International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:7, No:1, 2013

# Probability of Globality

Eva Eggeling, Dieter W. Fellner, Torsten Ullrich

*Abstract*—The objective of global optimization is to find the globally best solution of a model. Nonlinear models are ubiquitous in many applications and their solution often requires a global search approach; i.e. for a function $f$ from a set $A \subset \mathbb{R}^n$ to the real numbers, an element $\mathbf{x}_0 \in A$ is sought-after, such that $\forall\, \mathbf{x} \in A \;:\; f(\mathbf{x}_0) \leq f(\mathbf{x})$. Depending on the field of application, the question whether a found solution $\mathbf{x}_0$ is not only a local minimum but a global one is very important.

This article presents a probabilistic approach to determine the probability of a solution being a global minimum. The approach is independent of the used global search method and only requires a limited, convex parameter domain $A$ as well as a Lipschitz continuous function $f$ whose Lipschitz constant is not needed to be known.

*Keywords*—global optimization, probability theory, probability of globality

## I. Introduction

AN optimization problem can be represented in the following way: For a function $f$ from a set $A$ to the real numbers, an element $\mathbf{x}_0 \in A$ is sought-after, such that

$$\forall\, \mathbf{x} \in A \;:\; f(\mathbf{x}_0) \leq f(\mathbf{x}). \tag{1}$$

Such a formulation is called a minimization problem and the element $\mathbf{x}_0$ is a global minimum. Without loss of generality, it is sufficient to investigate minimization problems; maximization problems can be transformed to minimization problems via duality. The maximization of a real-valued function $g(\mathbf{x})$ can be regarded as the minimization of the transformed function

$$f(\mathbf{x}) = (-1) \cdot g(\mathbf{x}). \tag{2}$$

Depending on the field of application, $f$ is called an objective function, cost function, energy function, or energy functional. A feasible solution that minimizes the objective function is called an optimal solution.

Typically, $A$ is some subset of the Euclidean space $\mathbb{R}^n$, often specified by a set of constraints (equalities or inequalities) that the members of $A$ have to satisfy. The domain $A$ of $f$ is often called search space or choice set, while the elements of $A$ are called candidate solutions or feasible solutions. In this article the domain $A$ shall always be limited and convex.

Generally, a function $f$ may have several local minima, where a local minimum $\mathbf{x}^\star$ satisfies the expression $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in A$ in a neighborhood of $\mathbf{x}^\star$:

$$||\mathbf{x} - \mathbf{x}^\star|| \leq \delta, \delta > 0. \tag{3}$$

E. Eggeling and T. Ullrich are with Fraunhofer Austria Reseach GmbH, Inffeldgasse 16c, 8010 Graz, Austria; torsten.ullrich@fraunhofer.at

D. W. Fellner is with Institute of ComputerGraphics and KnowledgeVisualization (CGV), TU Graz, Austria and with GRIS, TU Darmstadt & Fraunhofer IGD, Darmstadt, Germany

Manuscript received August 31, 2012.

In other words, on some region around $\mathbf{x}^\star$ all function values are greater than or equal to the value at $\mathbf{x}^\star$. The occurrence of multiple extrema makes problem solving in (nonlinear) optimization very hard. The global (best) minimizer is difficult to obtain without supplying global information, which in turn is usually unavailable for a nontrivial case. Since there is no easy algebraic characterization of global optimality, global optimization is a difficult area, at least in higher dimensions for non-trivial problems.

As most global search techniques cannot ensure that a found solution $\mathbf{x}_0$ is a global minimum, the question whether $\mathbf{x}_0$ is only a local minimum or a global one remains open. In this context a probabilistic approach to estimate, whether a solution is the globally best solution, is a valuable tool. The presented approach does no more than that. It uses the function evaluations $f(\mathbf{x}_1), f(\mathbf{x}_2), f(\mathbf{x}_3), \ldots f(\mathbf{x}_m)$ returned by any global search method and analyzes the domain's sampling density. If the parameter domain $A$ is convex, and if the function $f$ is Lipschitz continuous, it is possible to determine the probability of the found solution being a global minimum. The Lipschitz constant is not needed to be known.

## II. Related Work

### A. Numerical Analysis

The presented approach is independent of the used algorithms and methods of global optimization which can be found in, for example, "Numerical Methods" [1], "Numerical Optimization" [2], "Introduction to Applied Optimization" [3], "Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation" [4], as well as in "Numerische Methoden der Analysis" (english: Numerical Methods of Analysis) [5] – just to name a few. Besides these introductions and overviews some books emphasize practical aspects – e.g. "Practical Optimization" [6], "Practical Methods of Optimization" [7], and "Global Optimization: Software, Test Problems, and Applications" [8].

Current overviews on the latest research results have been published in "Large-Scale Nonlinear Constrained Optimization: A Current Survey Algorithms for continuous optimization: the state of the art" [9], "Numerical methods for large-scale nonlinear optimization" [10], and in "Robust optimization – A comprehensive survey" [11].

### B. Verified Computing

Unfortunately, most global search techniques cannot ensure that a found solution is a global minimum. Very often the question, whether a better solution exists, remains open. An exception are verified computation techniques for global

optimization [12]. These techniques often use interval arithmetic [13] to calculate an optimal solution, respectively an inclusion, with a guaranteed error bound [14].

As a consequence, the global optimization methods can be classified into two groups: verified optimization with guaranteed results on the one hand and non-verified optimization without any indications on optimality on the other hand. The presented, probabilistic approach in this article is in-between. It determines the probability of a solution being a global minimum.

### C. Probability Theory

Probability theory is concerned with random phenomena and stochastic processes. As it is the mathematical foundation for statistics, probability theory is essential to many fields of applications that involve quantitative analysis of data [15], [16].

Let $\Omega$ be a nonempty set. Then $\mathcal{F}$ is called a $\sigma$-algebra, if

1) $\mathcal{F}$ contains the empty set $\emptyset \in \mathcal{F}$.
2) If $A$ is in $\mathcal{F}$, then its complement (in $\Omega$) $\overline{A} = \Omega \backslash A$ also belongs to $\mathcal{F}$.
3) For an arbitrary sequence $(A_n)_n$ of subsets of $\mathcal{F}$ the union $\bigcup_{i \in N} A_i$ is in $\mathcal{F}$.

In a sample space $\Omega$ each subset $A \subset \Omega$ which belongs to $\mathcal{F}$ is called an event and is associated with a probability measure $P$, which obeys the axioms of probability:

1) $\forall\, A \in \mathcal{F} \;:\; P(A) \geq 0$,
2) $P(\Omega) = 1$,
3) $P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$ for all sequences $(A_n)_{n \in \mathbb{N}}$ of pairwise disjunct events.

The triple $(\Omega, \mathcal{F}, P)$ is called probability space. The conditional probability of an event $A$ assuming that $B$ has occurred is denoted $P(A|B)$. It can be calculated via

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}. \tag{4}$$

If the conditional probability $P(A|B)$ of an event $A$ assuming event $B$ satisfies the equation $P(A|B) = P(A)$, the events $A$ and $B$ are called statistically independent. In this case

$$P(AB) = P(A \cap B) = P(A) \cdot P(B). \tag{5}$$

A real function whose domain is the probability space and for which

1) the set $\{X \leq x\}$ is an event for any real number $x$ and for which
2) the probability of the events $\{X = -\infty\}$ and $\{X = \infty\}$ equals zero,

is called random variable. Its probability distribution describes the range of possible values it can attain and the probability that the value of the random variable is within any measurable subset of that range.

A $\sigma$-algebra $\mathcal{F}$ is discrete, if a set of subsets $(A_i)_{i \in I}$ of $\Omega$ exists with

1) the index set $I$ is not empty $\emptyset \neq I \subset \mathbb{N}$,
2) $\forall\, i \neq j \;:\; A_i \cap A_j = \emptyset$, and
3) $\Omega = \bigcup_{i \in I} A_i$,

so that every element $A \in \mathcal{F}$ can be described by a union of some $A_j$. Furthermore, if the set $\{A_i | i \in I\}$ is finite, the probability space $(\Omega, \mathcal{F}, P)$ is called finite.

For random variables in discrete probability spaces the characteristic values of expectation and variance are of importance. If for a random variable $X : \Omega \to \mathbb{R}$ of a discrete probability space $(\Omega, \mathcal{F}, P)$ the sum $\sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) < \infty$ converges, the expectation value of $X$ is defined by

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}). \tag{6}$$

The variance of a probability variable $X$ of a discrete probability space $(\Omega, \mathcal{F}, P)$ is

$$V(X) = E((X - E(X))^2). \tag{7}$$

In stochastics and statistics it is convenient to omit brackets, if the short notation does not cause any confusion. The variance is then written $V(X) = E(X - EX)^2$. The square root of $V(X)$ is called standard deviation. It is noted

$$\sigma(X) = \sqrt{V(X)}. \tag{8}$$

The theory of probability and statistics contains many theorems which allow approximating various probability terms. According to the Tschebyshev inequality for any random variable $X$ with existing expectation value and variance, the equation

$$P\left(|X - EX| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \cdot V(X), \varepsilon > 0 \tag{9}$$

is satisfied.

The central limit theorem by JARL WALDEMAR LINDEBERG and PAUL LÉVY states that the sum of independent random variables will approach a normal distribution regardless of the distribution of the individual variables themselves. More precisely, if $(X_n)_{n \geq 1}$ is a sequence of independent and identically distributed random variables with positive, finite variance $\sigma^2 = V(X_1)$ and expectation value $\mu = EX_1$, the limit

$$\lim_{n \to \infty} P\left(a \leq \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a), \tag{10}$$

$(-\infty \leq a < b < \infty)$ converges to the differences $\Phi(b) - \Phi(a)$ of the cumulative Gaussian distribution function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, \mathrm{d}\, t$. According to HANS BANDEMER and ANDREAS BELLMANN this approximation by the standard normal distribution is practicable for sequences with $n > 30$, [17].

### D. Statistical Estimation

According to ERIC WEISSTEIN an estimate is an educated guess for an unknown quantity or outcome based on known information. The making of estimates is an important part of statistics, since care is needed to provide as accurate an estimate as possible using as little input data as possible. Often, an estimate for the uncertainty of an estimate can also be determined statistically. A rule that tells how to calculate an estimate based on the measurements contained in a sample is called an estimator [18].

More formal, the initial situation consists of a probability space $(\Omega, \mathcal{F}, P)$, a random variable $X$ and a realization (an observed value) $x$ of $X$; e.g. $x = X(\omega)$ for an element $\omega \in \Omega$. The distribution is not known completely. The class of distribution is assumed and its free parameters, which shall be estimated, are denoted by $\vartheta$. To indicate the dependency on the distribution, its expectation value, its variance, etc. are written $P_\vartheta$, $E_\vartheta$, and $V_\vartheta$. The set of all possible parameters $\vartheta$ is the parameter space $\Theta$. For some data $x$ the probability $P_\vartheta(X = x)$ can be interpreted as a function of $\vartheta$. This function

$$L_x : \begin{cases} \Theta & \to & [0,1] \\ \vartheta & \mapsto & P_\vartheta(X = x) \end{cases} \tag{11}$$

maps each parameter $\vartheta$ to the probability to obtain the observed data $x$. $L_x$ is called likelihood function. If $L_x$ reaches a maximum value $\hat{\vartheta}$ for each $x$, the function $\hat{\vartheta}(x)$ is called maximum likelihood estimator.

A maximum likelihood estimator is a so-called point estimator. Its estimate is a single point / a single datum. As the estimated value $T(x)$ of an estimator $T$ and the unknown parameter $\vartheta$ may differ significantly, confidence regions have been introduced: A confidence region $C(x)$ for $\vartheta$ is a subset of all possible parameters $\Theta$. Is $C(x)$ an interval in $\Theta \subset \mathbb{R}$ with endpoints $l(x)$ and $L(x)$, it is called confidence interval. Confidence intervals are a form of interval estimation. In contrast to point estimation it indicates the precision with which the parameter $\vartheta$ is estimated. A confidence interval $C(x) = [l(x), L(x)]$ for $\vartheta$ is said to have confidence level $1 - \alpha$, $0 < \alpha < 1$, if

$$P_\vartheta\left(\{x \in \Omega : \vartheta \in C(x)\}\right) \geq 1 - \alpha, \forall\, \vartheta \in \Theta \ . \tag{12}$$

For $n$ independent and identically distributed random variables $X_i$ with unknown mean $\mu$ and unknown variance $\sigma^2$, the sample mean $\overline{X}_n$ and the unadjusted sample variance can be calculated via

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2, \tag{13}$$

which are used to define interval estimators of mean and variance: The interval estimator of the mean $\mu$ is

$$T_n^\mu = \left[\overline{X}_n - \sqrt{\frac{S_n^2}{n}}\, z_0\,,\, \overline{X}_n + \sqrt{\frac{S_n^2}{n}}\, z_0\right] \tag{14}$$

with $0 < z_0$. The probability that the unknown mean $\mu$ is within $T_n^\mu$ is

$$P\left(\mu \in T_n^\mu\right) = P\left(-\sqrt{\frac{n-1}{n}}\, z_0 < Z_{n-1} < +\sqrt{\frac{n-1}{n}}\, z_0\right), \tag{15}$$

where $Z_{n-1}$ is a standard Student's $t$-random variable with $n-1$ degrees of freedom. Similarly, the interval estimator of the unknown variance is

$$T_n^{\sigma^2} = \left[\frac{n}{z_2}\, S_n^2\,,\, \frac{n}{z_1}\, S_n^2\right] \tag{16}$$

with $0 < z_1 < z_2$. The probability that the unknown variance is within $T_n^{\sigma^2}$ is

$$P\left(\sigma^2 \in T_n^{\sigma^2}\right) = P\left(z_1 < Z < z_2\right), \tag{17}$$

where $Z$ is a $\chi^2$ random variable with $n$ degrees of freedom. For practical applications, density functions and their quantiles are tabulated in standard textbooks [19]. If only an upper or lower bound is needed, asymmetric quantiles can be chosen.

## III. CO-DOMAIN ESTIMATION

Let $f$ be a real-valued, Lipschitz continuous function defined on a limited, convex set $A \subset \mathbb{R}^n$. During an optimization process, a sequence of parameters $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and corresponding evaluations $f(\mathbf{x}_1)$, $f(\mathbf{x}_2)$, ..., $f(\mathbf{x}_m)$ have led to a minimum candidate $\mathbf{x}_0$. The probability of $\mathbf{x}_0$ being a global minimum can be estimated in a two-step process.

1) For each pair $(\mathbf{x}_k, f(\mathbf{x}_k))$ exists a confidence region $R_k$ with confidence level $1 - \alpha_k$, which does not contain a parameter with a function value less than the minimum candidate; i.e.

$$P\left(\exists \mathbf{x} \in R_k : f(\mathbf{x}) < f(\mathbf{x}_0)\right) = \alpha. \tag{18}$$

2) With the assumption that the distribution of global minima is uniform, the probability of globality is

$$P\left(\forall \mathbf{x} \in A : f(\mathbf{x}) > f(\mathbf{x}_0)\right) = \frac{\sum_{i=1}^{m}(1-\alpha_i)\cdot \mathrm{vol}(R_i)}{\mathrm{vol(A)}}, \tag{19}$$

with disjunct regions $R_i$, whose volume is $\mathrm{vol}(R_i)$. Due to the limitation of $A$, the volume $\mathrm{vol}(A) < \infty$ as well as the volumes of its subsets $R_i$ exist.

### A. Lipschitz Estimator

In local coordinates with origin at $(\mathbf{x}_k, f(\mathbf{x}_k))$ the sequence of parameters $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_m$ and their corresponding function values can be described by local vectors

$$\Delta_{k,i} = \mathbf{x}_i - \mathbf{x}_k \quad \text{and} \quad f_{k,i}^\Delta = f(\mathbf{x}_i) - f(\mathbf{x}_k). \tag{20}$$

As $\mathbf{x}_0$ is a minimum candidate, $f_{k,0}^\Delta$ is a lower bound of all $f_{k,i}^\Delta, i = 1, \ldots, m$. Furthermore, as $f$ is assumed to be Lipschitz continuous, the quotient of function values and parameters is limited:

$$\exists\, L \in \mathbb{R}, L \geq 0 \ : \ \frac{|f_{k,i}^\Delta|}{||\Delta_{k,i}||} < L. \tag{21}$$

This bound $L$ can be approximated via interval estimators $T_m$ of mean and variance of $|f^{\Delta}_{k,i}| \,/\, ||\Delta_{k,i}||$ and the fact that within normally distributed values, realizations of random variables are within the intervals $\mu \pm \sigma$ (in 68.3%), $\mu \pm 2\sigma$ (in 95.4%), $\mu \pm 3\sigma$ (in 99.7%), respectively, $\mu \pm z\sigma$ (in $\Phi_{0;1}(z) - \Phi_{0;1}(-z)$ with $\Phi_{0;1}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \,\mathrm{d}t$) using the standard deviation $\sigma = \sqrt{V(f^{\Delta})}$.

Consequently, $L$ can be limited probabilistically by the combined probability of the estimators confidence levels and the distribution of function value realizations: e.g. to ensure a confidence level of $92.69\%$ for $L$ the interval estimators for $\mu$, $\sigma$, and $\mu \pm z\sigma$ may have each a confidence level of $97.5\%$, which would lead – as a conditional probability – to the desired result

$$\underbrace{(1 - 0.0731)}_{=0.9269} = \underbrace{(1 - 0.025)}_{=0.975} \cdot (1 - 0.025) \cdot (1 - 0.025). \quad (22)$$

### B. Probability

Having an probabilistic, upper bounded Lipschitz constant $L_k$ with combined probability $1 - \alpha_k$, the function $f$ can be limited probabilistically by a linear function $f^L$. Then, the region $R_k$ is defined by a sphere with radius $r_k$, so that all values $\mathbf{x} \in R_k$ have $f^L(\mathbf{x}_k) > f((x)_0)$. As $L_k$ is finite, $f(\mathbf{x}_k) = f^L(\mathbf{x}_k)$, and $f(\mathbf{x}_k) > f(\mathbf{x}_0)$, the sphere $R_k$ has a positive radius. If all regions $R_i$ are pairwise disjunct, the probability of globality is

$$P\left(\forall \mathbf{x} \in A : f(\mathbf{x}) > f(\mathbf{x}_0)\right) = \frac{\sum_{i=1}^{m} (1 - \alpha_i) \cdot \mathrm{vol}(R_i)}{\mathrm{vol(A)}}. \quad (23)$$

## IV. Implementation

### A. Lipschitz Estimator

The mean and variance estimators of $|f^{\Delta}_{k,i}| \,/\, ||\Delta_{k,i}||$ do not utilize the complete set of function evaluations but a sufficient subset (with cardinality $> 30$, see [17]) of nearest neighbors for each $(\mathbf{x}_k, f(\mathbf{x}_k))$. The number of nearest neighbors has two contradicting effects: on the one hand, the more neighbors the estimation uses, the better the estimation result; on the other hand, the more neighbors the estimation uses, the worse the approximation of $f$'s local properties becomes (i.e. areas, in which $f$ is constantly zero, becomes approximated by a large Lipschitz constant underestimating the region $R_k$).

This problem can be solved by a greedy algorithm, that starts with a statistically significant number of nearest neighbors and expands the neighborhood radius as long as $(1-\alpha_i) \cdot \mathrm{vol}(R_i)$ increases.

### B. Probability

The calculation of the combined probabilities relies on pairwise disjunct regions $R_i$. Our implementation uses $\varepsilon$-sets [20] and discrete oriented polytopes [21] to restrict and limit overlapping regions; i.e. each region is represented by its center $\mathbf{x}_k \in \mathbb{R}^n$ and a fixed set of rays. The ray directions are precalulated by a spherical code distribution [22], [23], [24]. Along each ray the maximum region expansion is determined and stored as a simple scalar value: the expansion along a direction is limited by the region's radius, the domain's boundary $\partial A$ and by neighboring regions. In this way, all geometric calculations can be reduced to ray-intersections.

## V. Conclusion

The global (best) minimizer is difficult to obtain without supplying global information, which in turn is usually unavailable for a nontrivial case. Due to the fact that there is no easy algebraic characterization of global optimality, global optimization is a difficult area, at least in higher dimensions. In Equation (19) this fact is quantified: in practice the domain $A$ is very often a hypercube (or without loss of generality, a scaled hypercube). While the hypercube's volume is constant and therefore independent of the search space's dimension $n$, the regions $R_i$ have a hyperspherical shape $S$ whose unit volume is $\mathrm{vol}(S) = \pi^{n/2}/\Gamma(\frac{n}{2} + 1)$. This function has a maximum at $\approx 5.28$ and converges to zero for increasing dimensions $n$. As a consequence, the probability to find the globally best solution converges to zero as well.

### A. Contribution

Most global search techniques cannot ensure that a found solution $\mathbf{x}_0$ is a global minimum. Consequently, the question whether $\mathbf{x}_0$ is only a local minimum or a global one remains open. In this context, the presented approach gives a probabilistic answer. It uses the function evaluations $f(\mathbf{x}_1), f(\mathbf{x}_2), f(\mathbf{x}_3), \ldots f(\mathbf{x}_m)$ returned by any global search method and analyzes the domain's sampling density. If the parameter domain $A$ is convex, and if the function $f$ is Lipschitz continuous, it determines the probability of the found solution being a global minimum (without having to know the Lipschitz constant).

### B. Benefit

The relation between undersampled regions and regions of high confidence level does not only determine the probability of globality; this probabilistic framework offers a new kind of analysis for the results of optimization processes. It can be interpreted as a probabilistic convergence criterion, according to which an optimization technique has to process the search space until an a-priori confidence level is reached. Moreover, it can identify undersampled regions within the search space in order to guide an optimization process. Especially stochastic- and random-based optimization processes – such as Differential Evolution [25], genetic algorithms [26], etc. – may benefit.

## REFERENCES

[1] W. Boehm and H. Prautzsch, *Numerical Methods*, W. Boehm and H. Prautzsch, Eds. Vieweg, 1993.
[2] J. Nocedal and S. J. Wright, *Numerical Optimization*, J. Nocedal and S. J. Wright, Eds. Springer, 1999.
[3] U. Diwekar, *Introduction to Applied Optimization*, ser. Applied Optimization, U. Diwekar, Ed. Springer, 2003, vol. 80.
[4] J. C. Nash, *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, second edition ed., J. C. Nash, Ed. Adam Hilger, 1990.
[5] K. Höllig, J. Höner, and M. Pfeil, *Numerische Methoden der Analysis*, K. Höllig, J. Höner, and M. Pfeil, Eds. Mathematik-Online, 2010.
[6] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, P. E. Gill, W. Murray, and M. H. Wright, Eds. Academic Press, 1982.
[7] R. Fletcher, *Practical Methods of Optimization*, R. Fletcher, Ed. Wiley, 2000.
[8] J. D. Pinter, "Global Optimization: Software, Test Problems, and Applications," *Handbook of Global Optimization, P.M. Pardalos and H.E. Romeijn (eds)*, vol. 2, pp. 515–569, 2002.
[9] A. Conn, N. I. M. Gould, and P. L. Toint, "Large-Scale Nonlinear Constrained Optimization: A Current Survey," *Algorithms for continuous optimization: the state of the art*, vol. 434, pp. 287–332, 1994.
[10] N. Gould, D. Orban, and P. Toint, "Numerical methods for large-scale nonlinear optimization," *Acta Numerica*, vol. 14, pp. 299–361, 2005.
[11] H.-G. Beyer and B. Sendhoff, "Robust optimization – A comprehensive survey," *Computer Methods in Applied Mechanics and Engineering*, vol. 196, pp. 3190–3218, 2007.
[12] M. Kieffer, M. C. Markót, H. Schichl, and E. Walter, "Verified global optimization for estimating the parameters of nonlinear models," *Modeling, Design, and Simulation of Systems with Uncertainties*, vol. 1, pp. 129–151, 2011.
[13] R. Hammer, M. Hocks, U. Kulisch, and D. Ratz, *C++ Toolbox for Verified Computing*, R. Hammer, M. Hocks, U. Kulisch, and D. Ratz, Eds. Springer, 1997.
[14] M. C. Markót and H. Schichl, "Comparison and Automated Selection of Local Optimization Solvers for Interval Global Optimization Methods," *SIAM Journal on Optimization*, vol. 21, pp. 1371–1391, 2011.
[15] N. Henze, *Stochastik – Einführung in die Wahrscheinlichkeitstheorie und Statistik (english: Stochastics – Introduction to probability calculus and statistics)*, N. Henze, Ed. Technische Universität Karlsruhe, 1995.
[16] ——, *Stochastik für Einsteiger (english: Stochastics for Beginners)*, N. Henze, Ed. Vieweg, 1997.
[17] H. Bandemer and A. Bellmann, *Statistische Versuchsplanung (english: Statistical Test Planning)*, H. Bandemer and A. Bellmann, Eds. Verlag H. Deutsch / BSB B. G. Teubner, 1979.
[18] E. Weisstein, *MathWorld – A Wolfram Web Resource*, E. Weisstein, Ed. Wolfram Research, 2009.
[19] M. P. McLaughlin, *A Compendium of Common Probability Distributions*, M. P. McLaughlin, Ed. Regress+, 1999.
[20] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, "Geometric Approximations via Coresets," *Combinatorial and Computational Geometry - MSRI Publications*, vol. 52, pp. 1–30, 2005.
[21] G. Zachmann, "Rapid Collision Detection by Dynamically Aligned DOP-Trees," *Proceedings of the Virtual Reality Annual International Symposium*, vol. 1, pp. 90–97, 1998.
[22] J. D. Pinter, "Globally Optimized Spherical Point Arrangements: Model Variants and Illustrative Results," *Annals of Operations Research*, vol. 104, pp. 213–230, 2001.
[23] A. Katanforoush and M. Shahshahani, "Distributing Points on a Sphere," *Experimental Mathematics*, vol. 12, pp. 199–208, 2003.
[24] C. Schinko, T. Ullrich, and D. W. Fellner, "Simple and Efficient Normal Encoding with Error Bounds," *Proceedings of Theory and Practice of Computer Graphics*, vol. 29, pp. 63–66, 2011.
[25] R. Storn and K. Price, "Differential Evolution: A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.
[26] Z. Michalewicz and M. Schoenauer, "Evolutionary Algorithms for Constrained Parameter Optimization Problems," *Evolutionary Computation*, vol. 4, pp. 1–32, 1996.

**Eva Eggeling** Eva Eggeling holds a PhD in Applied Mathematics from University of Cologne. She has developed numerical simulations in the application fields of meteorological data assimilation and grain growth simulation for polycrystals. She has experience in modeling and simulation in virtual reality, visualization of scientific data and differential and parametric optimization. Eva Eggeling is head of the business area Visual Computing of Fraunhofer Austria.

**Dieter W. Fellner** Dieter W. Fellner is the director of the Fraunhofer Institute of Computer Graphics and a professor of computer science at the Darmstadt University of Technology, with a joint affiliation with the Graz University of Technology. His research interests include computer graphics, modeling, and graphics in digital libraries. Fellner has a PhD from the Graz University of Technology. He's a member of the ACM, Eurographics, and the Gesellschaft für Informatik. He's also on the editorial board of IEEE Computer Graphics and Applications.

**Torsten Ullrich** Torsten Ullrich is working for Fraunhofer Austria. His research involves computer-aided geometric-design topics, including modeling and numerical optimization. He has a PhD in computer science from the Graz University of Technology and an MSc in mathematics from the Karlsruhe Institute of Technology.