.296	0.31	0.312
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.85
8	$\bar{x}$	0.41	0.41	0.43	0.44	0.46	0.48	0.51	0.54	0.57	0.61	0.63	0.66	0.67
	σ	0.492	0.374	0.328	0.305	0.285	0.276	0.265	0.261	0.259	0.268	0.27	0.277	0.275
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.6	0.7	0.72	0.78	0.79
9	$\bar{x}$	0.39	0.42	0.45	0.46	0.48	0.52	0.54	0.57	0.6	0.62	0.64	0.65	0.68
	σ	0.487	0.381	0.347	0.315	0.304	0.287	0.278	0.275	0.273	0.279	0.272	0.286	0.282
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.7	0.73	0.72	0.77	0.79
10	$\bar{x}$	0.39	0.42	0.44	0.45	0.47	0.49	0.53	0.56	0.57	0.61	0.63	0.66	0.67
	σ	0.488	0.387	0.345	0.323	0.312	0.297	0.285	0.281	0.279	0.279	0.275	0.277	0.286
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.78	0.8
11	$\bar{x}$	0.4	0.42	0.45	0.47	0.49	0.5	0.54	0.56	0.59	0.61	0.63	0.65	0.66
	σ	0.491	0.392	0.356	0.333	0.325	0.302	0.292	0.283	0.281	0.279	0.283	0.276	0.285
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.72	0.75	0.77
12	$\bar{x}$	0.39	0.42	0.45	0.47	0.49	0.52	0.56	0.59	0.62	0.65	0.65	0.69	0.72
	σ	0.487	0.366	0.321	0.302	0.29	0.271	0.268	0.272	0.276	0.276	0.298	0.292	0.29
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.87
13	$\bar{x}$	0.41	0.46	0.48	0.5	0.52	0.55	0.59	0.62	0.64	0.67	0.69	0.69	0.72
	σ	0.491	0.371	0.33	0.307	0.296	0.287	0.283	0.285	0.285	0.296	0.29	0.314	0.303
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.88
14	$\bar{x}$	0.39	0.44	0.46	0.47	0.49	0.54	0.57	0.59	0.62	0.64	0.66	0.69	0.7
	σ	0.488	0.381	0.336	0.319	0.304	0.287	0.281	0.283	0.275	0.286	0.295	0.287	0.295
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
15	$\bar{x}$	0.4	0.44	0.46	0.5	0.52	0.55	0.58	0.62	0.64	0.66	0.67	0.68	0.69
	σ	0.491	0.386	0.343	0.322	0.311	0.299	0.287	0.287	0.286	0.291	0.299	0.304	0.311
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.73	0.75	0.8	0.8	0.83	0.85

Table D.20: Local update for 5-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.47	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.356	0.298	0.263	0.237	0.21	0.185	0.161	0.144	0.138	0.126	0.122	0.115
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.49
2	$\bar{x}$	0.46	0.46	0.47	0.47	0.48	0.49	0.5	0.51	0.53	0.56	0.59	0.63	0.66
	σ	0.499	0.358	0.305	0.269	0.244	0.216	0.194	0.174	0.167	0.164	0.169	0.168	0.179
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.68	0.71
3	$\bar{x}$	0.47	0.48	0.5	0.51	0.53	0.55	0.58	0.61	0.64	0.67	0.7	0.74	0.76
	σ	0.499	0.364	0.303	0.268	0.25	0.226	0.208	0.2	0.197	0.202	0.197	0.202	0.194
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.75	0.8	0.82
4	$\bar{x}$	0.46	0.46	0.46	0.48	0.47	0.47	0.48	0.49	0.5	0.53	0.55	0.58	0.61
	σ	0.498	0.362	0.302	0.268	0.247	0.214	0.193	0.172	0.166	0.157	0.157	0.159	0.17
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.57	0.57	0.62	0.66
5	$\bar{x}$	0.46	0.46	0.48	0.49	0.5	0.51	0.54	0.56	0.6	0.63	0.65	0.67	0.7
	σ	0.498	0.363	0.313	0.281	0.268	0.234	0.214	0.2	0.187	0.185	0.188	0.201	0.191
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.73	0.76
6	$\bar{x}$	0.46	0.47	0.47	0.48	0.5	0.52	0.54	0.57	0.59	0.63	0.66	0.7	0.72
	σ	0.498	0.361	0.305	0.271	0.246	0.222	0.201	0.186	0.187	0.191	0.19	0.197	0.208
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.77	0.8
7	$\bar{x}$	0.46	0.5	0.52	0.55	0.56	0.58	0.62	0.65	0.67	0.71	0.74	0.75	0.77
	σ	0.499	0.362	0.301	0.275	0.257	0.228	0.21	0.205	0.204	0.197	0.195	0.212	0.217
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.73	0.75	0.77	0.8	0.82	0.85
8	$\bar{x}$	0.46	0.46	0.47	0.47	0.48	0.49	0.5	0.53	0.55	0.59	0.62	0.64	0.68
	σ	0.499	0.363	0.305	0.272	0.248	0.225	0.202	0.191	0.183	0.173	0.179	0.199	0.181
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.63	0.65	0.7	0.74
9	$\bar{x}$	0.47	0.49	0.5	0.52	0.52	0.54	0.57	0.6	0.62	0.65	0.67	0.7	0.72
	σ	0.499	0.37	0.322	0.291	0.264	0.243	0.225	0.206	0.205	0.191	0.195	0.184	0.198
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.75	0.78
10	$\bar{x}$	0.47	0.47	0.47	0.49	0.49	0.5	0.52	0.55	0.56	0.6	0.62	0.65	0.68
	σ	0.499	0.367	0.311	0.282	0.264	0.237	0.219	0.201	0.199	0.189	0.193	0.192	0.197
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.6	0.63	0.68	0.7	0.72
11	$\bar{x}$	0.46	0.47	0.48	0.5	0.51	0.52	0.55	0.57	0.58	0.61	0.63	0.66	0.67
	σ	0.498	0.378	0.332	0.309	0.285	0.266	0.243	0.226	0.221	0.211	0.203	0.195	0.196
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.68	0.7	0.72
12	$\bar{x}$	0.46	0.47	0.48	0.49	0.49	0.5	0.52	0.54	0.58	0.61	0.65	0.68	0.72
	σ	0.498	0.362	0.3	0.268	0.25	0.22	0.2	0.19	0.184	0.177	0.186	0.194	0.196
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.6	0.6	0.67	0.7	0.73	0.79
13	$\bar{x}$	0.46	0.49	0.51	0.54	0.55	0.58	0.62	0.65	0.68	0.71	0.73	0.76	0.77
	σ	0.498	0.364	0.305	0.272	0.255	0.232	0.212	0.208	0.2	0.215	0.213	0.221	0.225
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.75	0.77	0.8	0.83	0.86
14	$\bar{x}$	0.47	0.47	0.48	0.49	0.5	0.53	0.54	0.58	0.61	0.66	0.69	0.72	0.74
	σ	0.499	0.363	0.309	0.277	0.26	0.232	0.214	0.197	0.202	0.186	0.191	0.196	0.206
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.78	0.81
15	$\bar{x}$	0.45	0.48	0.52	0.54	0.55	0.59	0.62	0.65	0.67	0.7	0.72	0.75	0.77
	σ	0.497	0.372	0.316	0.285	0.269	0.242	0.221	0.21	0.21	0.207	0.207	0.213	0.217
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.67	0.7	0.77	0.79	0.83	0.85

Table D.21: Local update for 10-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.47	0.49	0.48	0.49	0.49	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.499	0.357	0.293	0.255	0.23	0.199	0.171	0.147	0.131	0.117	0.11	0.098	0.098
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.49
2	$\bar{x}$	0.48	0.49	0.49	0.49	0.48	0.49	0.49	0.5	0.51	0.51	0.52	0.54	0.57
	σ	0.5	0.358	0.291	0.258	0.236	0.199	0.173	0.148	0.132	0.122	0.113	0.116	0.11
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.57	0.59
3	$\bar{x}$	0.48	0.49	0.5	0.51	0.52	0.54	0.56	0.59	0.61	0.65	0.68	0.72	0.74
	σ	0.5	0.358	0.293	0.258	0.235	0.207	0.177	0.162	0.156	0.143	0.143	0.139	0.147
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.75	0.78
4	$\bar{x}$	0.47	0.49	0.47	0.48	0.48	0.48	0.49	0.49	0.49	0.5	0.5	0.51	0.53
	σ	0.499	0.36	0.299	0.256	0.236	0.203	0.177	0.155	0.133	0.121	0.117	0.109	0.112
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
5	$\bar{x}$	0.48	0.49	0.48	0.49	0.51	0.51	0.53	0.55	0.57	0.59	0.62	0.66	0.68
	σ	0.499	0.364	0.302	0.267	0.244	0.217	0.193	0.166	0.155	0.152	0.143	0.13	0.144
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.6	0.6	0.65	0.68	0.71
6	$\bar{x}$	0.48	0.47	0.49	0.49	0.49	0.5	0.51	0.53	0.55	0.58	0.6	0.64	0.67
	σ	0.5	0.358	0.295	0.259	0.236	0.202	0.179	0.153	0.134	0.129	0.125	0.128	0.139
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.7
7	$\bar{x}$	0.48	0.51	0.52	0.54	0.56	0.58	0.6	0.64	0.66	0.69	0.71	0.74	0.76
	σ	0.5	0.36	0.301	0.259	0.239	0.207	0.184	0.161	0.155	0.148	0.152	0.153	0.162
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.75	0.77	0.8
8	$\bar{x}$	0.49	0.48	0.49	0.49	0.49	0.49	0.5	0.51	0.52	0.54	0.56	0.59	0.62
	σ	0.5	0.36	0.298	0.261	0.235	0.207	0.18	0.157	0.147	0.133	0.13	0.126	0.119
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.57	0.6	0.64
9	$\bar{x}$	0.47	0.5	0.5	0.52	0.52	0.54	0.57	0.59	0.61	0.63	0.66	0.68	0.69
	σ	0.499	0.367	0.308	0.276	0.252	0.222	0.196	0.174	0.163	0.148	0.145	0.143	0.144
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.68	0.7	0.72
10	$\bar{x}$	0.47	0.48	0.49	0.49	0.49	0.49	0.5	0.51	0.52	0.53	0.55	0.58	0.6
	σ	0.499	0.357	0.301	0.265	0.241	0.209	0.185	0.163	0.152	0.145	0.143	0.141	0.137
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.57	0.6	0.63
11	$\bar{x}$	0.49	0.49	0.49	0.49	0.5	0.51	0.52	0.54	0.55	0.57	0.59	0.61	0.63
	σ	0.5	0.374	0.319	0.287	0.27	0.242	0.22	0.193	0.182	0.169	0.163	0.157	0.145
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.63	0.66
12	$\bar{x}$	0.47	0.47	0.48	0.48	0.49	0.49	0.5	0.5	0.52	0.53	0.55	0.58	0.6
	σ	0.499	0.354	0.297	0.257	0.236	0.204	0.176	0.152	0.139	0.134	0.125	0.129	0.139
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.57	0.6	0.63
13	$\bar{x}$	0.49	0.49	0.51	0.52	0.54	0.56	0.59	0.62	0.65	0.68	0.71	0.74	0.77
	σ	0.5	0.361	0.299	0.263	0.238	0.208	0.184	0.169	0.156	0.158	0.148	0.153	0.152
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.65	0.7	0.75	0.78	0.81
14	$\bar{x}$	0.49	0.49	0.49	0.49	0.5	0.5	0.51	0.54	0.55	0.59	0.61	0.66	0.69
	σ	0.5	0.357	0.296	0.264	0.236	0.207	0.182	0.16	0.15	0.138	0.138	0.142	0.138
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.63	0.68	0.71
15	$\bar{x}$	0.48	0.51	0.52	0.54	0.55	0.57	0.6	0.63	0.65	0.68	0.71	0.74	0.76
	σ	0.5	0.362	0.302	0.268	0.244	0.217	0.192	0.17	0.164	0.155	0.148	0.143	0.163
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.7	0.7	0.75	0.77	0.8

Table D.22: Local update for 20-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.5	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	σ	0.5	0.353	0.289	0.255	0.229	0.196	0.164	0.14	0.127	0.103	0.098	0.078	0.083
	m	1	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.48	0.49	0.49	0.5	0.49	0.49	0.49	0.49	0.5	0.5	0.5	0.51	0.51
	σ	0.5	0.356	0.293	0.256	0.229	0.197	0.165	0.142	0.125	0.106	0.098	0.085	0.078
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.53
3	$\bar{x}$	0.47	0.48	0.5	0.51	0.51	0.52	0.54	0.56	0.58	0.61	0.63	0.67	0.69
	σ	0.499	0.358	0.289	0.254	0.226	0.197	0.169	0.145	0.128	0.11	0.11	0.094	0.106
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.6	0.63	0.65	0.68	0.71
4	$\bar{x}$	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.49	0.5	0.5
	σ	0.5	0.36	0.292	0.254	0.23	0.199	0.163	0.139	0.123	0.108	0.104	0.084	0.077
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.51
5	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.53	0.55	0.56	0.59	0.62	0.64
	σ	0.5	0.354	0.294	0.258	0.232	0.202	0.175	0.149	0.138	0.124	0.112	0.099	0.102
	m	0	0.5	0.67	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.62	0.65
6	$\bar{x}$	0.49	0.48	0.49	0.49	0.49	0.5	0.5	0.51	0.51	0.53	0.54	0.57	0.58
	σ	0.5	0.354	0.291	0.253	0.229	0.201	0.168	0.138	0.128	0.11	0.099	0.088	0.097
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.57	0.59
7	$\bar{x}$	0.5	0.51	0.52	0.53	0.54	0.57	0.59	0.61	0.62	0.66	0.68	0.71	0.73
	σ	0.5	0.357	0.292	0.254	0.229	0.197	0.167	0.145	0.134	0.112	0.103	0.113	0.107
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.73	0.75
8	$\bar{x}$	0.49	0.48	0.49	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.54	0.55
	σ	0.5	0.352	0.295	0.253	0.232	0.2	0.168	0.143	0.128	0.109	0.101	0.089	0.087
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.5	0.5	0.53	0.53	0.56
9	$\bar{x}$	0.49	0.49	0.51	0.52	0.52	0.54	0.55	0.57	0.58	0.6	0.63	0.65	0.67
	σ	0.5	0.364	0.296	0.268	0.237	0.208	0.181	0.155	0.139	0.122	0.116	0.103	0.097
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.6	0.65	0.65	0.68
10	$\bar{x}$	0.49	0.48	0.49	0.49	0.49	0.49	0.49	0.5	0.5	0.51	0.51	0.52	0.53
	σ	0.5	0.356	0.287	0.256	0.23	0.201	0.17	0.144	0.132	0.117	0.105	0.095	0.095
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.53	0.52	0.53
11	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.52	0.54	0.55	0.56	0.57
	σ	0.5	0.363	0.306	0.279	0.251	0.221	0.198	0.173	0.162	0.141	0.135	0.131	0.121
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.53	0.55	0.57	0.57
12	$\bar{x}$	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.49	0.5	0.51	0.51	0.52	0.52
	σ	0.5	0.355	0.294	0.256	0.231	0.196	0.167	0.141	0.125	0.109	0.101	0.086	0.082
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.53
13	$\bar{x}$	0.5	0.5	0.51	0.51	0.52	0.53	0.55	0.57	0.6	0.62	0.65	0.69	0.72
	σ	0.5	0.357	0.291	0.257	0.233	0.197	0.173	0.148	0.134	0.123	0.114	0.104	0.095
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.7	0.74
14	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.49	0.51	0.51	0.52	0.53	0.55	0.57	0.59
	σ	0.5	0.354	0.291	0.256	0.231	0.197	0.168	0.144	0.127	0.112	0.098	0.103	0.104
	m	1	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.55	0.58	0.6
15	$\bar{x}$	0.48	0.5	0.52	0.53	0.54	0.56	0.58	0.61	0.62	0.66	0.68	0.71	0.73
	σ	0.5	0.355	0.298	0.258	0.231	0.202	0.174	0.148	0.135	0.121	0.109	0.109	0.112
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.73	0.75

Table D.23: Local update for 40-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.49	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.49	0.49	0.49	0.5	0.5
	σ	0.5	0.359	0.292	0.25	0.223	0.193	0.161	0.134	0.118	0.095	0.085	0.07	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.49	0.49	0.5	0.5	0.5	0.5
	σ	0.5	0.355	0.297	0.251	0.224	0.192	0.164	0.135	0.117	0.101	0.084	0.075	0.07
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
3	$\bar{x}$	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.54	0.55	0.57	0.59	0.61	0.64
	σ	0.5	0.354	0.29	0.254	0.228	0.192	0.162	0.134	0.12	0.101	0.093	0.082	0.077
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.62	0.65
4	$\bar{x}$	0.51	0.5	0.5	0.5	0.5	0.49	0.49	0.5	0.5	0.5	0.49	0.5	0.5
	σ	0.5	0.352	0.288	0.252	0.228	0.192	0.16	0.133	0.119	0.098	0.089	0.074	0.072
	m	1	0.5	0.33	0.5	0.4	0.43	0.5	0.5	0.5	0.5	0.5	0.5	0.5
5	$\bar{x}$	0.5	0.49	0.5	0.5	0.5	0.51	0.51	0.52	0.52	0.54	0.55	0.58	0.59
	σ	0.5	0.355	0.294	0.257	0.224	0.197	0.166	0.138	0.123	0.107	0.098	0.083	0.074
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
6	$\bar{x}$	0.49	0.51	0.49	0.49	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.53
	σ	0.5	0.354	0.292	0.254	0.227	0.191	0.163	0.134	0.118	0.097	0.088	0.077	0.077
	m	0	0.5	0.67	0.5	0.6	0.43	0.5	0.53	0.5	0.5	0.53	0.53	0.54
7	$\bar{x}$	0.49	0.51	0.52	0.52	0.53	0.55	0.57	0.58	0.6	0.63	0.64	0.67	0.69
	σ	0.5	0.353	0.291	0.256	0.229	0.191	0.161	0.137	0.115	0.101	0.095	0.086	0.072
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.68	0.7
8	$\bar{x}$	0.5	0.49	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51
	σ	0.5	0.352	0.29	0.253	0.225	0.195	0.16	0.136	0.118	0.101	0.098	0.08	0.075
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.53
9	$\bar{x}$	0.5	0.5	0.52	0.52	0.52	0.53	0.54	0.56	0.57	0.58	0.59	0.61	0.63
	σ	0.5	0.36	0.296	0.256	0.237	0.203	0.169	0.144	0.13	0.106	0.096	0.094	0.078
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.62	0.64
10	$\bar{x}$	0.49	0.49	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51
	σ	0.5	0.353	0.289	0.252	0.232	0.193	0.165	0.138	0.12	0.098	0.093	0.077	0.068
	m	0	0.5	0.33	0.5	0.6	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.51
11	$\bar{x}$	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.52	0.52	0.53
	σ	0.5	0.364	0.303	0.264	0.248	0.215	0.186	0.158	0.142	0.127	0.123	0.108	0.098
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.53	0.53	0.54
12	$\bar{x}$	0.49	0.51	0.49	0.5	0.5	0.49	0.49	0.5	0.49	0.5	0.5	0.5	0.5
	σ	0.5	0.351	0.287	0.255	0.225	0.194	0.161	0.136	0.115	0.1	0.086	0.074	0.067
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.5	0.51
13	$\bar{x}$	0.51	0.49	0.5	0.5	0.51	0.51	0.53	0.53	0.55	0.56	0.59	0.62	0.64
	σ	0.5	0.355	0.29	0.252	0.226	0.199	0.165	0.14	0.126	0.111	0.104	0.088	0.091
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.62	0.65
14	$\bar{x}$	0.49	0.49	0.5	0.49	0.5	0.5	0.5	0.5	0.51	0.51	0.51	0.53	0.53
	σ	0.5	0.353	0.298	0.249	0.229	0.191	0.161	0.134	0.119	0.098	0.086	0.077	0.071
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.53	0.53	0.54
15	$\bar{x}$	0.5	0.51	0.52	0.53	0.54	0.55	0.56	0.58	0.6	0.62	0.63	0.67	0.69
	σ	0.5	0.351	0.292	0.256	0.224	0.195	0.164	0.137	0.123	0.101	0.099	0.086	0.081
	m	1	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.67	0.7

Table D.24: Local update for 80-dimensional tablet function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.5	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.36	0.289	0.252	0.223	0.19	0.16	0.131	0.113	0.092	0.084	0.071	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
2	$\bar{x}$	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.355	0.289	0.245	0.221	0.191	0.163	0.133	0.116	0.092	0.083	0.069	0.062
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
3	$\bar{x}$	0.5	0.5	0.5	0.5	0.51	0.5	0.51	0.52	0.53	0.54	0.55	0.58	0.59
	σ	0.5	0.354	0.294	0.254	0.224	0.191	0.161	0.132	0.115	0.095	0.085	0.066	0.069
	m	1	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.58	0.6
4	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.49	0.5	0.5	0.49	0.5	0.5	0.5	0.5
	σ	0.5	0.355	0.29	0.251	0.224	0.192	0.158	0.132	0.113	0.093	0.08	0.07	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
5	$\bar{x}$	0.51	0.49	0.5	0.5	0.5	0.51	0.5	0.51	0.51	0.52	0.52	0.54	0.55
	σ	0.5	0.36	0.29	0.251	0.225	0.195	0.162	0.134	0.12	0.097	0.09	0.073	0.065
	m	1	0.5	0.67	0.5	0.4	0.57	0.5	0.53	0.5	0.53	0.53	0.53	0.55
6	$\bar{x}$	0.48	0.5	0.5	0.51	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51
	σ	0.5	0.354	0.287	0.25	0.224	0.191	0.16	0.131	0.113	0.094	0.084	0.069	0.064
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.52	0.51
7	$\bar{x}$	0.51	0.52	0.52	0.52	0.52	0.54	0.55	0.56	0.57	0.59	0.6	0.63	0.65
	σ	0.5	0.352	0.288	0.252	0.225	0.191	0.161	0.129	0.116	0.095	0.086	0.068	0.069
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.63	0.65
8	$\bar{x}$	0.49	0.5	0.5	0.5	0.49	0.5	0.49	0.5	0.49	0.5	0.5	0.5	0.51
	σ	0.5	0.355	0.286	0.249	0.228	0.193	0.163	0.131	0.114	0.093	0.084	0.069	0.063
	m	0	0.5	0.67	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.51
9	$\bar{x}$	0.5	0.51	0.52	0.51	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.59	0.6
	σ	0.5	0.356	0.295	0.257	0.235	0.2	0.169	0.138	0.121	0.1	0.091	0.081	0.069
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.6
10	$\bar{x}$	0.5	0.49	0.49	0.49	0.5	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.352	0.289	0.252	0.225	0.188	0.158	0.133	0.116	0.095	0.084	0.069	0.059
	m	1	0.5	0.33	0.5	0.4	0.57	0.5	0.47	0.5	0.5	0.5	0.5	0.5
11	$\bar{x}$	0.49	0.49	0.5	0.5	0.5	0.5	0.51	0.5	0.51	0.51	0.51	0.51	0.52
	σ	0.5	0.356	0.299	0.265	0.239	0.204	0.176	0.148	0.13	0.115	0.102	0.09	0.087
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.5	0.5	0.52	0.51
12	$\bar{x}$	0.49	0.51	0.5	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	σ	0.5	0.353	0.29	0.254	0.224	0.191	0.162	0.131	0.113	0.093	0.084	0.072	0.06
	m	0	0.5	0.67	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
13	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.56	0.57
	σ	0.5	0.354	0.285	0.252	0.221	0.192	0.159	0.135	0.115	0.1	0.089	0.08	0.075
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.53	0.55	0.57	0.57
14	$\bar{x}$	0.5	0.5	0.5	0.5	0.5	0.5	0.49	0.5	0.5	0.5	0.5	0.51	0.51
	σ	0.5	0.355	0.289	0.25	0.227	0.189	0.16	0.134	0.114	0.094	0.086	0.073	0.06
	m	1	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.5	0.5	0.5	0.5	0.51
15	$\bar{x}$	0.49	0.5	0.52	0.52	0.53	0.54	0.55	0.56	0.57	0.59	0.61	0.63	0.65
	σ	0.5	0.355	0.291	0.25	0.223	0.191	0.162	0.132	0.117	0.095	0.084	0.076	0.068
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.63	0.65

D.5 Ellipsoid Function

Table D.25: Local update for 2-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.4	0.4	0.39	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.41	0.4
	σ	0.49	0.358	0.305	0.274	0.251	0.227	0.206	0.186	0.171	0.161	0.16	0.155	0.144
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.39	0.42	0.43	0.45	0.46	0.49	0.51	0.54	0.56	0.6	0.61	0.63	0.66
	σ	0.489	0.366	0.314	0.287	0.276	0.262	0.255	0.261	0.266	0.273	0.288	0.295	0.292
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.75	0.78	0.81
3	$\bar{x}$	0.41	0.43	0.45	0.47	0.49	0.53	0.55	0.58	0.59	0.63	0.64	0.68	0.66
	σ	0.491	0.37	0.324	0.299	0.292	0.279	0.271	0.275	0.276	0.286	0.291	0.295	0.309
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.78	0.83	0.82
4	$\bar{x}$	0.39	0.41	0.42	0.42	0.44	0.45	0.47	0.51	0.54	0.57	0.6	0.61	0.64
	σ	0.488	0.367	0.318	0.295	0.277	0.261	0.25	0.247	0.252	0.257	0.26	0.277	0.28
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.7	0.73	0.76
5	$\bar{x}$	0.4	0.43	0.43	0.44	0.46	0.48	0.51	0.54	0.56	0.59	0.62	0.65	0.65
	σ	0.49	0.375	0.328	0.304	0.291	0.273	0.268	0.262	0.268	0.27	0.273	0.27	0.281
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.67	0.72	0.75	0.78
6	$\bar{x}$	0.38	0.43	0.45	0.48	0.49	0.52	0.56	0.59	0.6	0.63	0.65	0.66	0.69
	σ	0.485	0.366	0.317	0.298	0.286	0.272	0.268	0.267	0.275	0.285	0.288	0.295	0.297
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.82	0.85
7	$\bar{x}$	0.4	0.45	0.48	0.5	0.53	0.56	0.59	0.62	0.65	0.66	0.67	0.68	0.72
	σ	0.49	0.375	0.327	0.311	0.295	0.284	0.282	0.284	0.277	0.292	0.295	0.307	0.289
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.83	0.88
8	$\bar{x}$	0.41	0.42	0.43	0.44	0.45	0.47	0.51	0.54	0.57	0.59	0.61	0.64	0.65
	σ	0.491	0.376	0.331	0.309	0.289	0.272	0.265	0.264	0.266	0.273	0.275	0.283	0.286
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.7	0.72	0.76	0.78
9	$\bar{x}$	0.4	0.43	0.45	0.47	0.49	0.52	0.55	0.58	0.6	0.62	0.64	0.67	0.67
	σ	0.491	0.384	0.336	0.317	0.304	0.288	0.278	0.27	0.267	0.281	0.277	0.276	0.285
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.65	0.7	0.75	0.78	0.79
10	$\bar{x}$	0.4	0.41	0.44	0.46	0.48	0.5	0.53	0.57	0.57	0.62	0.63	0.65	0.67
	σ	0.491	0.386	0.342	0.323	0.309	0.291	0.283	0.277	0.28	0.271	0.28	0.286	0.285
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.77	0.78
11	$\bar{x}$	0.39	0.43	0.45	0.47	0.48	0.51	0.54	0.57	0.58	0.6	0.64	0.65	0.67
	σ	0.488	0.395	0.351	0.334	0.321	0.304	0.293	0.285	0.283	0.281	0.278	0.287	0.279
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.72	0.75	0.79
12	$\bar{x}$	0.4	0.43	0.45	0.47	0.49	0.52	0.55	0.59	0.61	0.64	0.66	0.68	0.69
	σ	0.49	0.369	0.32	0.301	0.287	0.276	0.269	0.272	0.276	0.283	0.288	0.295	0.302
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
13	$\bar{x}$	0.4	0.45	0.48	0.51	0.53	0.56	0.6	0.62	0.66	0.67	0.68	0.7	0.69
	σ	0.49	0.37	0.33	0.309	0.295	0.287	0.282	0.284	0.284	0.287	0.295	0.301	0.315
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.8	0.8	0.82	0.85	0.86
14	$\bar{x}$	0.4	0.43	0.45	0.48	0.48	0.53	0.55	0.6	0.62	0.64	0.67	0.68	0.7
	σ	0.49	0.38	0.338	0.319	0.305	0.287	0.287	0.282	0.282	0.288	0.289	0.3	0.297
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.7	0.77	0.8	0.82	0.86
15	$\bar{x}$	0.4	0.44	0.47	0.49	0.52	0.55	0.58	0.61	0.63	0.66	0.66	0.68	0.71
	σ	0.49	0.388	0.346	0.321	0.308	0.297	0.29	0.285	0.288	0.288	0.3	0.292	0.294
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.73	0.75	0.8	0.8	0.83	0.86

Table D.26: Local update for 5-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.46	0.45	0.46	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.45	0.46	0.46
	σ	0.498	0.359	0.298	0.265	0.239	0.211	0.185	0.161	0.149	0.134	0.128	0.116	0.112
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.48	0.49
2	$\bar{x}$	0.44	0.46	0.46	0.46	0.47	0.48	0.49	0.51	0.52	0.56	0.59	0.62	0.66
	σ	0.497	0.361	0.297	0.267	0.242	0.214	0.19	0.175	0.167	0.162	0.161	0.172	0.181
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.67	0.71
3	$\bar{x}$	0.45	0.48	0.49	0.51	0.52	0.54	0.58	0.62	0.64	0.68	0.7	0.73	0.75
	σ	0.498	0.363	0.302	0.269	0.251	0.225	0.208	0.195	0.189	0.189	0.193	0.194	0.201
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.8	0.82
4	$\bar{x}$	0.43	0.46	0.45	0.46	0.47	0.47	0.48	0.49	0.5	0.52	0.54	0.58	0.62
	σ	0.495	0.359	0.3	0.265	0.244	0.216	0.195	0.172	0.163	0.155	0.16	0.164	0.165
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.57	0.62	0.66
5	$\bar{x}$	0.46	0.46	0.48	0.48	0.49	0.51	0.53	0.56	0.59	0.62	0.65	0.68	0.7
	σ	0.498	0.365	0.312	0.279	0.261	0.238	0.213	0.199	0.194	0.185	0.171	0.187	0.179
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.65	0.67	0.7	0.73	0.75
6	$\bar{x}$	0.45	0.46	0.48	0.48	0.5	0.51	0.53	0.56	0.6	0.64	0.66	0.7	0.73
	σ	0.498	0.36	0.303	0.268	0.247	0.223	0.204	0.186	0.176	0.174	0.183	0.189	0.19
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.72	0.77	0.8
7	$\bar{x}$	0.46	0.49	0.52	0.54	0.56	0.59	0.62	0.65	0.68	0.7	0.73	0.76	0.78
	σ	0.499	0.363	0.305	0.268	0.25	0.227	0.212	0.201	0.19	0.2	0.192	0.196	0.186
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.67	0.75	0.77	0.8	0.82	0.85
8	$\bar{x}$	0.45	0.46	0.46	0.47	0.48	0.49	0.5	0.53	0.54	0.58	0.61	0.65	0.69
	σ	0.498	0.362	0.305	0.275	0.253	0.225	0.203	0.186	0.176	0.178	0.175	0.176	0.168
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.63	0.65	0.7	0.74
9	$\bar{x}$	0.45	0.48	0.49	0.51	0.51	0.54	0.56	0.6	0.62	0.65	0.67	0.7	0.71
	σ	0.498	0.373	0.319	0.291	0.27	0.242	0.224	0.208	0.195	0.183	0.183	0.181	0.186
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.7	0.72	0.75	0.76
10	$\bar{x}$	0.45	0.47	0.46	0.47	0.49	0.49	0.51	0.54	0.56	0.59	0.61	0.65	0.67
	σ	0.497	0.362	0.312	0.281	0.262	0.235	0.22	0.205	0.196	0.195	0.194	0.182	0.187
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.6	0.63	0.65	0.7	0.72
11	$\bar{x}$	0.45	0.47	0.48	0.49	0.5	0.52	0.54	0.56	0.58	0.6	0.63	0.65	0.67
	σ	0.498	0.378	0.334	0.306	0.287	0.262	0.243	0.226	0.219	0.202	0.193	0.193	0.188
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.68	0.72
12	$\bar{x}$	0.46	0.46	0.47	0.48	0.49	0.5	0.51	0.54	0.57	0.6	0.63	0.69	0.71
	σ	0.498	0.357	0.304	0.267	0.25	0.221	0.199	0.184	0.178	0.174	0.18	0.17	0.189
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.6	0.63	0.68	0.73	0.78
13	$\bar{x}$	0.45	0.48	0.51	0.53	0.54	0.58	0.61	0.65	0.68	0.72	0.74	0.76	0.78
	σ	0.498	0.363	0.306	0.27	0.259	0.228	0.213	0.205	0.197	0.2	0.198	0.202	0.202
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.75	0.77	0.8	0.83	0.86
14	$\bar{x}$	0.44	0.46	0.48	0.48	0.49	0.52	0.54	0.58	0.61	0.65	0.68	0.71	0.74
	σ	0.496	0.365	0.305	0.277	0.255	0.23	0.215	0.2	0.195	0.189	0.188	0.192	0.193
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.78	0.81
15	$\bar{x}$	0.46	0.49	0.51	0.53	0.55	0.58	0.62	0.65	0.67	0.7	0.72	0.75	0.76
	σ	0.498	0.371	0.318	0.286	0.264	0.244	0.218	0.205	0.208	0.203	0.197	0.2	0.212
	m	0	0.5	0.67	0.5	0.6	0.57	0.7	0.67	0.7	0.77	0.8	0.82	0.84

Table D.27: Local update for 10-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47
	σ	0.499	0.356	0.29	0.256	0.227	0.192	0.167	0.137	0.121	0.103	0.095	0.083	0.074
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.47	0.47
2	$\bar{x}$	0.47	0.47	0.47	0.47	0.47	0.47	0.48	0.49	0.49	0.5	0.51	0.54	0.55
	σ	0.499	0.356	0.289	0.254	0.225	0.196	0.165	0.139	0.125	0.105	0.098	0.089	0.084
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.56
3	$\bar{x}$	0.48	0.48	0.49	0.5	0.51	0.53	0.55	0.58	0.61	0.64	0.67	0.71	0.74
	σ	0.5	0.355	0.293	0.254	0.229	0.197	0.17	0.143	0.13	0.112	0.103	0.097	0.086
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.6	0.67	0.68	0.73	0.76
4	$\bar{x}$	0.47	0.47	0.47	0.47	0.47	0.47	0.48	0.47	0.48	0.49	0.49	0.5	0.51
	σ	0.499	0.358	0.288	0.255	0.229	0.195	0.163	0.141	0.125	0.107	0.096	0.094	0.086
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.53
5	$\bar{x}$	0.47	0.47	0.48	0.48	0.49	0.5	0.52	0.54	0.56	0.59	0.62	0.66	0.68
	σ	0.499	0.36	0.296	0.259	0.237	0.207	0.18	0.157	0.141	0.126	0.112	0.102	0.098
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69
6	$\bar{x}$	0.46	0.48	0.48	0.48	0.48	0.48	0.5	0.52	0.54	0.57	0.59	0.63	0.67
	σ	0.499	0.355	0.29	0.258	0.226	0.197	0.167	0.142	0.128	0.106	0.103	0.094	0.084
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.69
7	$\bar{x}$	0.48	0.5	0.52	0.53	0.55	0.57	0.6	0.63	0.66	0.69	0.72	0.74	0.77
	σ	0.5	0.355	0.294	0.256	0.23	0.195	0.169	0.141	0.126	0.107	0.104	0.098	0.088
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.65	0.7	0.72	0.77	0.79
8	$\bar{x}$	0.47	0.47	0.48	0.47	0.47	0.48	0.48	0.5	0.5	0.52	0.54	0.59	0.61
	σ	0.499	0.358	0.294	0.255	0.233	0.198	0.171	0.145	0.134	0.117	0.108	0.094	0.091
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.55	0.6	0.63
9	$\bar{x}$	0.47	0.48	0.49	0.51	0.52	0.53	0.56	0.59	0.6	0.63	0.65	0.67	0.7
	σ	0.499	0.365	0.307	0.272	0.246	0.216	0.185	0.161	0.143	0.125	0.113	0.103	0.097
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.6	0.63	0.65	0.68	0.71
10	$\bar{x}$	0.47	0.47	0.46	0.48	0.47	0.48	0.48	0.49	0.49	0.51	0.52	0.54	0.57
	σ	0.499	0.353	0.294	0.261	0.236	0.202	0.176	0.155	0.142	0.129	0.123	0.116	0.109
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.53	0.55	0.57
11	$\bar{x}$	0.46	0.48	0.47	0.48	0.48	0.49	0.5	0.51	0.52	0.54	0.55	0.56	0.58
	σ	0.498	0.377	0.318	0.286	0.263	0.238	0.208	0.191	0.175	0.154	0.148	0.138	0.131
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.55	0.57	0.59
12	$\bar{x}$	0.47	0.46	0.47	0.47	0.47	0.47	0.48	0.49	0.5	0.51	0.52	0.54	0.56
	σ	0.499	0.359	0.291	0.248	0.225	0.196	0.166	0.138	0.118	0.104	0.096	0.084	0.081
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.57
13	$\bar{x}$	0.46	0.48	0.49	0.51	0.51	0.53	0.56	0.59	0.62	0.67	0.69	0.74	0.76
	σ	0.499	0.351	0.293	0.259	0.231	0.198	0.171	0.148	0.139	0.116	0.107	0.092	0.094
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.6	0.65	0.67	0.7	0.75	0.78
14	$\bar{x}$	0.46	0.47	0.47	0.48	0.48	0.49	0.51	0.52	0.54	0.57	0.6	0.65	0.68
	σ	0.499	0.356	0.292	0.255	0.235	0.199	0.169	0.145	0.134	0.117	0.111	0.1	0.093
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.63	0.67	0.7
15	$\bar{x}$	0.47	0.49	0.51	0.54	0.55	0.57	0.6	0.63	0.65	0.68	0.71	0.75	0.77
	σ	0.499	0.362	0.302	0.263	0.237	0.205	0.171	0.146	0.132	0.12	0.108	0.097	0.089
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.67	0.65	0.7	0.72	0.77	0.79

Table D.28: Local update for 20-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.48	0.47	0.47	0.47	0.47	0.48	0.48	0.47	0.48	0.47	0.47	0.48	0.47
	σ	0.5	0.353	0.287	0.251	0.224	0.188	0.159	0.128	0.116	0.094	0.082	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.47
2	$\bar{x}$	0.48	0.48	0.48	0.48	0.48	0.47	0.48	0.48	0.48	0.48	0.49	0.49	0.5
	σ	0.5	0.352	0.284	0.25	0.224	0.194	0.158	0.13	0.116	0.092	0.081	0.068	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
3	$\bar{x}$	0.49	0.48	0.49	0.49	0.5	0.51	0.52	0.55	0.56	0.6	0.62	0.66	0.69
	σ	0.5	0.35	0.284	0.251	0.223	0.191	0.161	0.13	0.114	0.093	0.081	0.067	0.058
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69
4	$\bar{x}$	0.48	0.48	0.47	0.47	0.48	0.47	0.47	0.48	0.48	0.48	0.48	0.49	0.49
	σ	0.5	0.35	0.286	0.248	0.228	0.192	0.159	0.13	0.115	0.094	0.084	0.069	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.48	0.47	0.48	0.49
5	$\bar{x}$	0.47	0.47	0.48	0.48	0.49	0.49	0.5	0.52	0.53	0.55	0.58	0.61	0.63
	σ	0.499	0.355	0.294	0.253	0.234	0.198	0.168	0.141	0.126	0.105	0.092	0.077	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.62	0.64
6	$\bar{x}$	0.48	0.49	0.48	0.48	0.48	0.48	0.49	0.5	0.51	0.51	0.53	0.55	0.57
	σ	0.499	0.35	0.285	0.254	0.225	0.191	0.159	0.129	0.119	0.093	0.084	0.067	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.57
7	$\bar{x}$	0.48	0.5	0.51	0.52	0.54	0.56	0.57	0.6	0.62	0.65	0.68	0.7	0.73
	σ	0.5	0.352	0.287	0.254	0.225	0.187	0.162	0.127	0.112	0.09	0.075	0.062	0.053
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.67	0.68	0.7	0.72
8	$\bar{x}$	0.47	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.49	0.49	0.51	0.52	0.53
	σ	0.499	0.356	0.292	0.251	0.225	0.19	0.16	0.133	0.115	0.099	0.086	0.074	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.53	0.54
9	$\bar{x}$	0.48	0.48	0.49	0.5	0.5	0.53	0.54	0.56	0.57	0.59	0.61	0.64	0.66
	σ	0.5	0.365	0.302	0.264	0.24	0.205	0.172	0.145	0.13	0.106	0.091	0.073	0.07
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.63	0.65	0.66
10	$\bar{x}$	0.47	0.47	0.48	0.48	0.47	0.48	0.48	0.48	0.48	0.48	0.48	0.49	0.49
	σ	0.499	0.353	0.292	0.25	0.232	0.191	0.165	0.135	0.12	0.105	0.091	0.08	0.076
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.49
11	$\bar{x}$	0.47	0.48	0.48	0.48	0.48	0.48	0.49	0.49	0.5	0.5	0.5	0.5	0.51
	σ	0.499	0.366	0.307	0.277	0.253	0.22	0.19	0.17	0.151	0.137	0.125	0.112	0.104
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.51
12	$\bar{x}$	0.47	0.47	0.48	0.47	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.499	0.35	0.289	0.247	0.222	0.189	0.158	0.133	0.116	0.093	0.079	0.07	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.49
13	$\bar{x}$	0.48	0.48	0.48	0.49	0.48	0.49	0.51	0.52	0.54	0.57	0.59	0.63	0.66
	σ	0.5	0.351	0.294	0.254	0.222	0.192	0.16	0.131	0.115	0.096	0.085	0.072	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.66
14	$\bar{x}$	0.48	0.48	0.47	0.48	0.49	0.49	0.49	0.49	0.5	0.52	0.53	0.56	0.58
	σ	0.5	0.353	0.289	0.248	0.222	0.189	0.16	0.132	0.116	0.094	0.083	0.071	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.53	0.57	0.57
15	$\bar{x}$	0.47	0.49	0.51	0.52	0.53	0.55	0.57	0.6	0.62	0.65	0.67	0.7	0.73
	σ	0.499	0.355	0.29	0.255	0.227	0.192	0.158	0.129	0.11	0.09	0.079	0.063	0.057
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.67	0.68	0.72	0.74

Table D.29: Local update for 40-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.49	0.48	0.47	0.48	0.48	0.48	0.49	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.5	0.355	0.29	0.25	0.226	0.187	0.158	0.129	0.11	0.094	0.08	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.47
2	$\bar{x}$	0.48	0.48	0.48	0.48	0.49	0.48	0.48	0.48	0.48	0.48	0.49	0.48	0.49
	σ	0.5	0.354	0.29	0.252	0.224	0.189	0.16	0.129	0.112	0.091	0.081	0.061	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.47	0.48	0.49
3	$\bar{x}$	0.48	0.48	0.49	0.5	0.49	0.5	0.51	0.52	0.53	0.55	0.58	0.6	0.63
	σ	0.5	0.354	0.29	0.249	0.224	0.189	0.157	0.129	0.112	0.093	0.079	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.63
4	$\bar{x}$	0.47	0.47	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.499	0.353	0.288	0.252	0.223	0.186	0.159	0.128	0.111	0.091	0.082	0.064	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.5	0.48	0.49
5	$\bar{x}$	0.5	0.48	0.49	0.48	0.48	0.49	0.5	0.5	0.51	0.52	0.54	0.57	0.58
	σ	0.5	0.356	0.289	0.255	0.225	0.195	0.162	0.135	0.118	0.099	0.09	0.073	0.066
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.53	0.57	0.59
6	$\bar{x}$	0.48	0.49	0.48	0.49	0.48	0.48	0.49	0.49	0.5	0.5	0.5	0.51	0.52
	σ	0.499	0.353	0.289	0.252	0.222	0.188	0.161	0.132	0.111	0.091	0.08	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.51
7	$\bar{x}$	0.48	0.5	0.5	0.52	0.52	0.54	0.55	0.57	0.59	0.61	0.63	0.66	0.68
	σ	0.5	0.357	0.286	0.251	0.223	0.188	0.16	0.127	0.109	0.089	0.077	0.062	0.05
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.6	0.63	0.67	0.69
8	$\bar{x}$	0.47	0.48	0.48	0.49	0.48	0.48	0.48	0.49	0.48	0.49	0.49	0.5	0.5
	σ	0.499	0.352	0.289	0.252	0.225	0.187	0.159	0.13	0.115	0.094	0.083	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
9	$\bar{x}$	0.49	0.49	0.49	0.5	0.51	0.51	0.53	0.54	0.56	0.57	0.59	0.6	0.62
	σ	0.5	0.36	0.294	0.261	0.235	0.2	0.167	0.141	0.124	0.098	0.088	0.071	0.065
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.6	0.63
10	$\bar{x}$	0.47	0.49	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.499	0.354	0.289	0.248	0.223	0.191	0.16	0.131	0.114	0.096	0.08	0.07	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.47
11	$\bar{x}$	0.48	0.49	0.47	0.48	0.49	0.48	0.48	0.48	0.48	0.48	0.49	0.48	0.49
	σ	0.5	0.364	0.3	0.267	0.243	0.211	0.181	0.157	0.14	0.123	0.109	0.094	0.091
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.47	0.47	0.48	0.49
12	$\bar{x}$	0.48	0.48	0.48	0.49	0.49	0.48	0.49	0.48	0.48	0.48	0.48	0.48	0.48
	σ	0.5	0.352	0.29	0.244	0.224	0.188	0.16	0.13	0.112	0.093	0.078	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.47	0.48	0.47
13	$\bar{x}$	0.48	0.48	0.48	0.49	0.48	0.49	0.49	0.5	0.5	0.51	0.52	0.53	0.55
	σ	0.5	0.349	0.292	0.249	0.223	0.188	0.156	0.129	0.109	0.091	0.082	0.065	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
14	$\bar{x}$	0.49	0.49	0.48	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.5	0.51	0.52
	σ	0.5	0.353	0.291	0.251	0.227	0.188	0.159	0.128	0.113	0.092	0.078	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.52	0.53
15	$\bar{x}$	0.47	0.5	0.51	0.52	0.52	0.53	0.55	0.57	0.59	0.61	0.63	0.66	0.68
	σ	0.499	0.351	0.293	0.251	0.227	0.191	0.159	0.126	0.112	0.09	0.074	0.061	0.051
	m	0	0.5	0.67	0.5	0.6	0.57	0.6	0.6	0.6	0.6	0.63	0.67	0.68

Table D.30: Local update for 80-dimensional ellipsoid function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.49	0.48	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.48	0.48	0.48	0.49
	σ	0.5	0.354	0.29	0.249	0.223	0.192	0.158	0.13	0.111	0.092	0.08	0.064	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.47	0.48	0.49
2	$\bar{x}$	0.5	0.48	0.49	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	σ	0.5	0.355	0.284	0.253	0.221	0.185	0.157	0.129	0.114	0.092	0.079	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.48	0.49
3	$\bar{x}$	0.48	0.49	0.49	0.49	0.49	0.5	0.5	0.51	0.51	0.53	0.54	0.57	0.58
	σ	0.5	0.356	0.288	0.247	0.224	0.19	0.16	0.131	0.114	0.093	0.078	0.067	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.54	0.57	0.57
4	$\bar{x}$	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	σ	0.5	0.352	0.289	0.252	0.225	0.191	0.157	0.131	0.111	0.093	0.079	0.066	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.48	0.49
5	$\bar{x}$	0.48	0.49	0.49	0.49	0.49	0.49	0.5	0.5	0.5	0.51	0.51	0.53	0.54
	σ	0.5	0.35	0.288	0.252	0.228	0.19	0.158	0.133	0.114	0.094	0.083	0.069	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.54
6	$\bar{x}$	0.5	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.5
	σ	0.5	0.355	0.287	0.248	0.226	0.189	0.156	0.13	0.111	0.092	0.078	0.066	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
7	$\bar{x}$	0.49	0.49	0.51	0.51	0.51	0.53	0.54	0.55	0.56	0.58	0.59	0.62	0.64
	σ	0.5	0.352	0.29	0.249	0.225	0.191	0.157	0.127	0.11	0.088	0.076	0.059	0.053
	m	0	0.5	0.67	0.5	0.6	0.57	0.5	0.53	0.55	0.6	0.6	0.62	0.64
8	$\bar{x}$	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.49
	σ	0.5	0.354	0.291	0.25	0.222	0.189	0.159	0.13	0.113	0.093	0.081	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
9	$\bar{x}$	0.49	0.49	0.5	0.5	0.51	0.51	0.52	0.53	0.54	0.55	0.56	0.58	0.59
	σ	0.5	0.36	0.292	0.257	0.233	0.195	0.163	0.136	0.117	0.098	0.086	0.07	0.062
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.57	0.58	0.59
10	$\bar{x}$	0.48	0.48	0.48	0.49	0.49	0.49	0.49	0.48	0.48	0.49	0.49	0.49	0.49
	σ	0.5	0.354	0.289	0.248	0.224	0.19	0.158	0.129	0.114	0.092	0.079	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.47	0.48	0.49
11	$\bar{x}$	0.48	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.48	0.49	0.49	0.48	0.49
	σ	0.5	0.365	0.299	0.26	0.239	0.206	0.175	0.147	0.133	0.111	0.098	0.088	0.078
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.48	0.49
12	$\bar{x}$	0.49	0.48	0.49	0.48	0.49	0.49	0.48	0.49	0.49	0.48	0.49	0.49	0.49
	σ	0.5	0.356	0.289	0.249	0.223	0.188	0.161	0.129	0.111	0.093	0.077	0.067	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.47	0.48	0.49
13	$\bar{x}$	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.5	0.5
	σ	0.5	0.355	0.287	0.252	0.222	0.191	0.158	0.127	0.112	0.092	0.078	0.062	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
14	$\bar{x}$	0.49	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.5	0.5
	σ	0.5	0.355	0.289	0.25	0.222	0.189	0.158	0.13	0.111	0.088	0.08	0.066	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.5
15	$\bar{x}$	0.5	0.5	0.49	0.52	0.51	0.52	0.54	0.55	0.56	0.58	0.6	0.62	0.64
	σ	0.5	0.356	0.295	0.25	0.226	0.191	0.159	0.128	0.109	0.09	0.077	0.065	0.052
	m	0	0.5	0.33	0.5	0.6	0.57	0.5	0.53	0.55	0.57	0.6	0.62	0.64

D.6 Parabolic Ridge Function

Table D.31: Local update for 2-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.41	0.4	0.4	0.4	0.4	0.39	0.4	0.4	0.4	0.41	0.4	0.4	0.4
	σ	0.491	0.362	0.303	0.273	0.253	0.227	0.201	0.182	0.173	0.162	0.156	0.155	0.149
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.45	0.45
2	$\bar{x}$	0.39	0.42	0.44	0.45	0.46	0.48	0.51	0.55	0.57	0.6	0.62	0.64	0.67
	σ	0.487	0.365	0.317	0.291	0.278	0.262	0.258	0.258	0.259	0.275	0.28	0.291	0.292
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.75	0.78	0.82
3	$\bar{x}$	0.4	0.43	0.46	0.48	0.5	0.52	0.55	0.58	0.61	0.62	0.65	0.67	0.68
	σ	0.49	0.364	0.319	0.299	0.286	0.279	0.274	0.272	0.271	0.284	0.287	0.292	0.298
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.8	0.82	0.84
4	$\bar{x}$	0.4	0.41	0.41	0.42	0.44	0.45	0.47	0.51	0.53	0.58	0.59	0.62	0.64
	σ	0.49	0.37	0.317	0.293	0.279	0.26	0.249	0.247	0.252	0.252	0.26	0.271	0.278
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.67	0.68	0.73	0.76
5	$\bar{x}$	0.39	0.42	0.43	0.44	0.46	0.48	0.51	0.55	0.56	0.6	0.62	0.66	0.65
	σ	0.489	0.377	0.33	0.308	0.293	0.276	0.268	0.261	0.263	0.263	0.268	0.261	0.277
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.6	0.67	0.72	0.77	0.78
6	$\bar{x}$	0.41	0.43	0.45	0.47	0.5	0.53	0.55	0.59	0.61	0.65	0.66	0.68	0.68
	σ	0.493	0.369	0.318	0.297	0.283	0.269	0.267	0.271	0.272	0.278	0.285	0.287	0.298
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
7	$\bar{x}$	0.39	0.45	0.48	0.51	0.53	0.56	0.59	0.61	0.63	0.66	0.67	0.7	0.71
	σ	0.489	0.369	0.324	0.307	0.294	0.286	0.284	0.282	0.292	0.29	0.289	0.29	0.296
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.87
8	$\bar{x}$	0.4	0.41	0.43	0.44	0.45	0.48	0.51	0.54	0.58	0.61	0.64	0.64	0.66
	σ	0.491	0.377	0.33	0.302	0.293	0.274	0.266	0.261	0.26	0.267	0.26	0.283	0.28
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.7	0.72	0.75	0.79
9	$\bar{x}$	0.41	0.42	0.45	0.47	0.48	0.51	0.54	0.58	0.59	0.62	0.64	0.66	0.67
	σ	0.492	0.379	0.339	0.32	0.305	0.289	0.279	0.267	0.272	0.272	0.273	0.282	0.283
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.65	0.73	0.75	0.78	0.8
10	$\bar{x}$	0.41	0.42	0.44	0.46	0.47	0.5	0.53	0.57	0.59	0.62	0.64	0.67	0.68
	σ	0.492	0.384	0.343	0.319	0.309	0.293	0.283	0.274	0.28	0.273	0.274	0.271	0.274
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.72	0.77	0.79
11	$\bar{x}$	0.39	0.42	0.45	0.47	0.49	0.51	0.53	0.56	0.59	0.61	0.63	0.65	0.66
	σ	0.489	0.393	0.356	0.333	0.322	0.305	0.292	0.284	0.283	0.283	0.282	0.279	0.292
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.72	0.73	0.78
12	$\bar{x}$	0.4	0.42	0.45	0.47	0.5	0.52	0.55	0.59	0.62	0.64	0.66	0.69	0.71
	σ	0.489	0.37	0.324	0.302	0.287	0.279	0.268	0.272	0.271	0.28	0.286	0.283	0.285
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
13	$\bar{x}$	0.39	0.45	0.48	0.51	0.53	0.56	0.6	0.63	0.65	0.66	0.69	0.69	0.73
	σ	0.488	0.374	0.327	0.312	0.299	0.285	0.282	0.279	0.281	0.288	0.291	0.299	0.285
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.88
14	$\bar{x}$	0.39	0.43	0.46	0.48	0.49	0.53	0.57	0.59	0.61	0.65	0.66	0.7	0.71
	σ	0.488	0.383	0.336	0.321	0.302	0.287	0.28	0.28	0.281	0.28	0.289	0.286	0.279
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
15	$\bar{x}$	0.41	0.44	0.47	0.5	0.52	0.56	0.58	0.61	0.63	0.66	0.67	0.69	0.69
	σ	0.492	0.388	0.345	0.325	0.31	0.297	0.289	0.289	0.287	0.287	0.29	0.286	0.304
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.75	0.77	0.8	0.82	0.85

Table D.32: Local update for 5-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.38	0.39	0.39	0.38	0.39	0.39	0.38	0.39	0.39	0.38	0.39	0.38	0.39
	σ	0.487	0.347	0.286	0.247	0.221	0.187	0.158	0.131	0.112	0.097	0.084	0.074	0.067
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.38	0.39
2	$\bar{x}$	0.38	0.39	0.39	0.39	0.41	0.41	0.42	0.44	0.47	0.49	0.52	0.56	0.59
	σ	0.487	0.348	0.286	0.248	0.223	0.196	0.163	0.139	0.125	0.11	0.098	0.088	0.086
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
3	$\bar{x}$	0.38	0.41	0.42	0.44	0.46	0.48	0.51	0.55	0.58	0.62	0.64	0.67	0.7
	σ	0.486	0.353	0.29	0.258	0.231	0.204	0.177	0.15	0.131	0.113	0.103	0.093	0.088
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.65	0.68	0.71
4	$\bar{x}$	0.38	0.39	0.39	0.39	0.39	0.4	0.41	0.42	0.43	0.46	0.48	0.52	0.55
	σ	0.485	0.354	0.286	0.249	0.228	0.195	0.167	0.14	0.132	0.114	0.105	0.095	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.53	0.56
5	$\bar{x}$	0.38	0.39	0.41	0.41	0.43	0.45	0.47	0.5	0.53	0.57	0.61	0.65	0.66
	σ	0.486	0.36	0.299	0.267	0.247	0.219	0.19	0.166	0.15	0.129	0.113	0.096	0.097
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.63	0.65	0.68
6	$\bar{x}$	0.37	0.4	0.4	0.41	0.43	0.44	0.47	0.5	0.53	0.57	0.6	0.64	0.67
	σ	0.483	0.35	0.289	0.253	0.231	0.197	0.17	0.145	0.131	0.112	0.102	0.088	0.089
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.68
7	$\bar{x}$	0.39	0.42	0.45	0.48	0.49	0.52	0.56	0.6	0.62	0.66	0.68	0.71	0.72
	σ	0.489	0.354	0.297	0.263	0.238	0.203	0.172	0.144	0.13	0.11	0.102	0.087	0.082
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.6	0.67	0.7	0.72	0.72
8	$\bar{x}$	0.39	0.4	0.4	0.4	0.4	0.42	0.43	0.46	0.48	0.52	0.55	0.6	0.63
	σ	0.489	0.353	0.29	0.256	0.238	0.204	0.179	0.157	0.143	0.125	0.114	0.097	0.091
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.62	0.64
9	$\bar{x}$	0.39	0.4	0.42	0.45	0.46	0.48	0.51	0.55	0.57	0.61	0.63	0.66	0.68
	σ	0.487	0.367	0.311	0.278	0.262	0.225	0.196	0.166	0.147	0.124	0.111	0.096	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.6	0.65	0.67	0.69
10	$\bar{x}$	0.38	0.39	0.39	0.4	0.4	0.41	0.42	0.44	0.45	0.48	0.5	0.53	0.56
	σ	0.486	0.353	0.294	0.261	0.241	0.218	0.192	0.172	0.166	0.149	0.139	0.126	0.122
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.5	0.53	0.57
11	$\bar{x}$	0.39	0.4	0.4	0.41	0.41	0.43	0.44	0.46	0.47	0.5	0.51	0.53	0.55
	σ	0.487	0.371	0.323	0.292	0.276	0.254	0.23	0.21	0.193	0.173	0.164	0.156	0.145
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.53	0.55	0.57
12	$\bar{x}$	0.38	0.39	0.4	0.4	0.41	0.41	0.42	0.44	0.46	0.49	0.51	0.56	0.58
	σ	0.486	0.347	0.282	0.249	0.222	0.191	0.163	0.136	0.125	0.105	0.093	0.079	0.076
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.59
13	$\bar{x}$	0.38	0.41	0.43	0.44	0.45	0.48	0.51	0.56	0.58	0.63	0.65	0.69	0.7
	σ	0.486	0.352	0.29	0.256	0.228	0.203	0.175	0.144	0.124	0.107	0.098	0.082	0.083
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.68	0.7	0.71
14	$\bar{x}$	0.39	0.39	0.41	0.42	0.43	0.44	0.47	0.51	0.53	0.58	0.61	0.65	0.67
	σ	0.487	0.35	0.286	0.264	0.234	0.205	0.18	0.155	0.141	0.117	0.105	0.093	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.63	0.66	0.69
15	$\bar{x}$	0.39	0.42	0.45	0.47	0.49	0.52	0.55	0.6	0.62	0.66	0.68	0.71	0.72
	σ	0.487	0.364	0.301	0.274	0.246	0.21	0.182	0.152	0.139	0.111	0.11	0.089	0.085
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.7	0.72	0.72

Table D.33: Local update for 10-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.4	0.41	0.41	0.41	0.4	0.41	0.4	0.41	0.4	0.4	0.4	0.41	0.4
	σ	0.491	0.35	0.277	0.249	0.221	0.186	0.155	0.128	0.108	0.092	0.081	0.063	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
2	$\bar{x}$	0.4	0.42	0.4	0.41	0.4	0.41	0.42	0.42	0.43	0.44	0.45	0.47	0.49
	σ	0.489	0.352	0.286	0.246	0.223	0.187	0.159	0.128	0.111	0.096	0.08	0.069	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.47	0.49
3	$\bar{x}$	0.4	0.42	0.43	0.44	0.45	0.46	0.49	0.52	0.55	0.58	0.61	0.65	0.68
	σ	0.491	0.348	0.289	0.253	0.225	0.192	0.165	0.134	0.118	0.099	0.086	0.07	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.63	0.65	0.69
4	$\bar{x}$	0.4	0.41	0.4	0.41	0.41	0.41	0.41	0.41	0.42	0.42	0.42	0.44	0.45
	σ	0.489	0.348	0.288	0.245	0.224	0.192	0.16	0.13	0.115	0.093	0.086	0.075	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
5	$\bar{x}$	0.4	0.41	0.42	0.42	0.43	0.44	0.46	0.48	0.5	0.53	0.56	0.6	0.63
	σ	0.491	0.351	0.292	0.26	0.231	0.2	0.174	0.148	0.133	0.113	0.1	0.08	0.071
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.57	0.6	0.63
6	$\bar{x}$	0.4	0.41	0.42	0.41	0.42	0.43	0.44	0.45	0.47	0.5	0.53	0.57	0.6
	σ	0.49	0.344	0.289	0.245	0.223	0.19	0.156	0.13	0.117	0.096	0.087	0.069	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.61
7	$\bar{x}$	0.41	0.44	0.46	0.47	0.48	0.51	0.54	0.57	0.6	0.64	0.66	0.69	0.71
	σ	0.491	0.354	0.29	0.255	0.227	0.197	0.165	0.133	0.114	0.093	0.078	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.68	0.68	0.71
8	$\bar{x}$	0.41	0.41	0.41	0.42	0.41	0.42	0.42	0.43	0.45	0.47	0.48	0.51	0.55
	σ	0.491	0.351	0.289	0.249	0.226	0.196	0.163	0.138	0.124	0.105	0.094	0.083	0.073
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
9	$\bar{x}$	0.4	0.42	0.43	0.45	0.45	0.47	0.5	0.52	0.54	0.58	0.6	0.64	0.65
	σ	0.49	0.362	0.303	0.271	0.247	0.213	0.184	0.153	0.136	0.112	0.097	0.077	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.65
10	$\bar{x}$	0.4	0.4	0.41	0.41	0.4	0.41	0.41	0.42	0.42	0.42	0.43	0.43	0.44
	σ	0.49	0.349	0.292	0.253	0.227	0.196	0.164	0.141	0.131	0.114	0.109	0.098	0.099
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
11	$\bar{x}$	0.41	0.4	0.4	0.42	0.41	0.41	0.41	0.42	0.43	0.43	0.44	0.45	0.46
	σ	0.491	0.364	0.314	0.283	0.259	0.231	0.208	0.184	0.17	0.157	0.145	0.135	0.124
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.45	0.45
12	$\bar{x}$	0.39	0.4	0.41	0.41	0.41	0.42	0.41	0.41	0.41	0.42	0.43	0.44	0.45
	σ	0.488	0.349	0.285	0.245	0.221	0.187	0.154	0.129	0.11	0.093	0.08	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.45
13	$\bar{x}$	0.4	0.41	0.41	0.42	0.42	0.43	0.45	0.47	0.49	0.53	0.56	0.6	0.64
	σ	0.491	0.347	0.282	0.249	0.219	0.19	0.158	0.132	0.116	0.097	0.085	0.07	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.6	0.64
14	$\bar{x}$	0.4	0.42	0.41	0.41	0.42	0.43	0.44	0.45	0.47	0.5	0.53	0.57	0.61
	σ	0.491	0.348	0.285	0.25	0.221	0.19	0.161	0.134	0.118	0.1	0.089	0.074	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.53	0.57	0.61
15	$\bar{x}$	0.42	0.44	0.46	0.46	0.49	0.51	0.54	0.57	0.6	0.63	0.66	0.69	0.71
	σ	0.494	0.358	0.294	0.254	0.23	0.199	0.168	0.134	0.116	0.093	0.081	0.065	0.054
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.7	0.71

Table D.34: Local update for 20-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.44	0.42	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.496	0.348	0.282	0.249	0.221	0.19	0.158	0.128	0.11	0.089	0.078	0.066	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
2	$\bar{x}$	0.43	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.45
	σ	0.495	0.351	0.288	0.246	0.223	0.186	0.154	0.129	0.112	0.092	0.08	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.45	0.45
3	$\bar{x}$	0.42	0.44	0.44	0.45	0.45	0.46	0.47	0.5	0.51	0.55	0.57	0.61	0.64
	σ	0.493	0.351	0.287	0.254	0.228	0.195	0.161	0.133	0.116	0.096	0.08	0.067	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.57	0.62	0.64
4	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44
	σ	0.494	0.351	0.284	0.245	0.22	0.183	0.158	0.131	0.112	0.091	0.08	0.062	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.44
5	$\bar{x}$	0.43	0.43	0.43	0.43	0.45	0.45	0.45	0.47	0.49	0.51	0.53	0.56	0.58
	σ	0.495	0.355	0.289	0.255	0.228	0.195	0.169	0.14	0.127	0.103	0.094	0.074	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.59
6	$\bar{x}$	0.42	0.42	0.44	0.43	0.43	0.44	0.44	0.45	0.45	0.47	0.47	0.5	0.52
	σ	0.494	0.351	0.288	0.253	0.223	0.187	0.156	0.133	0.111	0.095	0.081	0.068	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
7	$\bar{x}$	0.43	0.45	0.46	0.47	0.48	0.5	0.52	0.55	0.57	0.6	0.63	0.66	0.69
	σ	0.495	0.349	0.286	0.255	0.224	0.194	0.161	0.132	0.114	0.092	0.079	0.063	0.05
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.65	0.69
8	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.45	0.47	0.48
	σ	0.495	0.351	0.289	0.245	0.224	0.188	0.159	0.131	0.113	0.094	0.083	0.07	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.47	0.49
9	$\bar{x}$	0.44	0.43	0.45	0.45	0.46	0.47	0.49	0.51	0.53	0.56	0.57	0.6	0.62
	σ	0.496	0.355	0.297	0.26	0.237	0.206	0.174	0.146	0.128	0.106	0.092	0.073	0.067
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.63
10	$\bar{x}$	0.42	0.43	0.43	0.43	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.494	0.353	0.284	0.251	0.223	0.194	0.16	0.132	0.118	0.098	0.088	0.076	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.42	0.42
11	$\bar{x}$	0.42	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.44	0.44	0.44
	σ	0.494	0.366	0.307	0.275	0.252	0.221	0.192	0.165	0.149	0.135	0.122	0.114	0.108
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.43	0.44
12	$\bar{x}$	0.44	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.496	0.351	0.29	0.247	0.221	0.19	0.158	0.126	0.111	0.09	0.08	0.063	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.42
13	$\bar{x}$	0.42	0.43	0.43	0.43	0.44	0.43	0.44	0.45	0.45	0.46	0.48	0.5	0.52
	σ	0.494	0.349	0.286	0.243	0.219	0.191	0.158	0.129	0.114	0.091	0.08	0.066	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
14	$\bar{x}$	0.42	0.43	0.42	0.43	0.43	0.43	0.44	0.45	0.45	0.47	0.48	0.51	0.52
	σ	0.494	0.346	0.285	0.244	0.219	0.191	0.158	0.131	0.112	0.092	0.081	0.067	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
15	$\bar{x}$	0.42	0.46	0.47	0.46	0.49	0.5	0.52	0.55	0.57	0.6	0.63	0.66	0.68
	σ	0.494	0.353	0.29	0.255	0.227	0.193	0.16	0.131	0.113	0.093	0.08	0.064	0.054
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69

Table D.35: Local update for 40-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45
	σ	0.497	0.351	0.291	0.249	0.22	0.184	0.161	0.129	0.112	0.09	0.08	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.43	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.495	0.352	0.287	0.254	0.223	0.186	0.156	0.126	0.112	0.089	0.079	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.45	0.45
3	$\bar{x}$	0.46	0.45	0.45	0.45	0.46	0.46	0.48	0.49	0.5	0.52	0.54	0.57	0.59
	σ	0.498	0.352	0.289	0.251	0.222	0.189	0.158	0.13	0.112	0.094	0.08	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.55	0.57	0.6
4	$\bar{x}$	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.497	0.35	0.292	0.248	0.223	0.19	0.158	0.128	0.109	0.09	0.081	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
5	$\bar{x}$	0.45	0.45	0.45	0.45	0.44	0.46	0.46	0.47	0.47	0.49	0.5	0.52	0.55
	σ	0.498	0.355	0.293	0.251	0.226	0.193	0.162	0.133	0.119	0.1	0.086	0.076	0.066
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.5	0.5	0.52	0.55
6	$\bar{x}$	0.45	0.45	0.46	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.47	0.48	0.48
	σ	0.498	0.352	0.286	0.248	0.223	0.188	0.156	0.131	0.113	0.092	0.081	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.49
7	$\bar{x}$	0.45	0.46	0.47	0.49	0.48	0.5	0.51	0.53	0.55	0.58	0.6	0.63	0.65
	σ	0.498	0.353	0.291	0.251	0.227	0.191	0.16	0.128	0.11	0.09	0.077	0.063	0.053
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.62	0.65
8	$\bar{x}$	0.44	0.44	0.44	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.46	0.47
	σ	0.496	0.349	0.284	0.251	0.225	0.191	0.158	0.13	0.111	0.092	0.079	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.47
9	$\bar{x}$	0.44	0.45	0.46	0.46	0.47	0.48	0.49	0.51	0.52	0.54	0.55	0.57	0.6
	σ	0.497	0.354	0.297	0.257	0.234	0.2	0.17	0.138	0.127	0.103	0.09	0.073	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.58	0.6
10	$\bar{x}$	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.355	0.287	0.247	0.227	0.188	0.159	0.13	0.114	0.092	0.08	0.068	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
11	$\bar{x}$	0.44	0.44	0.44	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.44	0.45
	σ	0.497	0.363	0.297	0.265	0.242	0.209	0.181	0.155	0.14	0.119	0.111	0.096	0.089
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
12	$\bar{x}$	0.44	0.45	0.45	0.44	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.497	0.35	0.287	0.249	0.222	0.187	0.157	0.129	0.111	0.091	0.078	0.065	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
13	$\bar{x}$	0.45	0.44	0.45	0.45	0.44	0.45	0.45	0.45	0.46	0.46	0.46	0.47	0.47
	σ	0.497	0.353	0.288	0.245	0.221	0.185	0.159	0.125	0.11	0.091	0.08	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.47	0.47
14	$\bar{x}$	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.47	0.48	0.49
	σ	0.498	0.35	0.287	0.252	0.22	0.188	0.156	0.133	0.111	0.092	0.078	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.47	0.49
15	$\bar{x}$	0.45	0.47	0.47	0.48	0.49	0.5	0.51	0.54	0.55	0.58	0.6	0.63	0.65
	σ	0.497	0.35	0.29	0.253	0.223	0.192	0.159	0.131	0.113	0.089	0.075	0.063	0.053
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65

Table D.36: Local update for 80-dimensional parabolic ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.352	0.29	0.249	0.228	0.192	0.158	0.128	0.112	0.09	0.079	0.063	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.45	0.46
2	$\bar{x}$	0.46	0.47	0.47	0.45	0.47	0.47	0.46	0.46	0.47	0.46	0.46	0.46	0.47
	σ	0.499	0.357	0.292	0.245	0.224	0.19	0.157	0.13	0.111	0.09	0.08	0.063	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
3	$\bar{x}$	0.46	0.47	0.46	0.47	0.47	0.47	0.47	0.48	0.49	0.5	0.52	0.53	0.55
	σ	0.498	0.355	0.288	0.249	0.225	0.191	0.157	0.125	0.111	0.094	0.083	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
4	$\bar{x}$	0.47	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.499	0.355	0.291	0.252	0.223	0.187	0.158	0.129	0.11	0.089	0.079	0.068	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
5	$\bar{x}$	0.46	0.46	0.46	0.46	0.47	0.47	0.47	0.47	0.47	0.49	0.5	0.5	0.52
	σ	0.499	0.355	0.289	0.251	0.225	0.192	0.158	0.133	0.115	0.097	0.085	0.071	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.5	0.5	0.5	0.53
6	$\bar{x}$	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.47	0.47	0.47
	σ	0.499	0.355	0.284	0.248	0.223	0.185	0.16	0.128	0.113	0.094	0.078	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
7	$\bar{x}$	0.47	0.48	0.48	0.48	0.49	0.5	0.51	0.53	0.54	0.56	0.57	0.6	0.61
	σ	0.499	0.352	0.288	0.247	0.223	0.19	0.161	0.132	0.111	0.092	0.08	0.061	0.053
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.61
8	$\bar{x}$	0.46	0.46	0.47	0.47	0.47	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.47
	σ	0.498	0.352	0.291	0.25	0.223	0.19	0.158	0.127	0.113	0.093	0.08	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
9	$\bar{x}$	0.46	0.47	0.47	0.48	0.48	0.48	0.49	0.5	0.51	0.53	0.54	0.55	0.57
	σ	0.499	0.355	0.293	0.255	0.231	0.197	0.168	0.138	0.121	0.1	0.085	0.071	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.55	0.56
10	$\bar{x}$	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.46	0.46	0.46	0.46
	σ	0.499	0.354	0.288	0.247	0.223	0.187	0.157	0.128	0.111	0.092	0.079	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
11	$\bar{x}$	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.499	0.361	0.298	0.264	0.234	0.204	0.177	0.145	0.13	0.11	0.099	0.088	0.082
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
12	$\bar{x}$	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.499	0.354	0.287	0.246	0.223	0.189	0.157	0.129	0.11	0.092	0.078	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
13	$\bar{x}$	0.47	0.47	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.47	0.47
	σ	0.499	0.352	0.285	0.249	0.22	0.186	0.161	0.128	0.113	0.09	0.08	0.061	0.053
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
14	$\bar{x}$	0.46	0.45	0.46	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.47	0.47	0.48
	σ	0.498	0.356	0.289	0.249	0.221	0.188	0.158	0.128	0.111	0.091	0.077	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
15	$\bar{x}$	0.46	0.49	0.48	0.49	0.49	0.5	0.51	0.53	0.54	0.56	0.57	0.6	0.61
	σ	0.499	0.354	0.288	0.252	0.226	0.191	0.16	0.129	0.108	0.089	0.079	0.065	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.61

D.7 Sharp Ridge Function

Table D.37: Local update for 2-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.4	0.39	0.4	0.4	0.41	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
	σ	0.49	0.361	0.3	0.272	0.249	0.224	0.204	0.183	0.177	0.164	0.157	0.147	0.155
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.43	0.45
2	$\bar{x}$	0.38	0.42	0.44	0.45	0.46	0.48	0.52	0.54	0.56	0.59	0.61	0.64	0.66
	σ	0.486	0.369	0.316	0.295	0.274	0.263	0.258	0.259	0.264	0.278	0.277	0.295	0.299
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.7	0.75	0.8	0.81
3	$\bar{x}$	0.4	0.42	0.46	0.47	0.49	0.53	0.55	0.58	0.6	0.62	0.65	0.66	0.67
	σ	0.489	0.368	0.324	0.304	0.289	0.278	0.271	0.276	0.278	0.282	0.29	0.303	0.307
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.73	0.78	0.82	0.84
4	$\bar{x}$	0.4	0.41	0.41	0.43	0.44	0.45	0.47	0.5	0.55	0.57	0.59	0.63	0.64
	σ	0.489	0.37	0.317	0.295	0.279	0.259	0.251	0.252	0.246	0.257	0.267	0.275	0.274
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.68	0.75	0.76
5	$\bar{x}$	0.39	0.42	0.44	0.44	0.46	0.48	0.51	0.54	0.56	0.59	0.61	0.63	0.64
	σ	0.489	0.375	0.33	0.309	0.294	0.279	0.265	0.265	0.268	0.273	0.278	0.284	0.295
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.67	0.72	0.75	0.78
6	$\bar{x}$	0.41	0.43	0.46	0.48	0.48	0.52	0.55	0.59	0.6	0.63	0.65	0.69	0.69
	σ	0.491	0.369	0.322	0.297	0.285	0.27	0.271	0.27	0.276	0.285	0.289	0.29	0.299
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.7	0.77	0.8	0.83	0.85
7	$\bar{x}$	0.41	0.44	0.48	0.5	0.53	0.55	0.58	0.62	0.64	0.66	0.68	0.71	0.71
	σ	0.492	0.369	0.326	0.305	0.295	0.284	0.283	0.279	0.285	0.284	0.289	0.292	0.298
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.86
8	$\bar{x}$	0.39	0.41	0.43	0.45	0.46	0.48	0.51	0.54	0.58	0.6	0.63	0.66	0.67
	σ	0.488	0.372	0.328	0.309	0.289	0.275	0.266	0.26	0.261	0.273	0.267	0.277	0.272
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.6	0.65	0.7	0.72	0.78	0.79
9	$\bar{x}$	0.4	0.42	0.45	0.47	0.48	0.52	0.54	0.57	0.6	0.63	0.63	0.68	0.67
	σ	0.49	0.384	0.339	0.317	0.305	0.285	0.279	0.274	0.274	0.27	0.281	0.277	0.287
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.7	0.73	0.75	0.8	0.8
10	$\bar{x}$	0.41	0.42	0.43	0.44	0.46	0.47	0.49	0.52	0.53	0.57	0.59	0.61	0.63
	σ	0.492	0.386	0.342	0.319	0.313	0.295	0.29	0.284	0.283	0.279	0.285	0.281	0.287
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.63	0.68	0.7	0.72
11	$\bar{x}$	0.41	0.43	0.43	0.44	0.45	0.48	0.5	0.53	0.55	0.55	0.57	0.6	0.61
	σ	0.492	0.396	0.355	0.338	0.329	0.314	0.303	0.295	0.292	0.286	0.29	0.285	0.284
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.6	0.63	0.67	0.69
12	$\bar{x}$	0.4	0.43	0.43	0.46	0.47	0.5	0.53	0.57	0.6	0.62	0.65	0.66	0.71
	σ	0.49	0.364	0.317	0.296	0.277	0.265	0.262	0.257	0.265	0.276	0.28	0.299	0.284
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.67	0.7	0.73	0.78	0.82	0.85
13	$\bar{x}$	0.4	0.44	0.46	0.49	0.51	0.53	0.57	0.6	0.63	0.65	0.68	0.7	0.7
	σ	0.49	0.369	0.324	0.302	0.287	0.283	0.274	0.277	0.286	0.293	0.283	0.294	0.301
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.75	0.77	0.8	0.85	0.86
14	$\bar{x}$	0.41	0.43	0.46	0.48	0.5	0.53	0.56	0.6	0.63	0.65	0.67	0.69	0.7
	σ	0.491	0.377	0.329	0.312	0.3	0.287	0.282	0.282	0.284	0.288	0.289	0.293	0.288
	m	0	0.5	0.33	0.5	0.6	0.57	0.6	0.67	0.75	0.77	0.8	0.83	0.85
15	$\bar{x}$	0.41	0.45	0.48	0.51	0.53	0.56	0.59	0.62	0.64	0.66	0.68	0.71	0.72
	σ	0.491	0.385	0.339	0.317	0.303	0.295	0.288	0.288	0.292	0.289	0.298	0.291	0.293
	m	0	0.5	0.33	0.5	0.6	0.57	0.7	0.73	0.75	0.8	0.82	0.85	0.88

Table D.38: Local update for 5-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.38	0.39
	σ	0.487	0.348	0.282	0.248	0.222	0.187	0.158	0.131	0.114	0.097	0.086	0.074	0.067
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.38	0.39
2	$\bar{x}$	0.38	0.4	0.39	0.4	0.41	0.41	0.42	0.44	0.46	0.5	0.52	0.56	0.6
	σ	0.486	0.346	0.29	0.254	0.224	0.193	0.165	0.137	0.129	0.106	0.102	0.093	0.085
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.61
3	$\bar{x}$	0.39	0.39	0.43	0.43	0.46	0.48	0.51	0.55	0.58	0.62	0.65	0.68	0.7
	σ	0.488	0.35	0.291	0.261	0.234	0.202	0.173	0.149	0.135	0.117	0.104	0.091	0.091
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.63	0.65	0.68	0.71
4	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.4	0.41	0.42	0.43	0.45	0.48	0.52	0.55
	σ	0.489	0.348	0.284	0.249	0.231	0.195	0.169	0.141	0.128	0.115	0.104	0.098	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.47	0.47	0.52	0.55
5	$\bar{x}$	0.39	0.39	0.41	0.42	0.42	0.44	0.46	0.5	0.54	0.57	0.61	0.64	0.66
	σ	0.488	0.353	0.298	0.263	0.247	0.22	0.191	0.165	0.147	0.128	0.112	0.107	0.093
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.68
6	$\bar{x}$	0.39	0.4	0.4	0.42	0.42	0.44	0.47	0.5	0.53	0.57	0.6	0.64	0.66
	σ	0.489	0.347	0.285	0.252	0.23	0.196	0.17	0.144	0.134	0.114	0.102	0.094	0.089
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.68
7	$\bar{x}$	0.38	0.42	0.45	0.47	0.49	0.53	0.56	0.6	0.62	0.66	0.68	0.7	0.72
	σ	0.485	0.354	0.296	0.266	0.237	0.206	0.172	0.147	0.125	0.112	0.096	0.099	0.088
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.7	0.72	0.72
8	$\bar{x}$	0.4	0.39	0.4	0.4	0.41	0.42	0.43	0.46	0.48	0.53	0.55	0.59	0.63
	σ	0.49	0.349	0.291	0.259	0.233	0.203	0.176	0.157	0.143	0.123	0.115	0.097	0.09
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.6	0.64
9	$\bar{x}$	0.39	0.4	0.42	0.44	0.45	0.48	0.51	0.55	0.58	0.6	0.63	0.67	0.68
	σ	0.489	0.364	0.309	0.28	0.262	0.226	0.197	0.168	0.148	0.127	0.114	0.094	0.091
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.6	0.6	0.65	0.68	0.69
10	$\bar{x}$	0.4	0.39	0.4	0.4	0.4	0.4	0.4	0.41	0.43	0.44	0.45	0.47	0.49
	σ	0.489	0.348	0.291	0.264	0.245	0.215	0.191	0.175	0.164	0.152	0.153	0.146	0.146
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.43	0.43	0.45	0.48	0.49
11	$\bar{x}$	0.39	0.39	0.39	0.41	0.4	0.41	0.41	0.42	0.43	0.44	0.45	0.48	0.49
	σ	0.487	0.372	0.323	0.298	0.277	0.255	0.229	0.215	0.203	0.188	0.182	0.173	0.16
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.48	0.5
12	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.4	0.4	0.42	0.43	0.45	0.46	0.5	0.52
	σ	0.488	0.348	0.279	0.247	0.223	0.188	0.16	0.135	0.118	0.1	0.091	0.077	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.46	0.5	0.53
13	$\bar{x}$	0.38	0.39	0.4	0.41	0.42	0.44	0.47	0.5	0.53	0.57	0.6	0.65	0.66
	σ	0.485	0.35	0.288	0.251	0.229	0.197	0.169	0.14	0.126	0.108	0.098	0.085	0.08
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.67
14	$\bar{x}$	0.38	0.39	0.41	0.42	0.43	0.44	0.46	0.5	0.53	0.58	0.61	0.64	0.67
	σ	0.486	0.35	0.291	0.257	0.233	0.203	0.173	0.146	0.13	0.115	0.104	0.093	0.084
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.63	0.65	0.69
15	$\bar{x}$	0.39	0.42	0.44	0.48	0.49	0.52	0.56	0.6	0.63	0.66	0.68	0.7	0.72
	σ	0.489	0.357	0.302	0.266	0.239	0.211	0.176	0.149	0.129	0.114	0.103	0.092	0.088
	m	0	0.5	0.33	0.5	0.4	0.57	0.6	0.6	0.65	0.67	0.7	0.72	0.74

Table D.39: Local update for 10-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.4	0.4	0.39	0.41	0.4	0.4	0.41	0.4	0.4	0.41	0.41	0.4	0.4
	σ	0.49	0.346	0.281	0.247	0.218	0.185	0.156	0.127	0.108	0.088	0.081	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
2	$\bar{x}$	0.4	0.41	0.41	0.41	0.41	0.41	0.41	0.42	0.43	0.44	0.45	0.47	0.49
	σ	0.489	0.347	0.284	0.247	0.218	0.185	0.158	0.127	0.114	0.097	0.081	0.068	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.47	0.5
3	$\bar{x}$	0.41	0.42	0.43	0.44	0.44	0.46	0.49	0.52	0.54	0.58	0.61	0.65	0.68
	σ	0.493	0.351	0.29	0.251	0.227	0.195	0.165	0.136	0.121	0.099	0.085	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.6	0.6	0.65	0.69
4	$\bar{x}$	0.41	0.4	0.41	0.41	0.41	0.41	0.41	0.42	0.42	0.43	0.43	0.44	0.46
	σ	0.492	0.348	0.281	0.244	0.22	0.191	0.16	0.13	0.117	0.097	0.084	0.071	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.46
5	$\bar{x}$	0.4	0.41	0.42	0.42	0.42	0.44	0.46	0.48	0.5	0.53	0.56	0.6	0.63
	σ	0.49	0.353	0.291	0.258	0.239	0.202	0.175	0.147	0.13	0.114	0.099	0.082	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.57	0.6	0.63
6	$\bar{x}$	0.41	0.41	0.41	0.42	0.41	0.43	0.43	0.46	0.47	0.5	0.53	0.57	0.6
	σ	0.493	0.345	0.285	0.252	0.218	0.19	0.16	0.131	0.113	0.097	0.087	0.07	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
7	$\bar{x}$	0.41	0.44	0.46	0.47	0.49	0.5	0.54	0.58	0.6	0.63	0.66	0.69	0.71
	σ	0.492	0.35	0.291	0.256	0.23	0.194	0.165	0.134	0.115	0.091	0.08	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.68	0.71
8	$\bar{x}$	0.41	0.4	0.41	0.41	0.41	0.42	0.42	0.43	0.44	0.46	0.48	0.52	0.55
	σ	0.493	0.352	0.29	0.248	0.226	0.19	0.162	0.141	0.126	0.104	0.093	0.085	0.073
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
9	$\bar{x}$	0.41	0.42	0.44	0.44	0.45	0.47	0.5	0.52	0.55	0.58	0.6	0.64	0.65
	σ	0.492	0.358	0.302	0.269	0.246	0.215	0.183	0.151	0.134	0.112	0.099	0.08	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.65	0.65
10	$\bar{x}$	0.41	0.41	0.4	0.4	0.41	0.4	0.41	0.41	0.41	0.41	0.41	0.42	0.42
	σ	0.492	0.347	0.29	0.257	0.223	0.197	0.167	0.144	0.128	0.117	0.108	0.103	0.1
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.42	0.42
11	$\bar{x}$	0.4	0.4	0.41	0.41	0.41	0.41	0.41	0.42	0.42	0.42	0.42	0.43	0.43
	σ	0.49	0.371	0.314	0.283	0.262	0.232	0.208	0.184	0.169	0.156	0.143	0.132	0.133
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
12	$\bar{x}$	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.42	0.42	0.43
	σ	0.493	0.353	0.285	0.247	0.221	0.185	0.156	0.127	0.11	0.093	0.079	0.068	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.42	0.42
13	$\bar{x}$	0.41	0.42	0.41	0.41	0.42	0.42	0.43	0.44	0.45	0.48	0.5	0.53	0.57
	σ	0.491	0.351	0.283	0.243	0.221	0.19	0.156	0.129	0.113	0.095	0.083	0.07	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.5	0.53	0.57
14	$\bar{x}$	0.4	0.41	0.41	0.42	0.42	0.42	0.44	0.46	0.47	0.5	0.53	0.57	0.6
	σ	0.49	0.347	0.287	0.25	0.221	0.194	0.164	0.133	0.118	0.098	0.086	0.073	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.53	0.57	0.6
15	$\bar{x}$	0.4	0.43	0.45	0.47	0.49	0.51	0.54	0.57	0.59	0.64	0.66	0.69	0.71
	σ	0.49	0.353	0.292	0.257	0.232	0.2	0.163	0.135	0.115	0.092	0.079	0.063	0.058
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.6	0.6	0.63	0.65	0.7	0.71

Table D.40: Local update for 20-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.42	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.494	0.35	0.286	0.246	0.221	0.186	0.158	0.13	0.111	0.089	0.08	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.42	0.42
2	$\bar{x}$	0.45	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.45
	σ	0.497	0.35	0.29	0.244	0.218	0.187	0.157	0.131	0.114	0.091	0.081	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.42	0.44	0.44	0.44	0.45	0.47	0.48	0.5	0.51	0.55	0.57	0.61	0.64
	σ	0.494	0.351	0.287	0.248	0.223	0.193	0.16	0.132	0.115	0.094	0.08	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.57	0.62	0.64
4	$\bar{x}$	0.43	0.43	0.43	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44
	σ	0.495	0.35	0.285	0.246	0.221	0.187	0.158	0.131	0.111	0.091	0.079	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.44
5	$\bar{x}$	0.42	0.42	0.43	0.44	0.44	0.45	0.45	0.47	0.48	0.51	0.53	0.57	0.59
	σ	0.494	0.355	0.29	0.253	0.229	0.197	0.167	0.138	0.125	0.104	0.094	0.081	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.57	0.59
6	$\bar{x}$	0.42	0.43	0.44	0.43	0.44	0.44	0.44	0.44	0.45	0.47	0.48	0.5	0.52
	σ	0.494	0.352	0.285	0.248	0.221	0.192	0.159	0.128	0.11	0.093	0.082	0.067	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
7	$\bar{x}$	0.44	0.44	0.46	0.48	0.49	0.5	0.53	0.55	0.57	0.6	0.63	0.66	0.68
	σ	0.496	0.351	0.286	0.253	0.227	0.191	0.163	0.131	0.112	0.092	0.075	0.062	0.053
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69
8	$\bar{x}$	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.46	0.47	0.48
	σ	0.494	0.349	0.289	0.247	0.222	0.188	0.16	0.132	0.116	0.095	0.081	0.073	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.47	0.49
9	$\bar{x}$	0.44	0.44	0.45	0.45	0.46	0.47	0.49	0.51	0.53	0.56	0.57	0.6	0.62
	σ	0.496	0.359	0.297	0.264	0.238	0.204	0.177	0.147	0.126	0.107	0.088	0.074	0.066
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.6	0.63
10	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.494	0.35	0.291	0.253	0.221	0.192	0.161	0.133	0.117	0.1	0.086	0.078	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
11	$\bar{x}$	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.495	0.36	0.31	0.272	0.249	0.218	0.193	0.167	0.152	0.138	0.123	0.113	0.104
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.42
12	$\bar{x}$	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	σ	0.493	0.35	0.287	0.246	0.219	0.188	0.152	0.13	0.11	0.09	0.078	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.43	0.42
13	$\bar{x}$	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45	0.46	0.46	0.48
	σ	0.494	0.354	0.284	0.256	0.223	0.187	0.158	0.13	0.112	0.093	0.078	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.47	0.47
14	$\bar{x}$	0.43	0.43	0.43	0.44	0.44	0.44	0.44	0.45	0.45	0.46	0.48	0.5	0.52
	σ	0.495	0.35	0.286	0.245	0.223	0.185	0.157	0.129	0.114	0.095	0.081	0.067	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.53
15	$\bar{x}$	0.43	0.45	0.46	0.47	0.49	0.5	0.53	0.55	0.57	0.61	0.63	0.66	0.69
	σ	0.495	0.356	0.293	0.249	0.227	0.195	0.162	0.13	0.115	0.091	0.079	0.062	0.053
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.6	0.63	0.67	0.69

Table D.41: Local update for 40-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.45
	σ	0.498	0.347	0.285	0.251	0.226	0.189	0.156	0.128	0.113	0.092	0.078	0.063	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
2	$\bar{x}$	0.44	0.45	0.45	0.45	0.44	0.44	0.45	0.45	0.44	0.45	0.45	0.45	0.45
	σ	0.496	0.348	0.287	0.245	0.222	0.189	0.159	0.129	0.111	0.092	0.082	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.46	0.45	0.46	0.47	0.46	0.46	0.48	0.49	0.5	0.52	0.54	0.57	0.59
	σ	0.498	0.351	0.292	0.25	0.224	0.187	0.159	0.13	0.113	0.097	0.081	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.53	0.53	0.57	0.6
4	$\bar{x}$	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.497	0.35	0.289	0.252	0.222	0.187	0.156	0.127	0.111	0.092	0.077	0.066	0.053
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.43	0.45	0.45	0.45
5	$\bar{x}$	0.44	0.45	0.45	0.45	0.46	0.46	0.46	0.47	0.47	0.49	0.5	0.53	0.55
	σ	0.496	0.353	0.29	0.254	0.223	0.195	0.163	0.136	0.12	0.1	0.09	0.075	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.5	0.5	0.53	0.55
6	$\bar{x}$	0.44	0.44	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.46	0.47	0.48	0.49
	σ	0.497	0.353	0.287	0.248	0.222	0.189	0.159	0.13	0.112	0.093	0.079	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.49
7	$\bar{x}$	0.46	0.46	0.47	0.48	0.49	0.5	0.52	0.54	0.55	0.58	0.6	0.63	0.65
	σ	0.498	0.354	0.287	0.251	0.227	0.189	0.16	0.13	0.11	0.093	0.078	0.061	0.053
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65
8	$\bar{x}$	0.44	0.45	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.46	0.45	0.46	0.47
	σ	0.496	0.361	0.289	0.249	0.222	0.188	0.159	0.13	0.112	0.091	0.082	0.071	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.45	0.46
9	$\bar{x}$	0.44	0.46	0.46	0.47	0.47	0.48	0.49	0.51	0.52	0.54	0.56	0.57	0.59
	σ	0.497	0.359	0.295	0.262	0.236	0.2	0.171	0.138	0.125	0.102	0.087	0.076	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.53	0.55	0.57	0.6
10	$\bar{x}$	0.45	0.45	0.44	0.45	0.45	0.45	0.44	0.44	0.45	0.44	0.44	0.45	0.45
	σ	0.498	0.355	0.285	0.25	0.226	0.19	0.16	0.128	0.116	0.093	0.082	0.068	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
11	$\bar{x}$	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.361	0.301	0.267	0.241	0.211	0.184	0.156	0.143	0.121	0.11	0.094	0.086
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
12	$\bar{x}$	0.46	0.44	0.44	0.45	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	σ	0.498	0.354	0.282	0.249	0.224	0.189	0.159	0.129	0.112	0.09	0.078	0.068	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.43	0.45	0.45	0.45
13	$\bar{x}$	0.46	0.45	0.44	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.46	0.46
	σ	0.498	0.352	0.29	0.248	0.224	0.19	0.156	0.13	0.113	0.09	0.077	0.064	0.054
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.45	0.45	0.46
14	$\bar{x}$	0.44	0.44	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.47	0.48	0.49
	σ	0.496	0.355	0.286	0.247	0.225	0.189	0.158	0.132	0.112	0.092	0.08	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.48	0.49
15	$\bar{x}$	0.45	0.47	0.48	0.48	0.49	0.5	0.52	0.54	0.55	0.58	0.6	0.63	0.65
	σ	0.498	0.353	0.291	0.255	0.225	0.189	0.159	0.128	0.114	0.091	0.078	0.061	0.051
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.6	0.63	0.65

Table D.42: Local update for 80-dimensional sharp ridge function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.46	0.45	0.46	0.47	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.499	0.352	0.287	0.25	0.227	0.189	0.158	0.129	0.111	0.088	0.08	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
2	$\bar{x}$	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.352	0.29	0.247	0.224	0.192	0.158	0.128	0.113	0.09	0.081	0.068	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
3	$\bar{x}$	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.48	0.49	0.5	0.51	0.54	0.56
	σ	0.499	0.356	0.291	0.253	0.222	0.189	0.161	0.131	0.113	0.092	0.08	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
4	$\bar{x}$	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.499	0.355	0.285	0.249	0.219	0.189	0.159	0.128	0.116	0.093	0.079	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
5	$\bar{x}$	0.46	0.47	0.46	0.47	0.47	0.46	0.47	0.47	0.48	0.49	0.49	0.51	0.52
	σ	0.499	0.354	0.286	0.249	0.223	0.189	0.158	0.134	0.115	0.094	0.087	0.068	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.5	0.5	0.53
6	$\bar{x}$	0.45	0.46	0.46	0.47	0.46	0.47	0.46	0.46	0.46	0.47	0.47	0.47	0.47
	σ	0.497	0.353	0.29	0.252	0.223	0.189	0.157	0.13	0.108	0.09	0.08	0.063	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
7	$\bar{x}$	0.48	0.46	0.48	0.48	0.49	0.51	0.51	0.53	0.54	0.56	0.57	0.6	0.62
	σ	0.499	0.351	0.287	0.247	0.228	0.19	0.16	0.131	0.114	0.092	0.081	0.062	0.051
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.61
8	$\bar{x}$	0.47	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.47	0.47
	σ	0.499	0.352	0.289	0.248	0.222	0.186	0.157	0.13	0.115	0.091	0.077	0.067	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
9	$\bar{x}$	0.46	0.47	0.47	0.47	0.48	0.49	0.49	0.51	0.51	0.53	0.54	0.55	0.57
	σ	0.499	0.353	0.295	0.257	0.229	0.197	0.165	0.137	0.116	0.099	0.085	0.072	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.53	0.55	0.57
10	$\bar{x}$	0.46	0.46	0.46	0.47	0.46	0.47	0.46	0.46	0.46	0.46	0.47	0.46	0.46
	σ	0.498	0.355	0.288	0.25	0.227	0.187	0.16	0.13	0.111	0.091	0.081	0.063	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
11	$\bar{x}$	0.45	0.47	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	σ	0.498	0.365	0.295	0.26	0.238	0.204	0.174	0.147	0.131	0.111	0.1	0.085	0.081
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.45	0.47	0.46
12	$\bar{x}$	0.45	0.45	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.498	0.355	0.288	0.251	0.221	0.187	0.161	0.127	0.113	0.089	0.079	0.065	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
13	$\bar{x}$	0.45	0.46	0.46	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.46
	σ	0.497	0.352	0.288	0.252	0.225	0.186	0.156	0.128	0.11	0.094	0.08	0.066	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.46
14	$\bar{x}$	0.46	0.46	0.46	0.47	0.47	0.46	0.47	0.46	0.47	0.47	0.47	0.47	0.47
	σ	0.499	0.352	0.289	0.25	0.225	0.189	0.157	0.129	0.112	0.091	0.078	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.47	0.47	0.47
15	$\bar{x}$	0.45	0.48	0.48	0.49	0.48	0.5	0.51	0.53	0.54	0.55	0.57	0.59	0.61
	σ	0.497	0.352	0.286	0.249	0.225	0.191	0.156	0.128	0.113	0.094	0.078	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.57	0.5	0.53	0.55	0.57	0.57	0.6	0.61

D.8 Generalized Rosenbrock Function

Table D.43: Local update for 2-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.32	0.31	0.31	0.32	0.32	0.31	0.31	0.32	0.31	0.31	0.31	0.31	0.31
	σ	0.466	0.354	0.313	0.288	0.268	0.246	0.235	0.219	0.21	0.206	0.202	0.197	0.196
	m	0	0	0.33	0.25	0.2	0.29	0.3	0.33	0.3	0.3	0.3	0.32	0.3
2	$\bar{x}$	0.31	0.33	0.33	0.34	0.34	0.34	0.35	0.37	0.38	0.39	0.39	0.4	0.41
	σ	0.463	0.362	0.316	0.296	0.286	0.27	0.269	0.267	0.274	0.277	0.281	0.283	0.295
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.33	0.35	0.37	0.35	0.36	0.38
3	$\bar{x}$	0.3	0.33	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.4	0.41	0.4	0.41
	σ	0.459	0.356	0.32	0.304	0.295	0.285	0.282	0.282	0.284	0.291	0.293	0.295	0.298
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.33	0.35	0.37	0.38	0.33	0.38
4	$\bar{x}$	0.31	0.31	0.33	0.32	0.32	0.33	0.34	0.36	0.37	0.38	0.4	0.4	0.42
	σ	0.463	0.357	0.319	0.294	0.279	0.269	0.258	0.259	0.253	0.265	0.275	0.28	0.285
	m	0	0	0.33	0.25	0.2	0.29	0.3	0.33	0.35	0.33	0.38	0.38	0.4
5	$\bar{x}$	0.31	0.32	0.33	0.33	0.34	0.34	0.35	0.37	0.38	0.41	0.4	0.41	0.42
	σ	0.462	0.363	0.326	0.301	0.29	0.28	0.275	0.274	0.274	0.279	0.281	0.284	0.292
	m	0	0	0.33	0.25	0.2	0.29	0.3	0.33	0.35	0.4	0.38	0.4	0.4
6	$\bar{x}$	0.32	0.33	0.33	0.35	0.35	0.36	0.38	0.39	0.41	0.42	0.44	0.43	0.43
	σ	0.465	0.361	0.323	0.308	0.295	0.289	0.287	0.286	0.288	0.293	0.3	0.306	0.306
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.35	0.4	0.42	0.38	0.41
7	$\bar{x}$	0.31	0.34	0.35	0.37	0.37	0.38	0.39	0.4	0.42	0.43	0.44	0.44	0.45
	σ	0.464	0.368	0.332	0.318	0.306	0.298	0.299	0.299	0.299	0.303	0.303	0.312	0.312
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.4	0.42	0.43	0.44
8	$\bar{x}$	0.32	0.32	0.32	0.33	0.34	0.35	0.36	0.38	0.39	0.41	0.41	0.43	0.44
	σ	0.465	0.36	0.323	0.304	0.29	0.279	0.276	0.279	0.278	0.281	0.285	0.292	0.299
	m	0	0	0.33	0.25	0.2	0.29	0.3	0.33	0.35	0.4	0.4	0.42	0.42
9	$\bar{x}$	0.32	0.32	0.33	0.33	0.35	0.36	0.37	0.39	0.42	0.41	0.43	0.43	0.44
	σ	0.465	0.364	0.334	0.311	0.301	0.294	0.287	0.285	0.287	0.288	0.293	0.297	0.304
	m	0	0	0.33	0.25	0.4	0.29	0.3	0.33	0.4	0.4	0.42	0.43	0.44
10	$\bar{x}$	0.32	0.32	0.34	0.34	0.35	0.36	0.38	0.39	0.39	0.42	0.43	0.44	0.43
	σ	0.468	0.366	0.332	0.316	0.301	0.293	0.29	0.285	0.282	0.289	0.287	0.294	0.3
	m	0	0	0.33	0.25	0.2	0.29	0.4	0.33	0.35	0.43	0.42	0.43	0.4
11	$\bar{x}$	0.32	0.32	0.35	0.34	0.35	0.37	0.37	0.39	0.4	0.43	0.42	0.43	0.44
	σ	0.465	0.371	0.344	0.319	0.312	0.302	0.293	0.286	0.29	0.287	0.294	0.293	0.303
	m	0	0	0.33	0.25	0.2	0.29	0.3	0.33	0.4	0.43	0.4	0.42	0.42
12	$\bar{x}$	0.33	0.32	0.33	0.35	0.35	0.37	0.38	0.39	0.41	0.42	0.43	0.45	0.42
	σ	0.47	0.363	0.321	0.305	0.298	0.287	0.288	0.287	0.287	0.298	0.301	0.314	0.308
	m	0	0	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.38	0.4	0.44	0.39
13	$\bar{x}$	0.31	0.34	0.35	0.36	0.37	0.38	0.4	0.42	0.43	0.44	0.44	0.45	0.45
	σ	0.464	0.37	0.333	0.318	0.306	0.298	0.298	0.298	0.299	0.305	0.308	0.313	0.314
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.43	0.42	0.43	0.44
14	$\bar{x}$	0.3	0.32	0.33	0.34	0.35	0.36	0.38	0.4	0.41	0.42	0.43	0.44	0.46
	σ	0.458	0.371	0.33	0.312	0.305	0.299	0.293	0.294	0.292	0.301	0.303	0.302	0.309
	m	0	0	0.33	0.25	0.4	0.29	0.3	0.33	0.4	0.4	0.4	0.42	0.45
15	$\bar{x}$	0.32	0.33	0.35	0.36	0.36	0.37	0.39	0.41	0.41	0.42	0.44	0.45	0.45
	σ	0.468	0.371	0.337	0.321	0.312	0.301	0.298	0.299	0.299	0.297	0.307	0.309	0.302
	m	0	0	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.4	0.42	0.43	0.45

Table D.44: Local update for 5-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.34	0.34	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
	σ	0.473	0.336	0.284	0.252	0.229	0.201	0.177	0.157	0.146	0.134	0.126	0.116	0.114
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.33	0.35	0.33	0.35	0.33	0.34
2	$\bar{x}$	0.34	0.34	0.33	0.34	0.34	0.35	0.36	0.37	0.38	0.41	0.43	0.46	0.48
	σ	0.474	0.342	0.288	0.252	0.234	0.208	0.189	0.172	0.165	0.164	0.166	0.171	0.171
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.4	0.4	0.43	0.45	0.48	0.51
3	$\bar{x}$	0.34	0.34	0.36	0.36	0.38	0.4	0.42	0.45	0.48	0.5	0.52	0.55	0.57
	σ	0.473	0.345	0.299	0.263	0.25	0.228	0.211	0.202	0.194	0.19	0.192	0.19	0.196
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.58	0.61
4	$\bar{x}$	0.33	0.33	0.34	0.34	0.33	0.34	0.34	0.35	0.36	0.38	0.39	0.43	0.45
	σ	0.47	0.344	0.292	0.257	0.231	0.211	0.188	0.167	0.159	0.155	0.152	0.16	0.163
	m	0	0.5	0.33	0.25	0.4	0.29	0.3	0.33	0.35	0.37	0.4	0.45	0.47
5	$\bar{x}$	0.35	0.34	0.35	0.35	0.36	0.37	0.39	0.41	0.44	0.47	0.5	0.53	0.53
	σ	0.476	0.347	0.299	0.269	0.251	0.228	0.215	0.201	0.194	0.185	0.181	0.176	0.178
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.5	0.53	0.57	0.56
6	$\bar{x}$	0.33	0.34	0.34	0.36	0.36	0.38	0.39	0.41	0.44	0.47	0.49	0.53	0.55
	σ	0.471	0.347	0.291	0.261	0.239	0.218	0.199	0.184	0.183	0.183	0.187	0.187	0.19
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.5	0.5	0.57	0.59
7	$\bar{x}$	0.34	0.36	0.38	0.39	0.42	0.43	0.47	0.49	0.51	0.54	0.55	0.58	0.58
	σ	0.475	0.353	0.306	0.275	0.263	0.238	0.219	0.21	0.203	0.196	0.196	0.189	0.199
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.57	0.62	0.63
8	$\bar{x}$	0.33	0.33	0.34	0.34	0.34	0.35	0.37	0.38	0.4	0.44	0.46	0.49	0.51
	σ	0.47	0.346	0.292	0.262	0.241	0.216	0.197	0.185	0.179	0.176	0.178	0.179	0.177
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.47	0.47	0.52	0.55
9	$\bar{x}$	0.32	0.35	0.36	0.36	0.38	0.4	0.42	0.45	0.48	0.5	0.52	0.54	0.56
	σ	0.468	0.357	0.313	0.282	0.264	0.247	0.222	0.204	0.2	0.184	0.178	0.176	0.177
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.57	0.59
10	$\bar{x}$	0.32	0.34	0.34	0.34	0.34	0.36	0.36	0.39	0.4	0.44	0.46	0.48	0.49
	σ	0.467	0.349	0.295	0.268	0.247	0.222	0.204	0.196	0.19	0.181	0.176	0.174	0.177
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.43	0.47	0.5	0.51
11	$\bar{x}$	0.34	0.35	0.35	0.35	0.38	0.38	0.39	0.42	0.43	0.44	0.47	0.5	0.5
	σ	0.473	0.364	0.315	0.288	0.277	0.25	0.23	0.215	0.209	0.194	0.187	0.182	0.178
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.53
12	$\bar{x}$	0.33	0.35	0.34	0.35	0.35	0.36	0.36	0.39	0.4	0.43	0.45	0.48	0.5
	σ	0.471	0.344	0.282	0.26	0.234	0.208	0.193	0.176	0.171	0.169	0.169	0.176	0.175
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.4	0.43	0.47	0.52	0.53
13	$\bar{x}$	0.33	0.35	0.37	0.38	0.39	0.42	0.45	0.48	0.5	0.53	0.55	0.57	0.59
	σ	0.47	0.346	0.296	0.267	0.251	0.228	0.209	0.203	0.198	0.194	0.196	0.195	0.199
	m	0	0.5	0.33	0.25	0.4	0.43	0.5	0.47	0.5	0.57	0.57	0.62	0.64
14	$\bar{x}$	0.33	0.34	0.34	0.35	0.36	0.37	0.39	0.42	0.44	0.48	0.49	0.53	0.54
	σ	0.472	0.35	0.288	0.269	0.251	0.224	0.21	0.196	0.189	0.183	0.191	0.191	0.194
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.5	0.53	0.57	0.57
15	$\bar{x}$	0.35	0.35	0.37	0.39	0.4	0.42	0.45	0.47	0.5	0.53	0.55	0.57	0.58
	σ	0.476	0.357	0.311	0.282	0.266	0.247	0.233	0.214	0.209	0.2	0.199	0.193	0.197
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.57	0.6	0.6	0.63

Table D.45: Local update for 10-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.37	0.35	0.36	0.36	0.37	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
	σ	0.482	0.344	0.281	0.246	0.222	0.185	0.162	0.134	0.117	0.102	0.094	0.087	0.079
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.33	0.35	0.37	0.36	0.37	0.36
2	$\bar{x}$	0.36	0.36	0.37	0.36	0.36	0.37	0.37	0.37	0.38	0.39	0.4	0.42	0.43
	σ	0.48	0.344	0.283	0.244	0.222	0.189	0.165	0.14	0.121	0.111	0.102	0.094	0.095
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.35	0.4	0.4	0.43	0.44
3	$\bar{x}$	0.36	0.37	0.38	0.4	0.4	0.41	0.44	0.46	0.48	0.52	0.54	0.57	0.6
	σ	0.48	0.347	0.289	0.256	0.231	0.202	0.178	0.153	0.145	0.126	0.124	0.112	0.106
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.58	0.63
4	$\bar{x}$	0.35	0.36	0.36	0.36	0.37	0.36	0.36	0.37	0.37	0.38	0.38	0.39	0.41
	σ	0.478	0.347	0.281	0.249	0.223	0.189	0.165	0.14	0.124	0.111	0.101	0.095	0.089
	m	0	0.5	0.33	0.25	0.4	0.29	0.4	0.4	0.35	0.37	0.38	0.4	0.41
5	$\bar{x}$	0.36	0.37	0.37	0.37	0.38	0.38	0.4	0.42	0.44	0.48	0.5	0.54	0.56
	σ	0.479	0.344	0.285	0.255	0.236	0.203	0.184	0.162	0.151	0.13	0.121	0.111	0.109
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.45	0.5	0.5	0.55	0.57
6	$\bar{x}$	0.36	0.37	0.37	0.37	0.38	0.38	0.39	0.4	0.42	0.45	0.47	0.5	0.53
	σ	0.481	0.344	0.284	0.253	0.224	0.197	0.17	0.145	0.132	0.119	0.11	0.106	0.104
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.43	0.47	0.52	0.55
7	$\bar{x}$	0.34	0.38	0.41	0.42	0.43	0.45	0.48	0.51	0.54	0.56	0.58	0.61	0.63
	σ	0.475	0.348	0.29	0.259	0.239	0.206	0.181	0.155	0.142	0.127	0.121	0.115	0.112
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.65
8	$\bar{x}$	0.37	0.36	0.36	0.36	0.36	0.37	0.38	0.39	0.39	0.41	0.43	0.46	0.48
	σ	0.483	0.342	0.287	0.25	0.224	0.194	0.173	0.147	0.13	0.121	0.112	0.109	0.107
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.47	0.49
9	$\bar{x}$	0.37	0.38	0.39	0.39	0.41	0.42	0.44	0.47	0.49	0.52	0.54	0.56	0.59
	σ	0.483	0.358	0.3	0.27	0.249	0.219	0.187	0.167	0.149	0.128	0.121	0.114	0.096
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.57	0.6
10	$\bar{x}$	0.36	0.36	0.36	0.36	0.36	0.37	0.37	0.37	0.38	0.39	0.4	0.42	0.43
	σ	0.48	0.349	0.282	0.251	0.226	0.199	0.174	0.15	0.135	0.126	0.12	0.116	0.112
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.35	0.4	0.4	0.42	0.44
11	$\bar{x}$	0.36	0.37	0.37	0.37	0.38	0.38	0.39	0.39	0.4	0.42	0.43	0.44	0.45
	σ	0.48	0.361	0.306	0.281	0.258	0.234	0.209	0.185	0.174	0.152	0.148	0.134	0.132
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.43	0.45
12	$\bar{x}$	0.37	0.37	0.36	0.36	0.36	0.37	0.37	0.37	0.38	0.39	0.39	0.41	0.43
	σ	0.484	0.339	0.278	0.245	0.221	0.187	0.163	0.136	0.123	0.108	0.1	0.089	0.086
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.42	0.42
13	$\bar{x}$	0.36	0.36	0.38	0.39	0.4	0.41	0.43	0.45	0.48	0.51	0.54	0.57	0.6
	σ	0.479	0.339	0.285	0.251	0.226	0.198	0.173	0.149	0.138	0.125	0.119	0.118	0.112
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.53	0.55	0.58	0.61
14	$\bar{x}$	0.36	0.36	0.38	0.37	0.37	0.38	0.39	0.4	0.42	0.44	0.47	0.5	0.53
	σ	0.48	0.345	0.289	0.247	0.225	0.197	0.17	0.148	0.139	0.121	0.112	0.113	0.109
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.47	0.47	0.52	0.55
15	$\bar{x}$	0.36	0.39	0.4	0.41	0.43	0.45	0.48	0.5	0.53	0.56	0.59	0.61	0.63
	σ	0.479	0.351	0.297	0.266	0.242	0.211	0.184	0.164	0.145	0.128	0.12	0.113	0.116
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.63	0.65

Table D.46: Local update for 20-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	σ	0.487	0.347	0.288	0.241	0.219	0.185	0.159	0.132	0.113	0.096	0.085	0.068	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.38	0.39
2	$\bar{x}$	0.39	0.39	0.4	0.39	0.39	0.4	0.39	0.39	0.4	0.4	0.41	0.41	0.41
	σ	0.487	0.346	0.286	0.245	0.218	0.187	0.159	0.129	0.113	0.092	0.082	0.07	0.062
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.41
3	$\bar{x}$	0.39	0.4	0.4	0.41	0.41	0.42	0.44	0.45	0.47	0.5	0.53	0.56	0.6
	σ	0.487	0.35	0.283	0.252	0.224	0.193	0.162	0.134	0.118	0.103	0.089	0.076	0.067
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.5	0.5	0.53	0.57	0.6
4	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.4	0.39	0.39	0.4	0.4	0.4
	σ	0.487	0.346	0.283	0.248	0.221	0.186	0.159	0.13	0.116	0.094	0.084	0.072	0.063
	m	0	0.5	0.33	0.25	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
5	$\bar{x}$	0.4	0.4	0.4	0.4	0.4	0.4	0.42	0.43	0.44	0.47	0.48	0.52	0.54
	σ	0.489	0.352	0.287	0.253	0.232	0.193	0.168	0.144	0.127	0.111	0.099	0.082	0.074
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
6	$\bar{x}$	0.39	0.39	0.39	0.39	0.4	0.4	0.4	0.41	0.42	0.43	0.44	0.46	0.48
	σ	0.487	0.345	0.286	0.245	0.219	0.189	0.159	0.13	0.115	0.095	0.085	0.075	0.068
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.47	0.49
7	$\bar{x}$	0.39	0.41	0.43	0.44	0.45	0.46	0.48	0.51	0.53	0.56	0.59	0.61	0.64
	σ	0.488	0.35	0.29	0.253	0.227	0.196	0.167	0.137	0.118	0.099	0.088	0.073	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.62	0.64
8	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.4	0.4	0.4	0.4	0.41	0.42	0.43	0.45
	σ	0.488	0.345	0.285	0.25	0.221	0.19	0.156	0.133	0.116	0.098	0.089	0.08	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.43	0.45
9	$\bar{x}$	0.4	0.4	0.4	0.42	0.42	0.44	0.45	0.47	0.48	0.52	0.53	0.56	0.58
	σ	0.489	0.354	0.296	0.263	0.236	0.208	0.178	0.15	0.132	0.11	0.097	0.082	0.072
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.57	0.59
10	$\bar{x}$	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.4	0.4	0.41
	σ	0.488	0.345	0.284	0.246	0.222	0.187	0.161	0.135	0.117	0.101	0.091	0.08	0.074
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
11	$\bar{x}$	0.39	0.4	0.39	0.4	0.39	0.4	0.4	0.4	0.41	0.41	0.41	0.42	0.42
	σ	0.487	0.357	0.303	0.266	0.249	0.218	0.191	0.164	0.151	0.133	0.125	0.111	0.106
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.42	0.42
12	$\bar{x}$	0.4	0.39	0.39	0.4	0.4	0.39	0.39	0.39	0.4	0.4	0.4	0.4	0.4
	σ	0.489	0.345	0.283	0.249	0.216	0.187	0.155	0.129	0.113	0.095	0.083	0.066	0.061
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.4	0.4	0.4
13	$\bar{x}$	0.38	0.38	0.4	0.4	0.4	0.41	0.42	0.43	0.44	0.46	0.48	0.51	0.54
	σ	0.486	0.347	0.282	0.245	0.22	0.188	0.156	0.134	0.115	0.1	0.088	0.078	0.07
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.47	0.47	0.52	0.55
14	$\bar{x}$	0.39	0.39	0.39	0.39	0.4	0.4	0.4	0.41	0.41	0.43	0.44	0.47	0.48
	σ	0.488	0.345	0.286	0.248	0.224	0.188	0.159	0.132	0.115	0.096	0.084	0.073	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.45	0.47	0.49
15	$\bar{x}$	0.4	0.41	0.42	0.43	0.45	0.46	0.48	0.51	0.53	0.56	0.58	0.61	0.64
	σ	0.49	0.353	0.293	0.256	0.23	0.196	0.165	0.139	0.122	0.105	0.092	0.075	0.066
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.55	0.57	0.6	0.62	0.64

Table D.47: Local update for 40-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.43	0.41	0.43	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	σ	0.494	0.345	0.288	0.246	0.219	0.189	0.158	0.128	0.108	0.091	0.079	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.42	0.41
2	$\bar{x}$	0.42	0.41	0.42	0.42	0.42	0.42	0.42	0.42	0.43	0.42	0.42	0.42	0.43
	σ	0.494	0.35	0.284	0.249	0.223	0.186	0.158	0.128	0.111	0.089	0.076	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.42
3	$\bar{x}$	0.42	0.42	0.43	0.42	0.43	0.44	0.44	0.46	0.47	0.49	0.51	0.54	0.57
	σ	0.493	0.353	0.288	0.248	0.227	0.191	0.159	0.131	0.117	0.095	0.084	0.071	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.5	0.5	0.55	0.57
4	$\bar{x}$	0.41	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	σ	0.491	0.346	0.283	0.247	0.217	0.188	0.155	0.129	0.113	0.092	0.078	0.063	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.42	0.42
5	$\bar{x}$	0.42	0.41	0.42	0.43	0.42	0.43	0.44	0.44	0.45	0.46	0.47	0.5	0.52
	σ	0.494	0.348	0.286	0.246	0.226	0.191	0.163	0.135	0.118	0.1	0.086	0.078	0.069
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.5	0.51
6	$\bar{x}$	0.42	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.45
	σ	0.494	0.35	0.29	0.243	0.218	0.186	0.158	0.127	0.112	0.092	0.082	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.45	0.45
7	$\bar{x}$	0.41	0.43	0.44	0.45	0.46	0.47	0.49	0.51	0.52	0.55	0.57	0.6	0.62
	σ	0.492	0.352	0.287	0.249	0.221	0.19	0.159	0.133	0.114	0.092	0.081	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.57	0.6	0.63
8	$\bar{x}$	0.41	0.41	0.42	0.42	0.42	0.42	0.43	0.42	0.42	0.43	0.43	0.43	0.44
	σ	0.492	0.345	0.285	0.25	0.222	0.188	0.156	0.131	0.112	0.092	0.084	0.066	0.059
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.43	0.44
9	$\bar{x}$	0.42	0.43	0.43	0.43	0.45	0.46	0.46	0.47	0.48	0.51	0.53	0.55	0.56
	σ	0.493	0.355	0.295	0.26	0.235	0.201	0.17	0.143	0.124	0.101	0.089	0.077	0.065
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.55	0.56
10	$\bar{x}$	0.43	0.41	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	σ	0.494	0.347	0.284	0.247	0.222	0.189	0.158	0.129	0.112	0.093	0.082	0.068	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.43	0.42	0.42	0.42
11	$\bar{x}$	0.43	0.42	0.43	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.43	0.43
	σ	0.495	0.36	0.296	0.263	0.24	0.206	0.179	0.154	0.137	0.119	0.107	0.1	0.085
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.42	0.42
12	$\bar{x}$	0.42	0.42	0.41	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	σ	0.494	0.348	0.286	0.252	0.222	0.187	0.154	0.129	0.11	0.091	0.079	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.4	0.4	0.42	0.42	0.41
13	$\bar{x}$	0.41	0.42	0.43	0.42	0.42	0.42	0.42	0.43	0.43	0.44	0.45	0.46	0.47
	σ	0.491	0.349	0.286	0.248	0.221	0.186	0.158	0.129	0.111	0.092	0.08	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.45	0.45	0.47
14	$\bar{x}$	0.42	0.42	0.42	0.42	0.43	0.42	0.43	0.42	0.43	0.43	0.44	0.45	0.45
	σ	0.494	0.347	0.285	0.244	0.221	0.188	0.157	0.13	0.111	0.096	0.081	0.065	0.06
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.4	0.45	0.43	0.42	0.45	0.45
15	$\bar{x}$	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.5	0.52	0.55	0.57	0.6	0.62
	σ	0.494	0.353	0.285	0.256	0.225	0.193	0.162	0.135	0.113	0.093	0.08	0.065	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.57	0.6	0.63

Table D.48: Local update for 80-dimensional Rosenbrock function

Method		λ												
		1	2	3	4	5	7	10	15	20	30	40	60	80
1	$\bar{x}$	0.44	0.43	0.44	0.44	0.45	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
	σ	0.496	0.347	0.286	0.249	0.221	0.187	0.157	0.128	0.11	0.092	0.077	0.065	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.44
2	$\bar{x}$	0.45	0.44	0.43	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
	σ	0.497	0.347	0.285	0.245	0.22	0.187	0.156	0.131	0.11	0.093	0.079	0.064	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
3	$\bar{x}$	0.44	0.45	0.44	0.45	0.45	0.45	0.45	0.46	0.47	0.48	0.5	0.51	0.53
	σ	0.496	0.354	0.286	0.253	0.225	0.187	0.158	0.127	0.11	0.092	0.082	0.067	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.45	0.47	0.5	0.52	0.53
4	$\bar{x}$	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.45	0.45	0.44
	σ	0.497	0.35	0.285	0.247	0.224	0.19	0.157	0.13	0.111	0.091	0.079	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
5	$\bar{x}$	0.43	0.43	0.44	0.44	0.44	0.45	0.45	0.45	0.46	0.46	0.47	0.49	0.5
	σ	0.495	0.353	0.287	0.25	0.222	0.189	0.158	0.131	0.117	0.095	0.084	0.073	0.063
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.47	0.47	0.48	0.5
6	$\bar{x}$	0.45	0.45	0.45	0.44	0.44	0.45	0.45	0.44	0.44	0.44	0.45	0.45	0.46
	σ	0.497	0.348	0.287	0.25	0.222	0.19	0.159	0.128	0.112	0.094	0.078	0.065	0.053
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.46
7	$\bar{x}$	0.44	0.45	0.46	0.46	0.47	0.48	0.49	0.5	0.52	0.53	0.55	0.57	0.59
	σ	0.497	0.35	0.29	0.252	0.224	0.192	0.16	0.13	0.113	0.091	0.079	0.064	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.58	0.6
8	$\bar{x}$	0.45	0.44	0.44	0.44	0.45	0.44	0.44	0.44	0.44	0.45	0.44	0.45	0.45
	σ	0.497	0.35	0.286	0.25	0.222	0.186	0.156	0.127	0.113	0.091	0.082	0.067	0.058
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
9	$\bar{x}$	0.43	0.45	0.45	0.46	0.46	0.46	0.47	0.48	0.49	0.51	0.52	0.54	0.55
	σ	0.495	0.354	0.291	0.253	0.232	0.197	0.165	0.139	0.123	0.099	0.089	0.07	0.064
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.47	0.5	0.5	0.53	0.53	0.55
10	$\bar{x}$	0.44	0.45	0.45	0.44	0.44	0.44	0.45	0.44	0.44	0.44	0.44	0.44	0.44
	σ	0.496	0.348	0.287	0.249	0.218	0.187	0.159	0.128	0.111	0.092	0.08	0.066	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.43	0.44
11	$\bar{x}$	0.44	0.45	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
	σ	0.497	0.361	0.299	0.256	0.234	0.203	0.174	0.148	0.129	0.11	0.099	0.085	0.078
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.43	0.44
12	$\bar{x}$	0.44	0.45	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
	σ	0.496	0.352	0.288	0.245	0.219	0.187	0.153	0.129	0.109	0.09	0.078	0.063	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.44
13	$\bar{x}$	0.43	0.44	0.43	0.44	0.44	0.44	0.45	0.44	0.45	0.44	0.45	0.45	0.45
	σ	0.495	0.353	0.285	0.251	0.22	0.189	0.16	0.127	0.111	0.094	0.079	0.064	0.056
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
14	$\bar{x}$	0.45	0.44	0.44	0.44	0.44	0.45	0.45	0.44	0.44	0.45	0.45	0.45	0.46
	σ	0.498	0.351	0.286	0.248	0.219	0.188	0.158	0.129	0.112	0.09	0.078	0.067	0.057
	m	0	0.5	0.33	0.5	0.4	0.43	0.4	0.47	0.45	0.43	0.45	0.45	0.45
15	$\bar{x}$	0.45	0.44	0.46	0.47	0.47	0.48	0.49	0.51	0.52	0.54	0.55	0.57	0.59
	σ	0.497	0.351	0.287	0.251	0.221	0.19	0.162	0.13	0.111	0.093	0.08	0.066	0.055
	m	0	0.5	0.33	0.5	0.4	0.43	0.5	0.53	0.5	0.53	0.55	0.58	0.59

List of Figures

List of Tables

Bibliography

M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 9th edition, 1972.

B. C. Arnold and R. J. Beaver. Skewed multivariate models related to hidden truncation and/or selective reporting. *Test. Sociedad de Estadística e Investigación Operativa*, 11(1):7—-54, 2002.

A. Auger. Convergence results for the (1, λ)-SA-ES using the theory of φ-irreducible Markov chains. *Theoretical Computer Science*, 334(1-3):35–69, 2005.

A. Auger, C. L. Bris, and M. Schoenauer. Dimension-independent convergence rate for non-isotropic (1,λ)-ES. In *GECCO*, pages 512–524, 2003.

A. Auger, M. Schoenauer, and N. Vanhaecke. LS-CMA-ES: A second-order algorithm for covariance matrix adaptation. In *Parallel Problem Solving from Nature*, volume 3242 of *Lecture Notes in Computer Science*, pages 182–191. Springer-Verlag, Berlin Heidelberg, 2004.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistic*, 12:171–178, 1985.

A. Azzalini. Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199–208, 1986.

A. Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistic*, 32(2):159–188, June 2005.

A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602, 1999.

A. Azzalini and A. Dalla Valle. The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996.

T. Bäck. Self-adaption in genetic algorithms. In F. Varela and P. Bourgine, editors, *Toward a Practice of Autonomous Systems: Proceedings of the 1st European Conference on Articial Life*, pages 263–271. MIT Press, 1992.

T. Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, New York, 1996.

T. Bäck and M. Schütz. Intelligent mutation rate control in canonical genetic algorithms. In Z. W. Rás and M. Michalewicz, editors, *Proc. of the Ninth Int. Symposium on Methodologies for Intelligent Systems, ISMIS'96*, volume 1079 of *Lecture Notes in Artificial Intelligence*, pages 158–167. Springer Verlag, Berlin, 1996.

T. Bäck and H.-P. Schwefel. An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation*, 1(1):1–23, 1993.

T. Bäck and H.-P. Schwefel. Evolutionary computation: An overview. In *Proc. of the IEEE International Conference on Evolutionary Computation (ICEC '96), May 20-22, 1996, Nayoya University, Japan.*, pages 20–29, 1996.

T. Bäck, F. Hoffmeister, and H.-P. Schwefel. A survey of evolution strategies. In R. K. Belew and L. B. Booker, editors, *Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA*, pages 2–9. Morgan Kaufmann, July 1991. ISBN 1-55860-208-9.

T. Bäck, G. Rudolph, and H.-P. Schwefel. Evolutionary programming and evolution strategies: Similarities and differences. In *Proceedings of the Second Annual Conference on Evolutionary Programming*, pages 11–22, 1993.

T. Bäck, D. B. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. Institute of Physics Publishing, Bristol and Oxford University Press, Oxford, 1997a.

T. Bäck, U. Hammel, and H.-P. Schwefel. Evolutionary Computation: Comments on the History and Current State. *IEEE Transactions on Evolutionary Computation*, 1(1):3–17, 1997b.

S. Berlik. Entwicklung einer Taxonomie der Evolutionsstrategien und verwandter, stochastischer Optimierverfahren. Diploma thesis, Universität Dortmund, Fachbereich Informatik, 2000.

S. Berlik. Optimierung von Schraubenmaschinenprofilen. 21. Workshop "Interdisziplinäre Methoden in der Informatik" Forschungsbericht Nr. 477, Universität Dortmund, 2001.

S. Berlik. A polymorphical mutation operator for evolution strategies. In M. Wagenknecht and R. Hampel, editors, *Proc. of the Int. Conf. in Fuzzy Logic and Technology, EUSFLAT, Sep. 10–12, 2003, Zittau, Germany*, pages 502–505. European Society for Fuzzy Logic and Technology, EUSFLAT, 2003a.

S. Berlik. Polymorphe Mutation. 22. Workshop "Interdisziplinäre Methoden in der Informatik" Forschungsbericht Nr. 783, Universität Dortmund, 2003b.

S. Berlik. A directed mutation framework for evolutionary algorithms. In R. Matoušek and P. Ošmera, editors, *Proc. of the 10th Int. Conf. on Soft Computing, MENDEL, Jun. 16–18, 2004, Brno, Czech Republic*, pages 45–50, 2004a.

S. Berlik. A step size preserving directed mutation operator. In K. Deb, editor, *Proc. of the Int. Genetic and Evolutionary Computation Conf., GECCO 2004, Jun. 26–30, 2004, Seattle, USA*, volume 1 of *Lecture Notes in Computer Science*. Springer-Verlag Berlin Heidelberg, 2004b.

S. Berlik and M. Fathi. Multi-objective optimization using directed mutation. In H. R. Arabnia and R. Joshua, editors, *Proc. of the Int. Conf. on Artificial Intelligence, ICAI 2005, June 27–30, 2005, Las Vegas, USA*, volume II, pages 870–875. CSREA Press, USA, 2005a.

S. Berlik and M. Fathi. A design and optimization tool for screw type machines. In *Proc. of the Int. Conf. on Systems, Man and Cybernetics, IEEE SMC 2005, Oct. 10–12, 2005, Hawaii, USA*. IEEE Press, 2005b.

S. Berlik and B. Reusch. Directed mutation by means of the skew-normal distribution. In B. Reusch, editor, *Proc. of the Int. Conf. on Computational Intelligence, FUZZY DAYS, Sep. 29–Oct. 1, 2004, Dortmund, Germany*, Advances in Soft Computing, pages 35–50. Springer-Verlag Berlin Heidelberg, 2004.

S. Berlik and B. Reusch. Directed mutation operators – an overview. In R. Khosla, R. J. Howlett, and L. C. Jain, editors, *Proc. of the 9th Int. Conf. on Knowledge-Based & Intelligent Information & Engineering Systems, KES 2005, Sept. 14–16, 2005, Melbourne, Australia*, volume 3683 of *Lecture Notes in Computer Science*, pages 1151–1159. Springer-Verlag Berlin Heidelberg, 2005.

S. Berlik and B. Reusch. *Foundations of Directed Mutation*, chapter in: Integrated Intelligent Systems for Engineering Design. IOS Press, 2006.

S. Berlik, M. Fathi, and A. Holland. Advances in optimizing screw-type machines. In *Proc. of the Int. Conf. on Systems, Man and Cybernetics, IEEE SMC 2006, Oct. 8–11, 2006, Taipei, Taiwan [to appear]*. IEEE Press, 2006.

H.-G. Beyer. Some aspects of the 'evolution strategiy' for solving tsp-like optimization problems appearing at the design studies of a 0.5 tev e^+e^--linear collider. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, PPSN-II, Brussels, Belgium, September 28-30*, pages 363–372. Elsevier, 1992.

H.-G. Beyer. *The Theory of Evolution Strategies*. Springer-Verlag, Berlin Heidelberg, 2001.

H.-G. Beyer and H.-P. Schwefel. Evolution strategies - a comprehensive introduction. *Natural Computing: an international journal*, 1(1):3–52, 2002.

P. Bienert, I. Rechenberg, and H.-P. Schwefel. Messung kleiner Wandschubspannungen bei turbulenten Grenzschichten in Ablösenähe. Technical Report of the Hermann Föttinger–Institute for Fluid Dynamics Wi 8/45, Technical University of Berlin, September 1966.

Z. W. Birnbaum. Effect of linear truncation on a multinormal population. *The Annals of Mathematical Statitics*, 21(272—279), 1950.

G. E. P. Box. Evolutionary operation: A method for increasing industrial productivity. *Applied Statistics*, 6:81–101, 1957.

G. E. P. Box and N. R. Draper. *Evolutionary operation. A statistical method for process improvement.* John Wiley & Sons, New York, 1969.

H. J. Bremermann. *Self?Organizing Systems*, chapter Optimization through evolution and recombination. Spartan Books, 1962.

C. G. Broyden. The convergence of a class of double rank minimization, Part 2: the new algorithm. *Journal of the Institute of Mathematics and Applications*, 6:222–231, 1970.

J. Cairns, J. Overbaugh, and S. Miller. The origin of mutants. *Nature*, 335: 142–145, 1988.

Y. S. Chow and H. Teicher. *Probability Theory.* Springer, New York, third edition, 1997.

W. J. Cody. Rational chebyshev approximations for the error function. *Mathematics of Computation*, pages 631–638, 1969.

C. A. C. Coello. A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information Systems*, 1(3):129–156, 1999.

C. Darwin. *On the Origin of Species.* John Murray, London, 1859.

K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons: Chichester, UK, 2001.

K. Deb, S. Agarwal, A. Pratap, and T. Meyarivan. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. Technical report, KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000.

L. T. DeCarlo. On the meaning and use of kurtosis. *Psychological Methods*, 2: 292—-307, 1997.

L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, New York, 1986.

Á. E. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 3(2): 124–141, 1999.

J. T. Ferreira and M. F. Steel. A constructive representation of univariate skewed distributions. *???*, 2004.

R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, 13:317–322, 1970.

D. B. Fogel. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. IEEE Press, Piscataway, NJ., 2nd edition, 1999.

L. J. Fogel. Autonomous automata. *Industrial Research*, 4:14–19, 1962.

L. J. Fogel. *On the Organization of Intellect*. PhD thesis, University of California,, Los Angeles, 1964.

G. J. Friedman. Digital simulation of an evolutionary process. *General Systems Yearbook*, 4:171–184, 1959.

D. J. Futuyma. *Evolutionary Biology*. Sinauer Associates, Sunderland, MA, 1998.

D. J. Futuyma. *Evolution*. Sinauer Associates, Sunderland, MA, 2005.

M. G. Genton, editor. *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC, 2004.

M. G. Genton. Discussion of "the skew-normal". *Scandinavian Journal of Statistic*, 32:189–98, 2005.

A. Ghozeil and D. B. Fogel. A preliminary investigation into directed mutations in evolutionary algorithms. In *Parallel Problem Solving from Nature*, volume 1141 of *Lecture Notes in Computer Science*, pages 329–335, 1996.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, 1989.

D. Goldfarb. A family of variable metric methods derived by variational means. *Mathematics of Computation*, 24:23–26, 1970.

A. K. Gupta, F. C. Chang, and W. J. Huang. Some skew-symmetric models. *Random Operators & Stochastic Equations*, 10(2):133–141, 2002.

R. C. Gupta and R. D. Gupta. Generalized skew normal model. *Test. Sociedad de Estadística e Investigación Operativa*, 13(2):501–524, 2004.

N. Hansen. The cma evolution strategy: A tutorial, 2005. URL www.bionik.tu-berlin.de/user/niko/cmatutorial.pdf.

N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In X. Yao, E. K. Burke, J. A. Lozano, J. Smith, J. J. M. Guervós, J. A. Bullinaria, J. E. Rowe, P. Tiño, A. Kabán, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature*, volume 3242 of *Lecture Notes in Computer Science*, pages 282–291. Springer-Verlag Berlin Heidelberg, 2004.

N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *IEEE International Conference on Evolutionary Computation*, pages 312–317, 1996.

N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

N. Hansen, S. D. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

M. Helpertz. *Methode zur stochastischen Optimierung von Schraubenrotorprofilen.* PhD thesis, Universität Dortmund, Fakultät Maschinenbau, 2003.

N. Henze. A probabilistic representation of the 'skew-normal' distribution. *Scandinavian Journal of Statistic*, 13:271–275, 1986.

L. Hildebrand. Schiefe - Ein Ansatz zur Erweiterung von Evolutionsstrategien um gerichtete Mutation. 16. Workshop Interdisziplinäre Methoden der Informatik, Universität Dortmund, 1996.

L. Hildebrand. *Asymmetrische Evolutionsstrategien.* PhD thesis, Universität Dortmund, Fachbereich Informatik, 2001.

G. W. Hill and A. W. Davis. Normal deviate [s14] (algorithm 442). *Communications of the ACM*, 16(1):51–52, 1973.

J. H. Holland. Outline for a logical theory of adaptive systems. *Journal of the Association for Computing Machinery*, 3:297–311, 1962.

J. H. Holland. A new kind of turnpike theorem. *Bulletin of the American Mathematical Society*, 75(6):1311–1317, 1969.

J. H. Holland. Genetic algorithms and the optimal allocation of trials. *SIAM Journal on Computing*, 2(2):88–105, 1973.

J. H. Holland. *Adaptation in Natural and Artificial Systems.* MIT Press, 1994.

M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods.* Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1999.

W. Hörmann, J. Leydold, and G. Derflinger. *Automatic Nonuniform Random Variate Generation.* Springer-Verlag, Berlin Heidelberg, 2004.

A. F. Karr. *Probability.* Springer-Verlag, Berlin, 1993.

K. Kauder, B. Reusch, M. Helpertz, and S. Berlik. Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 1. *Schraubenmaschinen*, 9:27–46, 2001.

K. Kauder, B. Reusch, M. Helpertz, and S. Berlik. Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 2. *Schraubenmaschinen*, 10:17–34, 2002a.

K. Kauder, B. Reusch, M. Helpertz, and S. Berlik. Optimisation methods for rotors of twin-screw compressors. *VDI Berichte*, 1715:29–50, 2002b.

K. Kauder, B. Reusch, M. Helpertz, and S. Berlik. Automatisierte Optimierung der Geometrie von Schraubenrotoren, Teil 3. *Schraubenmaschinen*, 11:15–29, 2003.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.

J. Klockgether and H.-P. Schwefel. Two-phase nozzle and hollow core jet experiments. In D. G. Elliott, editor, *Proc. of the 11th Symp. Engineering Aspects of Magnetohydrodynamics, California Institute of Technology, Pasadena CA*, pages 141–148, 1970.

T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3): 385–482, 2003.

J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, 1992.

F. Kursawe. Towards self-adapting evolution strategies. In *IEEE Conf. on Evolutionary Computation*, volume 1, pages 283–288, Piscataway, NJ, 1995. IEEE Service Center.

J. Lederberg and E. M. Lederberg. Replica plating and indirect selection of bacterial mutants. *J. Bacteriol.*, 65:399–406, 1951.

H. J. Lichtfuss. Evolution eines Rohrkrümmers. Master's thesis, Institut für Strömungstechnik, Technische Universität Berlin, March 1965.

S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28:491–511, 1943.

H. B. Mann and D. R. Whitney. On a test of whether one of two variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18: 50–60, 1947.

D. J. Montana and L. Davis. Training Feedforward Neural Networks Using Genetic Algorithms. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence IJCAI-89*, pages 762–767, San Mateo, CA, 1989. Morgan Kaufmann.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

A. Ostermeier. An evolution strategy with momentum adaptation of the random number distribution. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature*, pages 197–206. Elsevier Science Publishers, 1992.

A. Ostermeier, A. Gawelczyk, and N. Hansen. Step-size adaptation based on non-local use of selection information. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature–PPSN IV, Proceedings*, pages 189–198. Springer Verlag, 1994.

V. Pareto. *Manuale di economia politica con una introduzione alla scienza sociale.* Società Editrice Libraria, Milano, Italy, 1906.

K. Pearson. Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder. *Biometrika*, 4:169–212, 1905.

W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing.* Cambridge University Press, Cambridge, 2nd edition, 1993.

D. Rasch. *Mathematische Statistik.* Johann Ambrosius Barth Verlag, Heidelberg Leipzig, 1995.

I. Rechenberg. Kybernetische Lösungsansteuerung einer experimentellen Forschungsaufgabe. Presented at the annual conference of the WGLR in Berlin, 1964.

I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog Verlag, Stuttgart, 1973.

I. Rechenberg. *Evolutionsstrategie '94.* Frommann-Holzboog Verlag, Stuttgart, 1994.

S. M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists.* Elsevier Academic Press, 2nd edition, 2000.

G. Rudolph. On correlated mutations in evolution strategies. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature*, pages 105–114. Elsevier, 1992.

G. Rudolph. *Convergence Properties of Evolutionary Algorithms.* PhD thesis, Universität Dortmund, Fachbereich Informatik, 1996.

G. Rudolph. Local convergence rates of simple evolutionary algorithms with Cauchy mutations. *IEEE Transactions on Evolutionary Computation*, 1(4): 249–258, 1998.

H.-P. Schwefel. Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik. Diploma thesis, Technical University of Berlin, Hermann Föttinger–Institute for Fluid Dynamics, March 1965.

H.-P. Schwefel. Adaptive Mechanismen in der biologischen Evolution und ihr Einfluß auf die Evolutionsgeschwindigkeit. Technical report, Technical University of Berlin, 1974.

H.-P. Schwefel. *Evolutionsstrategie und numerische Optimierung*. PhD thesis, Technische Universität Berlin, Fachbereich Verfahrenstechnik, 1975.

H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*, volume 26 of *Interdisciplinary Systems Research*. Birkhäuser, Basle, 1977.

H.-P. Schwefel. *Numerical Optimization of Computer Models*. John Wiley & Sons, Chichester, 1981.

H.-P. Schwefel. Collective phenomena in evolutionary systems. *Problems of Constancy and Change – the Complementarity of Systems Approaches to Complexity, Papers presented at the 31st Annual Meeting of the Int'l Soc. for General System Research*, 2:1025–1033, 1987.

H.-P. Schwefel. *Evolution and Optimum Seeking*. John Wiley & Sons, New York, 1995.

H.-P. Schwefel and G. Rudolph. Contemporary evolution strategies. In F. Morán, A. Moreno, J. J. M. Guervós, and P. Chacón, editors, *Advances in Artificial Life, Third European Conference on Artificial Life*, volume 929 of *Lecture Notes in Computer Science*, pages 893–907. Springer, 1995.

D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24:647–657, 1970.

J. Spanier and K. B. Oldham. *An Atlas of Functions*. Hemisphere Publishing Corporation, 1987.

W. Spendley, G. R. Hext, and F. R. Himsworth. Sequential application of simplex design in optimization and evolutionary operation. *Technometrics*, 4:441–461, 1962.

M. D. Springer. *The Algebra of Random Variables*. John Wiley & Sons, New York, 1979.

D. A. van Veldhuizen and G. B. Lamont. Multiobjective evolutionary algorithm research: A history and analysis. Technical Report TR-98-03, Department of Electrical and Computer Engineering, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 1998.

D. A. van Veldhuizen and G. B. Lamont. Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary Computation*, 8(2):125–147, 2000.

H.-M. Voigt and T. Anheyer. Modal Mutations in Evolutionary Algorithms. In Z. Michalewicz, J. Schaffer, H.-P. Schwefel, D. Fogel, and H. Kitano, editors, *Proc. of the first IEEE International Conference on Evolutionary Computation*, pages 88–92, Piscataway, NJ, 1994. IEEE Press.

P. von Sengbusch. *Botanik*. McGraw-Hill Book Company, Hambug, 1989.

J. Wang, J. Boyer, and M. G. Genton. A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, 14(4):1259–1270, 2004a.

J. Wang, J. Boyer, and M. G. Genton. A note on an equivalence between chi-square and generalized skew-normal distributions. *Statistics & Probability Letters*, 66:395–398, 2004b.

K. Weicker. Problem Diffculty in Real-Valued Dynamic Problems. In B. Reusch, editor, *Fuzzy Days*, volume 2206 of *Lecture Notes in Computer Science*, pages 313–325. Springer, 2001.

K. Weicker and N. Weicker. On Evolution Strategy Optimization in Dynamic Environments. In *Congress on Evolutionary Computation*, pages 2039–2046, Piscataway, NJ, 1999. IEEE Service Center.

F. Wilcoxon. Individual comparison by ranking methods. *Biometrics Bulletin*, 1:80–83, 1945.

M. H. Wright. Direct search methods: Once scorned, now respectable. In D. Griffiths and G. Watson, editors, *Proc. of the 1995 Biennial Conference on Numerical Analysis, Numerical Analysis 1995*, pages 191–208. Addison Wesley, Redwood City, 1995.

X. Yao and Y. Liu. Fast Evolution Strategies. *Control and Cybernetics*, 26(3): 467–496, 1997.

S. Zacks. *Parametric statistical inference*. Pergamon Press, Oxford, 1981.

Q. Zheng. Mathematical issues arising from the directed mutation controversy. *Genetics*, 164:373–379, 2003.

Index