UNIVERSITY OF CAMBRIDGE

Numerical Analysis Reports

A radial basis function method for global optimization

H.-M. Gutmann

DAMTP 1999/NA22
December, 1999

Department of Applied Mathematics and Theoretical Physics
Silver Street
Cambridge CB3 9EW
England

DAMTP 1999/NA22

A radial basis function method for global optimization

H.-M. Gutmann

Abstract:
We introduce a method that aims to find the global minimum of a continuous nonconvex function on a compact subset of $\mathbb{R}^d$. It is assumed that function evaluations are expensive and that no additional information is available. Radial basis function interpolation is used to define a utility function. The maximizer of this function is the next point where the objective function is evaluated. We show that, for most types of radial basis functions that are considered in this paper, convergence can be achieved without further assumptions on the objective function. Besides, it turns out that our method is closely related to a statistical global optimization method, the P-algorithm. A general framework for both methods is presented. Finally, a few numerical examples show that on the set of Dixon-Szegö test functions our method yields favourable results in comparison to other global optimization methods.

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge,
Silver Street,
Cambridge CB3 9EW,
England.

December, 1999.

1 Introduction

Global optimization has attracted a lot of attention in the last 20 years. In many applications, the objective function is nonlinear and nonconvex. Often, the number of local minima is large. Therefore standard nonlinear programming methods may fail to locate the global minimum.

In the most general way, the Global Optimization Problem can be stated as

$$\text{(GOP)} \qquad \text{find} \quad x^* \in \mathcal{D} \quad \text{such that} \quad f(x^*) \leq f(x),\ x \in \mathcal{D},$$

where $\mathcal{D} \subset I\!R^d$ is compact, and $f : \mathcal{D} \to I\!R$ is a continuous function defined on $\mathcal{D}$. Under these assumptions, (GOP) is solvable, because f attains its minimum on $\mathcal{D}$.

Numerous methods to solve (GOP) have been developed (see e.g. Horst and Pardalos [4] and Törn and Žilinskas [19]). Stochastic methods like simulated annealing and genetic algorithms which use only function values are very popular among users, although their rate of convergence is usually rather slow. Deterministic methods like Branch-and-Bound, however, assume that one can compute a lower bound of f on a subset of $\mathcal{D}$. This can be done, for example, when the Lipschitz constant on f is available. The further assumptions make these methods very powerful, but often they are not satisfied or it is too expensive to provide the necessary information.

For the method investigated in this paper, we have in mind problems when the only information available is the possibility to evaluate the objective function, and each evaluation is very expensive. This may mean that it takes several hours to calculate a function value. For example, a function evaluation at a point may be done by building an experiment, by running a long computer simulation or by using a finite element method. Therefore, as the duration of the optimization process is dominated by the function evaluations, our goal is to require as few function evaluations as possible to find an adequate estimate of the global minimum.

The method is based on a general technique proposed by Jones [8]. Let $\mathcal{A}$ be a linear space of functions, and assume that, for $s \in \mathcal{A}$, $\sigma(s)$ is a measure of the "bumpiness" of s. Now assume that we have calculated $x_1, \ldots, x_n$ and the function values $f(x_1), \ldots, f(x_n)$. A target value f^* is chosen that can be regarded as an estimate of the optimal value, but it might be very crude. For each $y \notin \{x_1, \ldots, x_n\}$, let $s_y \in \mathcal{A}$ be defined by the interpolation conditions

$$\begin{aligned} s_y(x_i) &= f(x_i), \qquad i = 1, \ldots, n, \\ s_y(y) &= f^*. \end{aligned} \tag{1.1}$$

The new point x_{n+1} is chosen to be the value of y that minimizes $\sigma(s_y)$, $y \notin \{x_1, \ldots, x_n\}$. Thus the view is taken that the "least bumpy" of the functions s_y

yields the most useful location of the new point.

We will use radial basis functions as interpolants. Their interpolation properties are very suitable. Specifically, the uniqueness of an interpolant is achieved under very mild conditions on the location of the interpolation points, and a measure of bumpiness is also available.

Close relations can be established between our method and one from statistical global optimization, namely the P-algorithm (Žilinskas [22]). Although being derived using a completely different approach, it is very similar to our method. One special case of a P-algorithm, developed by Kushner [12], is even equivalent to a special case of our radial basis function method.

Other global optimization methods based on radial basis functions have been developed. Alotto et. al. [1] use interpolation by multiquadrics to accelerate a simulated annealing method. Ishikawa et. al. [6], [7] employ radial basis functions to estimate the global minimizer and run an SQP algorithm to locate it.

The properties of radial basis functions that are necessary for the description of our method are introduced in the following section. In particular, we address the question of interpolation and introduce a suitable measure of "bumpiness". The global optimization method is described in detail in Section 3. Convergence of the method is the subject of Section 4. Subsection 4.1 contains various convergence results, and the proof of the main theorem can be found in Subsection 4.2. The relation between our method and the P-algorithm is addressed in Section 5. The final section deals with search strategies, but a complete analysis is beyond the scope of this paper.

2 Interpolation by radial basis functions and a measure of bumpiness

Let n pairwise different points $x_1, \ldots, x_n \in \mathbb{R}^d$ and data $f_1, \ldots, f_n \in \mathbb{R}$ be given, where n and d are any positive integers. We seek a function s of the form

$$s(x) = \sum_{i=1}^{n} \lambda_i \phi(\|x - x_i\|) + p(x), \quad x \in \mathbb{R}^d, \tag{2.1}$$

that interpolates the data $(x_1, f_1), \ldots, (x_n, f_n)$. The coefficients λ_i, $i = 1, \ldots, n$, are real numbers, and p is from Π_m, the space of polynomials of degree less than or equal to m. The norm $\|.\|$ is the Euclidean norm in $\mathbb{R}^d$. The following choices of ϕ are considered:

$$\left.\begin{array}{lcll} \phi(r) & = & r & \text{(linear)}, \\ \phi(r) & = & r^3 & \text{(cubic)}, \\ \phi(r) & = & r^2 \log r & \text{(thin plate spline)}, \\ \phi(r) & = & \sqrt{r^2 + \gamma^2} & \text{(multiquadric)}, \\ \phi(r) & = & e^{-\gamma r^2} & \text{(Gaussian)}, \end{array}\right\} \quad r \geq 0, \tag{2.2}$$

where γ is a prescribed positive constant. Let the matrix $\Phi \in \mathbb{R}^{n\times n}$ be defined by

$$(\Phi)_{ij} := \phi(\|x_i - x_j\|), \quad i,j = 1,\ldots,n. \tag{2.3}$$

Further, we introduce the linear space $\mathcal{V}_m \subset \mathbb{R}^n$ containing all $\lambda \in \mathbb{R}^n$ that satisfy

$$\sum_{i=1}^{n} \lambda_i q(x_i) = 0 \quad \forall\, q \in \Pi_m. \tag{2.4}$$

Formally, we set $\mathcal{V}_{-1} := \mathbb{R}^n$. Obviously, $\mathcal{V}_{m+1} \subset \mathcal{V}_m$ for all $m \geq -1$. Powell [15] shows that, in the cubic and thin plate spline cases

$$\lambda^T \Phi \lambda > 0 \quad \forall\, \lambda \in \mathcal{V}_1 \setminus \{0\}, \tag{2.5}$$

in the linear and multiquadric cases

$$\lambda^T \Phi \lambda < 0 \quad \forall\, \lambda \in \mathcal{V}_0 \setminus \{0\}, \tag{2.6}$$

and in the Gaussian case

$$\lambda^T \Phi \lambda > 0 \quad \forall\, \lambda \in \mathbb{R}^n \setminus \{0\}. \tag{2.7}$$

We let m_0 be 1 in the cubic and thin plate spline cases, 0 in the linear and multiquadric cases and -1 in the Gaussian case. Then the inequalities (2.5) – (2.7) can be merged into

$$(-1)^{m_0+1} \lambda^T \Phi \lambda > 0 \quad \forall\, \lambda \in \mathcal{V}_{m_0} \setminus \{0\}. \tag{2.8}$$

After choosing ϕ, we let m be an integer that is not less than m_0, and λ is confined to $\mathcal{V}_m$.

Let $\hat{m}$ be the dimension of Π_m, let $p_1, \ldots, p_{\hat{m}}$ be a basis of this linear space, and let P be the matrix

$$P := \begin{pmatrix} p_1(x_1) & \cdots & p_{\hat{m}}(x_1) \\ \vdots & & \vdots \\ p_1(x_n) & \cdots & p_{\hat{m}}(x_n) \end{pmatrix}. \tag{2.9}$$

Then it can be shown (see [15]) that the matrix

$$A = \begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \in \mathbb{R}^{(n+\hat{m})\times(n+\hat{m})} \tag{2.10}$$

is nonsingular if and only if $x_1, \ldots, x_n$ satisfy

$$q \in \Pi_m \quad \text{and} \quad q(x_i) = 0, \quad i = 1,\ldots,n, \quad \Longrightarrow \quad q \equiv 0. \tag{2.11}$$

In the Gaussian case with $m = -1$, P and condition (2.11) are omitted. Therefore the coefficients of the function s in (2.1) are defined uniquely by the system

$$\left.\begin{array}{rcl} s(x_i) & = & f_i, \quad i = 1, \dots, n \\ \sum_{i=1}^{n} \lambda_i p_j(x_i) & = & 0, \quad j = 1, \dots, \hat{m} \end{array}\right\}. \tag{2.12}$$

Let F be the vector whose entries are the data values $f_1, \dots, f_n$. Then the system (2.12) becomes

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} F \\ 0_{\hat{m}} \end{pmatrix}, \tag{2.13}$$

where $\lambda = (\lambda_1, \dots, \lambda_n)^T \in \mathbb{R}^n$, $c \in \mathbb{R}^{\hat{m}}$ and $0_{\hat{m}}$ is the zero in $\mathbb{R}^{\hat{m}}$. The components of c are the coefficients of the polynomial p with respect to the basis $p_1, \dots, p_{\hat{m}}$.

The motivation for the measurement of the bumpiness of a radial basis function interpolant can be developed from the theory of natural cubic splines in one dimension. They can be written in the form (2.1), where $\phi(r) = r^3$, $\lambda \in \mathcal{V}_1$ and $p \in \Pi_1$. It is well known (e.g. Powell [14]) that the interpolant s that is defined by the system (2.12) minimizes $I(g) := \int_{\mathbb{R}} [g''(x)]^2 dx$ among all functions $g : \mathbb{R} \to \mathbb{R}$ that satisfy the interpolation conditions $g(x_i) = f_i$, $i = 1, \dots, n$, and for which $I(g)$ exists and is finite. Therefore $I(g)$ is a suitable measure of bumpiness. The second derivative s'' is piecewise linear and vanishes outside a bounded interval. Thus one obtains by integration by parts

$$\begin{aligned} I(s) &= \int_{\mathbb{R}} [s''(x)]^2 \, dx = 12 \sum_{i=1}^{n} \lambda_i s(x_i) \\ &= 12 \sum_{i=1}^{n} \lambda_i \left(\sum_{j=1}^{n} \lambda_j |x_i - x_j|^3 + p(x_i) \right) = 12 \sum_{i,j=1}^{n} \lambda_i \lambda_j \Phi_{ij}, \end{aligned}$$

where the last equation follows from $\lambda \in \mathcal{V}_1$. This relation suggests that expression (2.8) can provide a semi-inner product and a semi-norm for each ϕ in (2.2) and $m \geq m_0$. Also, the semi-norm will be the measure of bumpiness of a radial basis function (2.1). A semi-inner product $\langle ., . \rangle$ satisfies the same properties as an inner product, except that $\langle s, s \rangle = 0$ need not imply $s = 0$. Similarly, for a semi-norm $\|.\|$, $\|s\| = 0$ does not imply $s = 0$.

We choose any radial basis function from (2.2) and $m \geq m_0$, and we define $\mathcal{A}_{\phi,m}$ to be the linear space of all functions of the form

$$\sum_{i=1}^{N} \lambda_i \phi(\|x - y_i\|) + p(x), \quad x \in \mathbb{R}^d,$$

where $N \in \mathbb{N}$, $y_1, \dots, y_N \in \mathbb{R}^d$, $p \in \Pi_m$, and $\lambda = (\lambda_1, \dots, \lambda_N)^T$ satisfies (2.4) for $n = N$. On this space, the semi-inner product and the semi-norm are defined as follows. Let s and u be any functions in $\mathcal{A}_{\phi,m}$, i.e.

$$s(x) = \sum_{i=1}^{N(s)} \lambda_i \phi(\|x - y_i\|) + p(x) \quad \text{and} \quad u(x) = \sum_{j=1}^{N(u)} \mu_j \phi(\|x - z_j\|) + q(x).$$

We let the semi-inner product be the expression

$$\langle s, u \rangle := (-1)^{m_0+1} \sum_{i=1}^{N(s)} \lambda_i u(y_i). \tag{2.14}$$

Clearly, it is bilinear. To show symmetry, we use

$$\sum_{i=1}^{N(s)} \lambda_i q(y_i) = 0 \quad \text{and} \quad \sum_{j=1}^{N(u)} \mu_j p(z_j) = 0,$$

to deduce

$$\begin{aligned}
\langle s, u \rangle &= (-1)^{m_0+1} \sum_{i=1}^{N(s)} \lambda_i \left(\sum_{j=1}^{N(u)} \mu_j \phi(\|y_i - z_j\|) + q(y_i) \right) \\
&= (-1)^{m_0+1} \sum_{i=1}^{N(s)} \sum_{j=1}^{N(u)} \lambda_i \mu_j \phi(\|y_i - z_j\|) \\
&= (-1)^{m_0+1} \sum_{j=1}^{N(u)} \mu_j \left(\sum_{i=1}^{N(s)} \lambda_i \phi(\|z_j - y_i\|) + p(z_j) \right) \\
&= (-1)^{m_0+1} \sum_{j=1}^{N(u)} \mu_j s(z_j) = \langle u, s \rangle.
\end{aligned}$$

By (2.8),

$$\langle s, s \rangle = (-1)^{m_0+1} \sum_{i=1}^{N(s)} \lambda_i s(y_i) = (-1)^{m_0+1} \sum_{i,j=1}^{N(s)} \lambda_i \lambda_j \phi(\|y_i - y_j\|) \tag{2.15}$$

is strictly positive, if $\lambda \in \mathcal{V}_m \setminus \{0\}$ and $m \geq m_0$, i.e. $s \in \mathcal{A}_{\phi,m}$ is not a polynomial in Π_m. Thus (2.14) is a semi-inner product on $\mathcal{A}_{\phi,m}$ that induces the semi-norm $\langle s, s \rangle$ with null space Π_m (for details see Powell [16] and Schaback [17]).

In analogy to the variational principle for cubic splines in one dimension, mentioned above, there is a theorem that states that the given interpolant is the solution to a minimization problem.

Theorem 1 (Schaback [17]) *Let ϕ be any radial basis function from (2.2), and let m be chosen such that $m \geq m_0$. Given are points $x_1, \ldots, x_n$ in $I\!R^d$ having the property (2.11) and values $f_1, \ldots, f_n$ in $I\!R$. Let s be the radial function of the form (2.1) that solves the system (2.12). Then s minimizes the semi-norm $\langle g, g \rangle^{1/2}$ on the set of functions $g \in \mathcal{A}_{\phi,m}$ that satisfy*

$$g(x_i) = f_i, \quad i = 1, \ldots, n. \tag{2.16}$$

3 A radial basis function method

It will be shown how radial basis functions can be used in the general method of Jones [8] to solve the problem (GOP) (cf. Powell [16]). As in Section 2, we pick ϕ from (2.2) and $m \geq m_0$. Let $p_1, \ldots, p_{\hat{m}}$ be a basis of Π_m, where $\hat{m} = \dim \Pi_m$. Assume we have chosen $x_1, \ldots, x_n \in \mathcal{D}$ that satisfy (2.11), and we know the function values $f(x_1), \ldots, f(x_n)$. Let the function

$$s_n(x) = \sum_{i=1}^{n} \lambda_i \phi(\|x - x_i\|) + p(x), \quad x \in I\!R^d,$$

interpolate $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$. Our task is to determine x_{n+1}. For a target value f_n^* and a point $y \in \mathcal{D} \setminus \{x_1, \ldots, x_n\}$ the radial basis function s_y that satisfies (1.1) can be written as

$$s_y(x) = s_n(x) + [f_n^* - s_n(y)]\, \ell_n(y, x), \quad x \in I\!R^d, \tag{3.1}$$

where $\ell_n(y, x)$ is the radial basis function solution to the interpolation conditions

$$\begin{aligned} \ell_n(y, x_i) &= 0, \qquad i = 1, \ldots, n, \\ \ell_n(y, y) &= 1. \end{aligned} \tag{3.2}$$

Therefore $\ell_n(y, .)$ can be expressed as

$$\ell_n(y, x) = \sum_{i=1}^{n} \alpha_i(y) \phi(\|x - x_i\|) + \mu_n(y) \phi(\|x - y\|) + \sum_{i=1}^{\hat{m}} b_i(y) p_i(x), \quad x \in I\!R^d, \tag{3.3}$$

where the coefficients of $\ell_n(y, .)$ are defined by the equations

$$A(y) \begin{pmatrix} \alpha(y) \\ \mu_n(y) \\ b(y) \end{pmatrix} = \begin{pmatrix} 0_n \\ 1 \\ 0_{\hat{m}} \end{pmatrix}. \tag{3.4}$$

Here $\alpha(y) = (\alpha_1(y), \ldots, \alpha_n(y))^T \in I\!R^n$, $b(y) = (b_1(y), \ldots, b_{\hat{m}}(y))^T \in I\!R^{\hat{m}}$, $\mu_n(y) \in I\!R$, 0_n and $0_{\hat{m}}$ denote the zero in $I\!R^n$ and $I\!R^{\hat{m}}$, respectively, and, as in equation (2.10), $A(y)$ has the form

$$A(y) := \begin{pmatrix} \Phi(y) & P(y) \\ P(y)^T & 0_{\hat{m}\times\hat{m}} \end{pmatrix}. \tag{3.5}$$

Specifically, letting $u(y)$ and $\pi(y)$ be the vectors

$$u(y) := (\phi(\|y - x_1\|), \ldots, \phi(\|y - x_n\|))^T$$

and

$$\pi(y) := (p_1(y), \ldots, p_{\hat{m}}(y))^T,$$

respectively, $\Phi(y)$ and $P(y)$ are the matrices

$$\Phi(y) := \begin{pmatrix} \Phi & u(y) \\ u(y)^T & \phi(0) \end{pmatrix} \quad \text{and} \quad P(y) := \begin{pmatrix} P \\ \pi(y)^T \end{pmatrix}. \tag{3.6}$$

The square of the semi-norm $\langle s_y, s_y \rangle$ of the new interpolant (3.1), as defined in the previous section, has the value

$$\langle s_y, s_y \rangle = \langle s_n, s_n \rangle + 2[f_n^* - s_n(y)]\langle s_n, \ell_n(y,.)\rangle + [f_n^* - s_n(y)]^2 \langle \ell_n(y,.), \ell_n(y,.)\rangle.$$

Equations (2.14) and (3.2) imply

$$\langle s_n, \ell_n(y,.)\rangle = (-1)^{m_0+1} \sum_{i=1}^{n} \lambda_i \ell_n(y, x_i) = 0,$$

and, using expressions (3.2) and (3.3), we find the expression

$$\begin{aligned} \langle \ell_n(y,.), \ell_n(y,.)\rangle &= (-1)^{m_0+1} \left[\sum_{i=1}^{n} \alpha_i(y)\ell_n(y, x_i) + \mu_n(y)\ell_n(y, y) \right] \\ &= (-1)^{m_0+1}\mu_n(y). \end{aligned} \tag{3.7}$$

Thus we deduce the formula

$$\langle s_y, s_y \rangle = \langle s_n, s_n \rangle + (-1)^{m_0+1}\mu_n(y)\,[f_n^* - s_n(y)]^2.$$

Further, we define the function $g_n : \mathcal{D} \setminus \{x_1, \ldots, x_n\} \to I\!R$ as the difference

$$g_n(y) := \langle s_y, s_y \rangle - \langle s_n, s_n \rangle = (-1)^{m_0+1}\mu_n(y)\,[f_n^* - s_n(y)]^2,$$

which is nonnegative. Since $\langle s_n, s_n \rangle$ is independent of y, the required minimization of $\langle s_y, s_y \rangle$ and the minimization of $g_n(y)$ are equivalent.

The choice of f_n^* determines the location of x_{n+1}. If

$$f_n^* \geq \min_{y\in\mathcal{D}} s_n(y),$$

then $g_n(y) = 0$ can be achieved. However, if

$$f_n^* < \min_{y\in\mathcal{D}} s_n(y),$$

then x_{n+1} will be away from the $x_i,\ i = 1, \dots, n$. In particular, for $f_n^* \to -\infty$, we make the following deduction.

Remark 2 *For $f_n^* < \min_{y\in\mathcal{D}} s_n(y)$ let $x(f_n^*)$ be the minimizer of g_n, i.e.*

$$(-1)^{m_0+1}\mu_n(x(f_n^*))[s_n(x(f_n^*)) - f_n^*]^2 \leq (-1)^{m_0+1}\mu_n(y)[s_n(y) - f_n^*]^2 \quad \forall\, y \in \mathcal{D} \setminus \{x_1, \dots, x_n\}.$$

This is equivalent to

$$(-1)^{m_0+1}\mu_n(x(f_n^*)) \leq (-1)^{m_0+1}\mu_n(y)\left[1 + \frac{s_n(y) - s_n(x(f_n^*))}{s_n(x(f_n^*)) - f_n^*}\right]^2.$$

As $f_n^* \to -\infty$, the boundedness of s_n on $\mathcal{D}$ implies

$$(-1)^{m_0+1}\mu_n(x(-\infty)) \leq (-1)^{m_0+1}\mu_n(y) \quad \forall\, y \in \mathcal{D} \setminus \{x_1, \dots, x_n\}.$$

Therefore, the choice $f_n^* = -\infty$ requires the minimization of $(-1)^{m_0+1}\mu_n(y)$. This process puts x_{n+1} in a large gap between $x_i,\ i = 1, \dots, n$, a property that is of fundamental importance to global optimization.

The following basic algorithm employs the given method.

Algorithm 3

Initial step: *Pick ϕ from (2.2) and $m \geq m_0$.*
Choose points $x_1, \dots, x_{n_0} \in \mathcal{D}$ that satisfy (2.11). Compute the radial function s_n that minimizes $\langle s, s\rangle$ on $\mathcal{A}_{\phi,m}$, subject to the interpolation conditions

$$s(x_i) = f(x_i),\ i = 1, \dots, n.$$

Iteration step: *$x_1, \dots, x_n$ are the points in $\mathcal{D}$ where the value of f is known, and s_n minimizes $\langle s, s\rangle$, subject to $s(x_i) = f(x_i)$, $i = 1, \dots, n$.*
Choose a target value $f_n^* \in [-\infty, \min_{y\in\mathcal{D}} s_n(y)]$. (The choice $f_n^* = \min s_n(y)$ is admissible only if none of the x_i is a global minimizer of s_n).
Calculate x_{n+1}, which is the value of y that minimizes the function

$$g_n(y) = (-1)^{m_0+1}\mu_n(y)\left[s_n(y) - f_n^*\right]^2, \quad y \in \mathcal{D} \setminus \{x_1, \dots, x_n\}. \tag{3.8}$$

Evaluate f at x_{n+1} and set $n := n + 1$.
Stop, if n is greater than a prescribed n_{max}.

The function g_n is infinitely differentiable on $\mathcal{D}\setminus\{x_1,\ldots,x_n\}$, but is not defined at the interpolation points. If $f_n^* = \min s_n(y)$, $y \in \mathcal{D}$, and if $s_n(x_i) > f_n^*$, $i = 1,\ldots,n$, then the global minimizers of g_n are the global minimizers of s_n. Thus one can minimize s_n, which is defined on the whole of $\mathcal{D}$, to obtain x_{n+1}. If $f_n^* < \min s_n(y)$, however, then $g_n(x)$ tends to infinity as x tends to x_i, $i = 1,\ldots,n$. Let $h_n : \mathcal{D} \to \mathbb{R}$ be defined as

$$h_n(y) := \begin{cases} \dfrac{1}{g_n(y)}, & y \notin \{x_1,\ldots,x_n\} \\ 0, & y = x_i,\ i = 1,\ldots,n \end{cases} . \tag{3.9}$$

The maximization of h_n on $\mathcal{D}$ is equivalent to the minimization of g_n. Further, h_n is infinitely differentiable on $\mathcal{D} \setminus \{x_1,\ldots,x_n\}$. It can also be shown, using the system (3.4), that it is in $C(\mathcal{D})$ in the linear case, in $C^1(\mathcal{D})$ in the thin plate spline case, in $C^2(\mathcal{D})$ in the cubic case, and in $C^\infty(\mathcal{D})$ in the multiquadric and Gaussian cases.

Under certain conditions on f and the values f_n^*, $n \to \infty$, it can be proved that a subsequence of the generated points $(x_n)_{n\in\mathbb{N}}$ converges to a global minimum. This is the subject of the following section.

4 Convergence of the method

Our aim is to prove convergence of the method for any continuous function f. A theorem by Törn and Zilinskas [19] tells us that the sequence that is generated by Algorithm 3 should be dense. Applied to our method, it states

Theorem 4 *The algorithm converges for every continuous function f if and only if it generates a sequence of points that is dense in $\mathcal{D}$.*

So our task is to establish the density of the sequence of generated points. Our main convergence result (Theorem 7), however, does not allow all choices of ϕ and m.

The first part of this section states Theorem 7 and derives a few corollaries. The second part gives the proof of the theorem.

4.1 Convergence results

Assume that f_n^* is set to $\min_{y\in\mathcal{D}} s_n(y)$ on each iteration and that this choice is admissible. Then, if there are large function values in a neighbourhood U of a global minimizer of f, the global minimizer of s_n might be outside U for every $n \geq n_0$, where n_0 is the number of initial points. In this case, U is not explored and the global optimum of f might be missed. Therefore, we have to assume that enough of the numbers $\min s_n(y) - f_n^*$ are sufficiently large. Specifically, let $\tau > 0$ and $\rho \geq 0$ be constants, where additionally $\rho < 1$ in the linear case and $\rho < 2$ in the thin plate spline and cubic cases, and define

$$\Delta_n := \min_{1 \leq i \leq n-1} \|x_n - x_i\|. \tag{4.1}$$

Then the condition

$$\min_{y \in \mathcal{D}} s_n(y) - f_n^* > \tau \Delta_n^{\rho/2} \|s_n\|_\infty, \tag{4.2}$$

for infinitely many $n \in I\!N$, where $\|.\|_\infty$ denotes the maximum norm of a function on $\mathcal{D}$, will lead to the required result. We note that the norms $\|s_n\|_\infty$ may diverge as $n \to \infty$.

Unfortunately, the choice of ϕ and m is restricted. In the proof of Theorem 7 we need the result that, for any $y \in \mathcal{D}$ and any neighbourhood U of y, $(-1)^{m_0+1}\mu_n(y)$ can be bounded above by a number that does not depend on n, if none of the points $x_1, \ldots, x_n$ is in U. This condition is achieved, if there is a function that takes the value 1 at y, that is identically zero outside U, and that is in the function space $\mathcal{N}_{\phi,m}(I\!R^d)$ as defined below.

Definition 5 *Let ϕ from (2.2) and $m \geq m_0$ be given. A continuous function $F : D \to I\!R$, $D \subset I\!R^d$, belongs to the function space $\mathcal{N}_{\phi,m}(D)$, if there exists a positive constant C such that, for any choice of interpolation points $x_1, \ldots, x_n \in D$ for which (2.11) holds, the interpolant $s_n \in \mathcal{A}_{\phi,m}$ to F at these points has the property*

$$\langle s_n, s_n \rangle \leq C.$$

The characterization of $\mathcal{N}_{\phi,m}(D)$ is rather abstract. In the linear, cubic and thin plate spline cases, the following proposition that is taken from Gutmann [3] provides a useful criterion to check whether it is satisfied. In the multiquadric and Gaussian cases, however, no such criterion is known.

Proposition 6 *Let $\phi(r) = r$, $\phi(r) = r^2 \log r$ or $\phi(r) = r^3$. Further, let $\kappa = 1$ in the linear case, $\kappa = 2$ in the thin plate spline case and $\kappa = 3$ in the cubic case, and choose the integer m such that $0 \leq m \leq d$ in the linear case, $1 \leq m \leq d+1$ in the thin plate spline case and $1 \leq m \leq d+2$ in the cubic case. Define $\nu := (d+\kappa)/2$ if $d+\kappa$ is even, and $\nu := (d+\kappa+1)/2$ otherwise. If $F \in C^\nu(I\!R^d)$ and has bounded support, then $F \in \mathcal{N}_{\phi,m}(I\!R^d)$.*

Global convergence will be established only for the cases covered by this proposition. It remains an open problem whether a similar property is achieved in other cases. Thus we have the following theorem.

Theorem 7 *Let $\phi(r) = r$, $\phi(r) = r^2 \log r$ or $\phi(r) = r^3$. Further, choose the integer m such that $0 \le m \le d$ in the linear case, $1 \le m \le d+1$ in the thin plate spline case and $1 \le m \le d+2$ in the cubic case. Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3, and s_n be the radial function that interpolates $(x_i, f(x_i)), i = 1, \ldots, n$. Assume that, for infinitely many $n \in \mathbb{N}$, the choice of f_n^* satisfies (4.2). Then the sequence (x_n) is dense in $\mathcal{D}$.*

A particular convergence result follows immediately from Theorems 4 and 7, because the right hand side of (4.2) is some real number.

Corollary 8 *Let the assumptions of Theorem 7 on ϕ and m hold. Further, let f be continuous, and, for infinitely many $n \in \mathbb{N}$, let $f_n^* = -\infty$. Then the method converges.*

An interesting question is to find conditions on f such that the maximum norm of an interpolant is uniformly bounded. If they hold, then the right-hand side of (4.2) can be replaced by $\tau\Delta_n^{\rho/2}$, so this constraint on f_n^* can be checked easily. We consider the special case of linear splines in one dimension, when $d = 1$, $\phi(r) = r$ and $m = 0$. For arbitrary points $x_1, \ldots, x_n$, the piecewise linear interpolant s_n attains its maximum and minimum values at interpolation points. Thus $\|s_n\|_\infty$ is bounded by $\|f\|_\infty$, a number that does not depend on the interpolation points. Therefore in this case the term $\|s_n\|_\infty$ may be dropped from (4.2).

For other radial basis functions and other dimensions this simplification may fail. It is shown in the next lemma, however, that the uniform boundedness of the semi-norm of an interpolant is sufficient for the uniform boundedness of the maximum norm. Thus, the second convergence result applies to functions f in $\mathcal{N}_{\phi,m}(\mathcal{D})$.

Lemma 9 *Let f be in $\mathcal{N}_{\phi,m}(\mathcal{D})$. Further, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{D}$ with pairwise different elements, such that (2.11) holds for $n = n_0$. For $n \ge n_0$, denote the radial basis function interpolant to f at $x_1, \ldots, x_n$ by s_n. Then $\|s_n\|_\infty$ is uniformly bounded by a number that only depends on $x_1, \ldots, x_{n_0}$.*

Proof:
We fix n, and we let y be any point of $\mathcal{D} \setminus \{x_1, \ldots, x_n\}$. Let $\tilde{s}_n$ be the radial function that interpolates $(y, f(y))$ and $(x_i, f(x_i))$, $i = 1, \ldots, n$. By analogy with equation (3.1), it can be written as

$$\tilde{s}_n(x) = s_n(x) + [f(y) - s_n(y)]\ell_n(y, x), \quad x \in \mathbb{R}^d,$$

where $\ell_n(y, .)$ is still the cardinal function that interpolates $(x_i, 0)$, $i = 1, \ldots, n$, and $(y, 1)$. Thus, as shown in Section 3,

$$\langle \tilde{s}_n, \tilde{s}_n \rangle = \langle s_n, s_n \rangle + [f(y) - s_n(y)]^2 (-1)^{m_0+1} \mu_n(y),$$

which gives the equation

$$[f(y) - s_n(y)]^2 = \frac{\langle \tilde{s}_n, \tilde{s}_n \rangle - \langle s_n, s_n \rangle}{(-1)^{m_0+1}\mu_n(y)}, \tag{4.3}$$

the value of $(-1)^{m_0+1}\mu_n(y)$ being strictly positive.

Next we show that $(-1)^{m_0+1}\mu_n(y)$ is bounded away from zero. Let $\ell_{n_0}(y,.)$ be the cardinal function that interpolates $(x_1, 0), \ldots, (x_{n_0}, 0)$ and $(y, 1)$. Then the semi-norm properties of $\langle .,. \rangle$ and Theorem 1 imply

$$\begin{aligned} 0 < (-1)^{m_0+1}\mu_{n_0}(y) &= \langle \ell_{n_0}(y,.), \ell_{n_0}(y,.) \rangle \\ &\le \langle \ell_n(y,.), \ell_n(y,.) \rangle = (-1)^{m_0+1}\mu_n(y). \end{aligned}$$

For $n = n_0$, let A and $A(y)$ be the matrices (2.10) and (3.5), respectively. By using Cramer's Rule to solve (3.4), we find

$$\mu_{n_0}(y) = \frac{\det A}{\det A(y)}.$$

Now $\det A$ is a nonzero constant, and $\det A(y)$ is bounded on $\mathcal{D}$. It follows that $(-1)^{m_0+1}\mu_{n_0}(y)$ is bounded away from zero. Therefore there exists a constant $\alpha > 0$ such that

$$(-1)^{m_0+1}\mu_n(y) \ge \alpha \quad \forall\, y \in \mathcal{D} \setminus \{x_1, \ldots, x_{n_0}\},\ n \ge n_0. \tag{4.4}$$

As $f \in \mathcal{N}_{\phi,m}(\mathcal{D})$, $\langle \tilde{s}_n, \tilde{s}_n \rangle$ is bounded above by a positive constant C. Further, $\langle s_n, s_n \rangle$ is nonnegative. It follows from (4.3) and (4.4) that

$$|f(y) - s_n(y)| \le \sqrt{\frac{C}{\alpha}}, \quad y \in \mathcal{D} \setminus \{x_1, \ldots, x_n\}.$$

Further, because f is bounded on $\mathcal{D}$, we obtain

$$|s_n(y)| \le \sqrt{\frac{C}{\alpha}} + \|f\|_\infty.$$

Note that the right-hand side is independent of n and y, as required. Alternatively, if $y \in \{x_1, \ldots, x_n\}$, we have

$$|s_n(y)| = |f(y)| \le \|f\|_\infty,$$

which completes the proof. □

Next, by applying Proposition 6, we obtain a criterion that ensures that f is in $\mathcal{N}_{\phi,m}(\mathcal{D})$.

Proposition 10 *Let ϕ, m and ν be defined as in Proposition 6, and let $f \in C^\nu(\mathcal{D})$, where $\mathcal{D} \subset I\!R^d$ is compact. Then $f \in \mathcal{N}_{\phi,m}(\mathcal{D})$.*

Proof:
By Whitney's theorem ([20]), f can be extended to a function $F \in C^\nu(I\!R^d)$ that is equal to f on $\mathcal{D}$. Now $\mathcal{D}$ is contained in a closed ball of radius δ, say, and there is an infinitely differentiable function g with $g(x) = 1$, $\|x\| \leq \delta$, and $g(x) = 0$, $\|x\| \geq 2\delta$. Thus $F \cdot g$ is in $C^\nu(I\!R^d)$, and by Proposition 6 it is in $\mathcal{N}_{\phi,m}(I\!R^d)$. Since $F \cdot g$ is equal to f on $\mathcal{D}$, it follows from the definition of the semi-norm that $f \in \mathcal{N}_{\phi,m}(\mathcal{D})$. $\square$

We complete this subsection by combining Theorem 7, Lemma 9 and Proposition 10.

Corollary 11 *Let the assumptions of Theorem 7 on ϕ and m hold. Let ν be as in Proposition 6, and let $f \in C^\nu(\mathcal{D})$. Further, assume that, for infinitely many $n \in I\!N$, f_n^* has the property*

$$\min_{y \in \mathcal{D}} s_n(y) - f_n^* \geq \tau \Delta_n^{\rho/2},$$

where $\tau > 0$ is a constant, and where Δ_n and ρ are as in the first paragraph of this subsection. Then the method converges.

4.2 Proof of Theorem 7

In order to establish Theorem 7, some lemmas are needed on the behaviour of the coefficients μ_n.

Lemma 12 *Let ϕ be any of the radial basis functions in (2.2), and let m_0 and $m \geq m_0$ be chosen as in Section 2. Let $\mathcal{D} \subset I\!R^d$ be compact, and let $(x_n)_{n \in I\!N}$ be a convergent sequence in $\mathcal{D}$ with pairwise different elements. Further, let $(y_n)_{n \in I\!N}$ be a sequence in $\mathcal{D}$ such that $y_n \neq x_n$, $n \in I\!N$, and $\lim_{n\to\infty} \|x_n - y_n\| = 0$. Choose k points $z_1, \ldots, z_k \in \mathcal{D}$ that satisfy condition (2.11). Assume (x_n) converges to $x^* \in \mathcal{D}\backslash\{z_1, \ldots, z_k\}$. For any $y \in \mathcal{D}\backslash\{z_1, \ldots, z_k, y_{n+1}\}$, let $\tilde{\ell}_y$ be the cardinal spline that interpolates the data $(z_1, 0), \ldots, (z_k, 0), (y_{n+1}, 0)$ and $(y, 1)$, and let $\tilde{\mu}_n(y)$ be the coefficient of $\tilde{\ell}_y$ that satisfies $(-1)^{m_0+1}\tilde{\mu}_n(y) = \langle \tilde{\ell}_y, \tilde{\ell}_y \rangle$. Then, for $0 \leq \rho < 1$ in the linear case and $0 \leq \rho < 2$ in the other cases,*

$$\lim_{n\to\infty} (-1)^{m_0+1} \|y_{n+1} - x_{n+1}\|^\rho \, \tilde{\mu}_n(x_{n+1}) = \infty. \tag{4.5}$$

Proof:
Let A_n and $A_n(x_{n+1})$ be the matrices of the form (2.10) for the interpolation points $z_1, \ldots, z_k, y_{n+1}$ and $z_1, \ldots, z_k, y_{n+1}, x_{n+1}$, respectively. For sufficiently large n, neither x_{n+1} nor y_{n+1} is in the set $\{z_1, \ldots, z_k\}$. Thus, A_n and $A_n(x_{n+1})$ are nonsingular. Cramer's Rule implies

$$\tilde{\mu}_n(x_{n+1}) = \frac{\det A_n}{\det A_n(x_{n+1})}. \tag{4.6}$$

Also, let A^* be the matrix of the form (2.10) for the interpolation points $z_1, \ldots, z_k, x^*$. By the continuity of the determinant,

$$\lim_{n\to\infty} \det A_n = \det A^* \neq 0. \tag{4.7}$$

In order to investigate the behaviour of $\|y_{n+1} - x_{n+1}\|^{-\rho} \det A_n(x_{n+1})$, we let

$$v(y) := \Big(\phi(\|y - z_1\|), \ldots, \phi(\|y - z_k\|)\Big)^T, \quad y \in \mathcal{D},$$

and

$$p(y) := \Big(p_1(y), \ldots, p_{\hat{m}}(y)\Big)^T, \quad y \in \mathcal{D},$$

where $\hat{m} = \dim \Pi_m$ and $p_1, \ldots, p_{\hat{m}}$ are as in Section 2. Thus $A_n(x_{n+1})$ is the matrix

$$\begin{pmatrix} \Phi & v(y_{n+1}) & v(x_{n+1}) & P \\ v(y_{n+1})^T & \phi(0) & \phi(\|y_{n+1} - x_{n+1}\|) & p(y_{n+1})^T \\ v(x_{n+1})^T & \phi(\|y_{n+1} - x_{n+1}\|) & \phi(0) & p(x_{n+1})^T \\ P^T & p(y_{n+1}) & p(x_{n+1}) & 0 \end{pmatrix}, \tag{4.8}$$

where Φ and P correspond to (2.3) and (2.9), respectively, if we set $n = k$ and $\{x_1, \ldots, x_n\} = \{z_1, \ldots, z_k\}$. Now the rows

$$\begin{array}{cccccc} & (\; v(y_{n+1})^T & \phi(0) & \phi(\|y_{n+1} - x_{n+1}\|) & p(y_{n+1})^T \;) \\ \text{and} & (\; v(x_{n+1})^T & \phi(\|y_{n+1} - x_{n+1}\|) & \phi(0) & p(x_{n+1})^T \;) \end{array}$$

have the same limit as $n \to \infty$, so $\det A_n(x_{n+1})$ tends to zero. Hence the properties (4.6) and (4.7) prove the assertion (4.5) for $\rho = 0$.

For $\rho > 0$, note that the value of the determinant of the matrix (4.8) does not change if we replace the second row by the difference between the second and third rows, and subsequently replace the second column by the difference between the second and third columns. Then $\det A_n(x_{n+1})$ becomes

$$\left|\begin{array}{cccc} \Phi & v(y_{n+1})-v(x_{n+1}) & v(x_{n+1}) & P \\ \begin{array}{c} v(y_{n+1})^T \\ -v(x_{n+1})^T \end{array} & 2[\phi(0)-\phi(\|y_{n+1}-x_{n+1}\|)] & \begin{array}{c} \phi(\|y_{n+1}-x_{n+1}\|) \\ -\phi(0) \end{array} & \begin{array}{c} p(y_{n+1})^T \\ -p(x_{n+1})^T \end{array} \\ v(x_{n+1})^T & \phi(\|y_{n+1}-x_{n+1}\|)-\phi(0) & \phi(0) & p(x_{n+1})^T \\ P^T & p(y_{n+1})-p(x_{n+1}) & p(x_{n+1}) & 0 \end{array}\right|. \tag{4.9}$$

We have to divide the determinant by $\|y_{n+1}-x_{n+1}\|^\rho$, so we divide the second row and then the second column of (4.9) by $\|y_{n+1}-x_{n+1}\|^{\rho/2}$. Then the following remarks are helpful.

For each choice of ϕ and each $j=1,\ldots,k$, the function $\phi(\|z_j-x\|)$, $x \in \mathcal{D}$, is Lipschitz continuous, so the components of $v(y_{n+1})-v(x_{n+1})$ satisfy

$$|\phi(\|z_j-y_{n+1}\|)-\phi(\|z_j-x_{n+1}\|)| \leq \text{const}\,\|x_{n+1}-y_{n+1}\|, \quad j=1,\ldots,k.$$

Thus for $\rho<2$ we have

$$\lim_{n\to\infty} \frac{1}{\|y_{n+1}-x_{n+1}\|^{\rho/2}}\Big[\phi(\|z_j-y_{n+1}\|)-\phi(\|z_j-x_{n+1}\|)\Big]=0, \tag{4.10}$$

Similarly, for $\rho<2$, the components of $p(y_{n+1})-p(x_{n+1})$ have the property

$$\lim_{n\to\infty} \frac{1}{\|y_{n+1}-x_{n+1}\|^{\rho/2}}\Big[p_i(y_{n+1})-p_i(x_{n+1})\Big]=0 \tag{4.11}$$

Finally, we deduce

$$\lim_{n\to\infty} \frac{1}{\|y_{n+1}-x_{n+1}\|^{\rho}}\Big[\phi(\|y_{n+1}-x_{n+1}\|)-\phi(0)\Big]=0, \tag{4.12}$$

for $\rho<1$ in the linear case, for $\rho<2$ in the thin plate spline, multiquadric and Gaussian cases, and for $\rho<3$ in the cubic case. This is clear in the linear, thin plate spline and cubic cases. In the other two cases it follows from second order Taylor expansion of ϕ, because $\phi'(0)=0$ and ϕ'' is bounded on $\mathbb{R}_+$. Thus (4.9) – (4.12) provide

$$\lim_{n\to\infty} \|y_{n+1}-x_{n+1}\|^{-\rho} \det A_n(x_{n+1})=0,$$

for the given values of ρ. Hence (4.6) and (4.7) imply (4.5). □

Now we obtain

Lemma 13 *Let ϕ, m_0 and m be chosen as in Lemma 12, and let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3. Further, let $0\leq\rho<1$ in the linear case and*

$0 \le \rho < 2$ in the other cases. Then, for every convergent subsequence $(x_{n_k})_{k\in I\!N}$ of (x_n),

$$\lim_{k\to\infty} (-1)^{m_0+1} \Delta_{n_k}^{\rho} \mu_{n_k-1}(x_{n_k}) = \infty,$$

where $\mu_n(.)$ is defined in Section 3 and Δ_{n_k} is expression (4.1) for $n = n_k$.

Proof:
For $n \ge 2$, define j_n to be the natural number j that minimizes $\|x_n - x_j\|$, $j < n$, so $\Delta_n = \|x_n - x_{j_n}\|$. Further, let $(y_n)_{n\in I\!N}$ be the sequence

$$y_n := \begin{cases} x_2, & n = 1, \\ x_{j_n}, & n \ge 2. \end{cases}$$

Let $(x_{n_k})_{k\in I\!N}$ be a subsequence of $(x_n)_{n\in I\!N}$, that converges to x^*, say. Convergence and the choice of $(y_n)_{n\in I\!N}$ imply $\lim_{k\to\infty} \|x_{n_k} - y_{n_k}\| = 0$.

The initial step of Algorithm 3 provides a finite number of points that satisfy (2.11), so the initial interpolation matrix (2.10) is nonsingular. If one of these points is x^*, we pick $x_{n_{k_0}}$ in a neighbourhood of x^* so that the interpolation matrix to $x_{n_{k_0}}$ and the other initial points is also nonsingular. Therefore there exist points $\hat{x}_1, \ldots, \hat{x}_l$ in $(x_n)_{n\in I\!N}$ such that their interpolation matrix is nonsingular, and $x^* \notin \{\hat{x}_1, \ldots, \hat{x}_l\}$.

For sufficiently large $k \in I\!N$, such that $y_{n_k} \notin \{\hat{x}_1, \ldots, \hat{x}_l\}$, and for any $y \in \mathcal{D} \setminus \{x_1, \ldots, x_{n_k-1}\}$, we let $\hat{\ell}_k(y, .)$ be the radial function that interpolates $(\hat{x}_1, 0), \ldots, (\hat{x}_l, 0), (y_{n_k}, 0)$ and $(y, 1)$, and we let $\ell_{n_k-1}(y, .)$ be the interpolant to $(x_1, 0), \ldots, (x_{n_k-1}, 0)$ and $(y, 1)$. Because $\ell_{n_k-1}(y, .)$ interpolates $(y_{n_k}, 0)$ and $(\hat{x}_i, 0)$, $i = 1, \ldots, l$, for sufficiently large k, (3.7) and Theorem 1 imply the inequality

$$\begin{aligned} (-1)^{m_0+1} \hat{\mu}_k(y) &= \langle \hat{\ell}_k(y, .), \hat{\ell}_k(y, .) \rangle \\ &\le \langle \ell_{n_k-1}(y, .), \ell_{n_k-1}(y, .) \rangle = (-1)^{m_0+1} \mu_{n_k-1}(y) \end{aligned} \tag{4.13}$$

for the coefficients $\hat{\mu}_k$ and μ_{n_k-1}.

We apply Lemma 12 in the case when $\{z_1, \ldots, z_k\}$ is the set $\{\hat{x}_1, \ldots, \hat{x}_l\}$ and $n = n_k - 1$. It follows that

$$\lim_{k\to\infty} (-1)^{m_0+1} \Delta_{n_k}^{\rho} \hat{\mu}_k(x_{n_k}) = \lim_{k\to\infty} (-1)^{m_0+1} \|x_{n_k} - y_{n_k}\|^{\rho} \hat{\mu}_k(x_{n_k}) = \infty,$$

with the choice of ρ stated in Lemma 12. Thus, setting $y = x_{n_k}$ in (4.13), we obtain that $(-1)^{m_0+1} \Delta_{n_k}^{\rho} \mu_{n_k-1}(x_{n_k})$ tends to infinity as $k \to \infty$. □

Finally we show, using Proposition 6, that the coefficients $\mu_n(y)$ are uniformly bounded, if y is bounded away from the points that are generated by the algorithm.

Lemma 14 *Let $\phi(r) = r$, $\phi(r) = r^2 \log r$ or $\phi(r) = r^3$. Further, choose the integer m such that $0 \le m \le d$ in the linear case, $1 \le m \le d+1$ in the thin plate spline case and $1 \le m \le d+2$ in the cubic case. Let $(x_n)_{n \in I\!N}$ be the sequence generated by Algorithm 3, and let n_0 be the number of points chosen in the initial step. Assume that there exist $y_0 \in \mathcal{D}$ and a neighbourhood $N_\delta := \{x \in I\!R^d : \|x - y_0\| < \delta\}$, $\delta > 0$, that does not contain any point of the sequence. Then there exists $K > 0$, that depends only on y_0 and δ, such that*

$$(-1)^{m_0+1} \mu_n(y_0) \le K \quad \forall\, n \ge n_0.$$

Proof:
For any $n \ge n_0$, let ℓ_n be the radial function that is defined by $\ell_n(x_i) = 0$, $i = 1, \ldots, n$, and $\ell_n(y_0) = 1$. There exists a compactly supported infinitely differentiable function F that takes the value 1 at y_0 and 0 on $I\!R^d \setminus N_\delta$. It follows from Proposition 6 that $F \in \mathcal{N}_{\phi,m}$. ℓ_n interpolates F at $x_1, \ldots, x_n$ and y_0. Therefore, there is a positive number K, depending on y_0 and δ, such that

$$(-1)^{m_0+1} \mu_n(y_0) = \langle \ell_n, \ell_n \rangle \le K, \quad n \ge n_0. \qquad \square$$

Now we are ready to prove Theorem 7.

Proof of Theorem 7:
Assume there is $y_0 \in \mathcal{D}$ and an open neighbourhood $U = \{x \in I\!R^d : \|x - y_0\| < \delta\}$, $\delta > 0$, that does not contain an interpolation point. The iteration step of Algorithm 3 gives

$$g_n(x_{n+1}) \le g_n(y_0), \quad n \ge n_0,$$

where n_0 is the number of points chosen in the initial step of the algorithm.
By assumption (4.2), there is a subsequence $(n_k)_{k \in I\!N}$ of the natural numbers such that

$$\min_{y \in \mathcal{D}} s_{n_k-1}(y) - f^*_{n_k-1} > \tau \Delta^{\rho/2}_{n_k-1} \|s_{n_k-1}\|_\infty \ge 0, \quad k \in I\!N, \tag{4.14}$$

with $\tau > 0$, Δ_{n_k-1} being the expression (4.1) for $n = n_k - 1$, $0 \le \rho < 1$ in the linear and $0 \le \rho < 2$ in the thin plate spline and cubic cases. The sequence $(x_{n_k})_{k \in I\!N}$ is a sequence in a compact set, thus it contains a convergent subsequence. Therefore, without loss of generality, we assume that $(x_{n_k})_{k \in I\!N}$ itself converges.

For all $k \in I\!N$, x_{n_k} is the minimizer of $g_{n_k-1}(x)$. Thus, if $f^*_{n_k-1} > -\infty$,

$$\begin{aligned} &(-1)^{m_0+1} \mu_{n_k-1}(x_{n_k}) [s_{n_k-1}(x_{n_k}) - f^*_{n_k-1}]^2 \\ &\qquad \le (-1)^{m_0+1} \mu_{n_k-1}(y_0) [s_{n_k-1}(y_0) - f^*_{n_k-1}]^2. \end{aligned} \tag{4.15}$$

If $\|s_{n_k-1}\|_\infty > 0$, this inequality, condition (4.14) and the definition of $\|.\|_\infty$ provide

$$\begin{aligned}
(-1)^{m_0+1}\mu_{n_k-1}(x_{n_k}) &\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[\frac{s_{n_k-1}(y_0)-f^*_{n_k-1}}{s_{n_k-1}(x_{n_k})-f^*_{n_k-1}}\right]^2 \\
&\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[1+\frac{|s_{n_k-1}(y_0)-s_{n_k-1}(x_{n_k})|}{s_{n_k-1}(x_{n_k})-f^*_{n_k-1}}\right]^2 \\
&\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[1+\frac{1}{\tau\Delta_{n_k}^{\rho/2}}\frac{|s_{n_k-1}(y_0)-s_{n_k-1}(x_{n_k})|}{\|s_{n_k-1}\|_\infty}\right]^2 \\
&\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[1+\frac{2}{\tau\Delta_{n_k}^{\rho/2}}\right]^2 .
\end{aligned}$$

If $\|s_{n_k-1}\|_\infty = 0$, $s_{n_k-1}(y) - f^*_{n_k-1}$ is a positive number independent of y, thus (4.15) gives

$$\begin{aligned}
(-1)^{m_0+1}\mu_{n_k-1}(x_{n_k}) &\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0) \\
&\leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[1+\frac{2}{\tau\Delta_{n_k}^{\rho/2}}\right]^2 ,
\end{aligned}$$

for any positive τ, as before. Remark 2 shows that this inequality holds also in the case $f^*_{n_k-1} = -\infty$. Multiplying both sides by $\Delta_{n_k}^{\rho}$ yields

$$\Delta_{n_k}^{\rho}(-1)^{m_0+1}\mu_{n_k-1}(x_{n_k}) \leq (-1)^{m_0+1}\mu_{n_k-1}(y_0)\left[\Delta_{n_k}^{\rho/2}+\frac{2}{\tau}\right]^2 . \tag{4.16}$$

By Lemma 13, the left-hand side of (4.16) tends to infinity as k tends to infinity. However, Lemma 14 states that $(-1)^{m_0+1}\mu_n(y_0)$ is bounded above by a constant that does not depend on n. Thus the right-hand side of (4.16) is bounded by a constant that is independent of k which contradicts (4.16). Therefore there is a point in the sequence that is an element of U. This implies that in each neighbourhood of an arbitrary $y_0 \in \mathcal{D}$ there are infinitely many elements of $(x_n)_{n\in\mathbb{N}}$, so the sequence is dense in $\mathcal{D}$. □

5 Relations to statistical global optimization

In this section we consider the similarities between the given radial basis function method and the P-algorithm. The idea of that method is proposed by Kushner [11], [12] for one-dimensional problems. Here the objective function is regarded as a realization of a Brownian motion stochastic process. If real numbers $x_1 < \ldots < x_n$

are given, and their function values $f(x_1), \dots, f(x_n)$ have been calculated, the model yields, for each x in the feasible set, a mean value $\mathrm{Mean}(x)$ and a variance $\mathrm{Var}(x)$. $\mathrm{Mean}(x)$ serves as a prediction of the true function value at x, while $\mathrm{Var}(x)$ is a measure of uncertainty. It turns out that Mean is the piecewise linear interpolant of the given data. The variance is piecewise quadratic on $[x_1, x_n]$, nonnegative and takes the value zero at $x_1, \dots, x_n$. For a real number x, let F_x be the normally distributed random variable F_x with mean $\mathrm{Mean}(x)$ and variance $\mathrm{Var}(x)$. Now a nonnegative ϵ_n is chosen, and the next point x_{n+1} will be the one that maximizes the utility function

$$P\left(F_x \le \min_{i=1,\dots,n} f(x_i) - \epsilon_n\right), \quad x \in \mathcal{D}, \tag{5.1}$$

where P denotes probability. One can show that maximizing (5.1) is equivalent to maximizing

$$\frac{\mathrm{Var}(x)}{[\mathrm{Mean}(x) - \min\{f(x_1), \dots, f(x_n)\} + \epsilon_n]^2}, \quad x \in \mathcal{D}. \tag{5.2}$$

Compare our method in one dimension with the choice of linear splines, i.e. $\phi(r) = r$ and $m = 0$, and with the target values $f_n^* = \min\{f(x_1), \dots, f(x_n)\} - \epsilon_n$. In this case, the interpolant s_n is identical to Mean. Further, except for a constant factor,

$$\mathrm{Var}(x) = -\frac{1}{\mu_n(x)}.$$

Therefore, Kushner's method and our method using linear splines are equivalent.

Žilinskas [21], [22] extends this approach to Gaussian random processes in several dimensions. He uses the selection rule (5.1) and introduces the name "P-algorithm". In addition, he gives an axiomatic description of the terms involved in (5.1), namely the mean value function, the variance function and the utility function. We relate these results to our method.

Consider a symmetric function $\sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, (x, z) \mapsto \sigma(x, z)$, and assume that σ is conditionally positive or negative definite of order m. This means, there exists $\alpha \in \{0, 1\}$ such that, given n different points $x_1, \dots, x_n \in \mathbb{R}^d$ and multipliers $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, we have

$$(-1)^\alpha \sum_{i,j=1}^{n} \lambda_i \lambda_j \sigma(x_i, x_j) > 0,$$

if the λ_i, $i = 1, \dots, n$, are not all zero and satisfy

$$\sum_{i=1}^{n} \lambda_i p(x_i) = 0, \quad p \in \Pi_m.$$

Denote the matrix with the elements $\sigma(x_i, x_j)$, $i, j = 1, \dots, n$, by Σ, and the matrix with the elements $p_j(x_i)$, $i = 1, \dots, n$, $j = 1, \dots, \hat{m}$ by P, where $\{p_j : j = 1, \dots, \hat{m}\}$ is a basis of Π_m and $\hat{m}$ its dimension. The analogue of expression (2.10) is the matrix

$$A = \begin{pmatrix} \Sigma & P \\ P^T & 0 \end{pmatrix}. \tag{5.3}$$

Further, we now let the interpolant s to the components of $F = (f(x_1), \dots, f(x_n))^T$ be the function

$$s(y) = \sum_{i=1}^{n} \lambda_i \sigma(y, x_i) + \sum_{j=1}^{\hat{m}} c_j p_j(y),$$

whose coefficients solve the system

$$A \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} F \\ 0_{\hat{m}} \end{pmatrix},$$

It can be written as

$$s(y) = v_m(y)^T A^{-1} \begin{pmatrix} F \\ 0_{\hat{m}} \end{pmatrix}, \quad y \in \mathcal{D}, \tag{5.4}$$

where $v_m(y)$ is the vector

$$v_m(y) := (\sigma(y, x_1), \dots, \sigma(y, x_n), p_1(y), \dots, p_{\hat{m}}(y))^T. \tag{5.5}$$

The nonnegative function

$$\mathrm{Var}(y) = |\sigma(y, y) - v_m(y)^T A^{-1} v_m(y)|, \quad y \in \mathcal{D}, \tag{5.6}$$

is assumed to be a measure of uncertainty. Note that $v_m(x_i)$ is the i-th column of A, $i = 1, \dots, n$, so we obtain

$$\begin{aligned} \mathrm{Var}(x_i) &= |\sigma(x_i, x_i) - v_m(x_i)^T A^{-1} v_m(x_i)| \\ &= |\sigma(x_i, x_i) - v_m(x_i)^T e_i| \\ &= 0. \end{aligned}$$

Thus there is no uncertainty at the interpolation points, which is meaningful because we know the true function values there.

For the P-algorithm, σ is interpreted as the correlation function of a Gaussian stochastic process. The use of a normal distribution, for example, gives $\sigma(x, y) := e^{-\|x-y\|^2/2}$, but other choices of σ are also considered in the literature. All of them are positive definite, so we set $m = -1$. The conditional mean and the conditional variance can be expressed as ([21])

$$\begin{aligned}
\mathrm{Mean}(y) &= \Big(f(x_1), \ldots, f(x_n)\Big)\Sigma^{-1}\begin{pmatrix}\sigma(y,x_1)\\ \vdots \\ \sigma(y,x_n)\end{pmatrix}, &(5.7)\\
\mathrm{Var}(y) &= \sigma(y,y) - \Big(\sigma(y,x_1), \ldots, \sigma(y,x_n)\Big)\Sigma^{-1}\begin{pmatrix}\sigma(y,x_1)\\ \vdots \\ \sigma(y,x_n)\end{pmatrix}, &(5.8)
\end{aligned}$$

which agree with (5.4) and (5.6).

For our method, given ϕ and m, it is suitable to define $\sigma(x,y) := \phi(\|x-y\|)$. Thus $\Sigma = \Phi$, and the coefficients of the interpolant s solve (2.13). This gives the form

$$s(y) = v_m(y)^T A^{-1}\begin{pmatrix}F\\ 0\end{pmatrix}, \tag{5.9}$$

which is the same as the function (5.4).

We have seen already that $|1/\mu_n|$ can be regarded as a variance in the case of linear splines in one dimension. An expression for it containing the matrix A and the vector $v_m(x)$ can be derived in the following way. For any $y \in \mathcal{D}\backslash\{x_1,\ldots,x_n\}$, consider the cardinal function (3.3). The second cardinality condition from (3.2) implies

$$\frac{1}{\mu_n(y)} = \phi(0) + \sum_{i=1}^{n}\frac{\alpha_i(y)}{\mu_n(y)}\phi(\|y-x_i\|) + \sum_{j=1}^{\hat{m}}\frac{b_j(y)}{\mu_n(y)}p_j(y). \tag{5.10}$$

The coefficients $\alpha(y)$, $\mu_n(y)$ and $b(y)$ solve the system (3.4). Moreover, the vector (5.5) contains the first n and the last $\hat{m}$ elements of the $(n+1)$-th column of $A(y)$. Therefore $\alpha(y)$ and $b(y)$ also solve

$$A\begin{pmatrix}\alpha(y)\\ b(y)\end{pmatrix} = -\mu_n(y)v_m(y),$$

which implies

$$\frac{1}{\mu_n(y)}\begin{pmatrix}\alpha(y)\\ b(y)\end{pmatrix} = -A^{-1}v_m(y).$$

Thus, replacing the terms $\alpha_i(y)/\mu_n(y)$ and $b_j(y)/\mu_n(y)$ in (5.10), we find

$$\frac{1}{\mu_n(y)} = \phi(0) - v_m(y)^T A^{-1} v_m(y). \tag{5.11}$$

It follows from $\sigma(x,x) = \phi(0)$, $x \in \mathcal{D}$, that expression (5.6) is equivalent to $|1/\mu_n(y)|$.

Finally, we consider the selection rule for the next point. For the P-algorithm, it has already been noted that the maximization of (5.1) is equivalent to the maximization of (5.2). For a given target value f^* define the function $U : I\!R^2 \to I\!R$ as

$$U(M,V) := \frac{V^2}{(M-f^*)^2}. \tag{5.12}$$

It is increasing in V and decreasing in M, if $f^* \leq M$. Also, it satisfies the axioms of rational search stated in Žilinskas [22]. Then, employing (5.4) and (5.6), both methods choose the point that maximizes

$$U\Big(s(y), \sqrt{\mathrm{Var}(y)}\Big), \quad y \in \mathcal{D}.$$

6 Search strategies and practical questions

Practical features of our method have received little attention in this paper. Several questions arise concerning the choice of parameters in Algorithm 3.

1. What radial basis function ϕ should be chosen, and what polynomial degree m?
2. What is a good strategy for the choice of the target values f_n^*?
3. Given a target value, how should the minimization of g_n in (3.8) (or the maximization of h_n in (3.9)) be carried out? Should we approximate the global optimum of g_n (or h_n) or compute a (possibly non-global) local minimum?

The first problem has not been investigated thoroughly. Experiments using cubic and thin plate splines on a few test functions suggest that one cannot say in general that one of them is better than the other. Experiments have not been tried yet for the other types.

The choice of target values is crucial for the performance of the method. The interpretation of the two extremal cases has been noted in Section 3. Specifically, the choice $f_n^* = \min_{y\in\mathcal{D}} s_n(y)$ means that we trust our model and assume that the minimizer of s_n is close to the global minimizer of f. In the other case, namely $f_n^* = -\infty$, we try to find a point in a region that has not been explored at all. It may be best to employ a mix between values of f_n^* that are suitable for convergence to a local minimizer and values that provide points in previously unexplored regions of the domain.

Research is going on for the third question. We prefer to maximize h_n, as this function is defined everywhere on $\mathcal{D}$. It might seem strange that we consider computing the global optimum of h_n, i.e. that a global optimization problem is replaced by another one. However, unlike f, h_n can be evaluated quickly, and derivatives are available as well. Also we know roughly where the local maxima

function	dimension	no. of local minima	no. of global minima	domain
Branin	2	3	3	$[-5,10]\times[0,15]$
Goldstein-Price	2	4	1	$[-2,2]^2$
Hartman 3	3	4	1	$[0,1]^3$
Shekel 5	4	5	1	$[0,10]^4$
Shekel 7	4	7	1	$[0,10]^4$
Shekel 10	4	10	1	$[0,10]^4$
Hartman 6	6	4	1	$[0,1]^6$

Table 1: *Dixon-Szegö test functions and their dimension, the domain and the number of local and global minima.*

of h_n lie. Thus the maximization of h_n is much easier than the minimization of f. In addition, as the problem (GOP) is very difficult under our assumptions, it would take too long to compute a global minimizer accurately. Therefore, from a practical point of view, we are interested in an approximate solution of (GOP). So it should suffice to determine an approximation to the maximizer of (3.9). As far as the second option in question 3 is concerned, we have to find a way to choose starting points or search regions in order to ensure fast convergence, which is still an open problem.

Experiments show that large differences between function values can cause the interpolant to oscillate very strongly. Thus its minimal value can be much below the least calculated function value. We have found in numerical computations that these inefficiencies are reduced if large function values are replaced by the median of all available function values.

Some experiments were performed using the test functions proposed by Dixon and Szegö [2]. Table 1 gives the name of each function, the dimension, the domain and the number of local and global minima in that domain. The maximization of (3.9) was carried out using a version of the tunneling method (Levy and Montalvo [13]).

The target values f_n^* are determined as follows. The idea is to perform cycles of $N+1$ iterations for some $N \in \mathbb{N}$, where each cycle employs a range of target values, starting with a low one (global search), and ending with a value of f_n^* that is close to $\min s_n(y)$ (local search). Then we go back to a global search, starting the cycle again. The results that we report have been obtained using the following strategy. We choose the cycle length $N = 5$. Let the number of initial points be n_0, let the cycle start at $n = \tilde{n}$, and let the function values be ordered, i.e. $f(x_1) \leq \ldots \leq f(x_n)$. If $f(x_1) = f(x_n)$, the interpolant is a constant function, because we pick $m \geq 0$, so the maximization of (3.9) is equivalent to the maximization of $|1/\mu_n|$, if $f_n^* < f(x_1)$. In this case, we choose $f_n^* = -\infty$. Otherwise, for $\tilde{n} \leq n \leq \tilde{n} + N - 1$, we set

$$f_n^* = \min_{y \in \mathcal{D}} s_n(y) - \left(\frac{N - n + \tilde{n}}{N}\right)^2 \left(f(x_{\sigma(n)}) - \min_{y \in \mathcal{D}} s_n(y)\right),$$

where $\sigma(\tilde{n}) = \tilde{n}$ and $\sigma(n) = \sigma(n-1) - \left\lfloor \dfrac{n - n_0}{N} \right\rfloor$, $\tilde{n} + 1 \leq n \leq \tilde{n} + N - 1$. When $n = \tilde{n} + N$ we set $f_n^* = \min s_n(y)$. However, we have to be careful here since this choice is only admissible if x_1 is not one of the global minimizers of s_n. Thus, we do not accept this choice, if $(f(x_1) - \min s_n(y)) < 10^{-4}|f(x_1)|$, provided $f(x_1)$ is nonzero, or if $f(x_1) - \min_{y \in \mathcal{D}} s_n(y) < 10^{-4}$, if $f(x_1) = 0$. In these cases, we set $f_n^* = \min s_n(y) - 10^{-2}|f(x_1)|$ and $f_n^* = \min s_n(y) - 10^{-2}$, respectively, to try to obtain a yet lower function value.

We use thin plate splines as radial basis function, except for the Hartman 3 function where we take cubic splines. The algorithm is stopped when the relative error $|f_{\text{best}} - f^*|/|f^*|$ becomes smaller than a fixed ϵ, where f^* is the global optimum and f_{best} the current best function value. The optimal values of all test functions in Table 1 are nonzero, so this stopping criterion is valid.

Table 2 reports the number of function evaluations needed to achieve a relative error less than 1% and 0.01%. RBF denotes our method using the target value strategy described above. DIRECT (Jones, Perttunen and Stuckman [9]) and MCS (Multilevel Coordinate Search, Huyer and Neumaier [5]) are recent methods that, according to the results presented in those papers, are more efficient than most of their competitors on the Dixon-Szegö testbed. DE (Differential Evolution, Storn and Price [18]) is an evolutionary method that operates only at the global level, which explains the large number of function evaluations. EGO (Efficient Global Optimization, Jones, Schonlau and Welch [10]) is the latest method known to us. Unfortunately, no tests are reported on the Shekel test functions. All the results from these methods are quoted from the papers mentioned above. For the DIRECT method, numbers of evaluations for both the 1% and the 0.01% stopping criterion are reported. For DE and EGO only results for the 1% criterion are available, whereas MCS only uses the 0.01% criterion. It should be noted that MCS uses a local search method at some stages of the algorithm, and in all the cases of Table 2 the first local minimum found is the global one.

7 Conclusions

Our global optimization method converges to the global minimizer of an arbitrary continuous function f, if we choose the sequence of target values carefully. If f is

	error < 1%				error< 0.01%		
Test function	RBF	DIRECT	DE	EGO	RBF	DIRECT	MCS
Branin	44	63	1190	28	64	195	41
Goldstein-Price	63	101	1018	32	76	191	81
Hartman 3	25	83	476	35	79	199	79
Shekel 5	76	103	6400	–	100	155	83
Shekel 7	76	97	6194	–	125	145	129
Shekel 10	51	97	6251	–	112	145	103
Hartman 6	112	213	7220	121	160	571	111

Table 2: *Number of function evaluations for our method in comparison to DIRECT, DE, EGO and MCS with two different stopping criteria.*

sufficiently smooth, there is even a suitable condition on this sequence that can be checked by the algorithm. However, it is unsatisfactory that the multiquadric and Gaussian cases are excluded from the statement of Theorem 7. It is believed that the convergence result is true also in these cases, although they are not covered by the analysis in [3].

Table 2 shows that the method is able to compete with other global optimization methods on the set of the Dixon-Szegö test functions. The test functions in this testbed, however, are of relatively low dimension, and the number of local and global minima is very small. Therefore, it is necessary to test the method on other sets of test functions and of course on real-world applications.

The relation to the P-algorithm is very interesting. It is hoped that the connections can be exploited further. In particular, the choice of the target values and the determination of the point of highest utility are common problems. Solutions to these problems may be developed that are useful for both methods.

Acknowledgement I am very grateful to my supervisor, Prof. M.J.D. Powell, for his constant help and his guidance. Also, I would like to thank the German Academic Exchange Service for supporting this research with a doctoral scholarship (HSP III) and the Engineering and Physical Sciences Research Council for further support with a research grant.

References

[1] P. Alotto, A. Caiti, G. Molinari, and M.Repetto. A Multiquadrics-based Algorithm for the Acceleration of Simulated Annealing Optimization Procedures. *IEEE Transactions on Magnetics*, 32(3):1198–1201, 1996.

[2] L.C.W. Dixon and G.P. Szegö. The Global Optimization Problem: An Introduction. In L.C.W. Dixon and G.P. Szegö, editors, *Towards Global Optimization 2*, pages 1–15. North-Holland, Amsterdam, 1978.

[3] H.-M. Gutmann. On the semi-norm of radial basis function interpolants. In preparation.

[4] R. Horst and P.M. Pardalos. *Handbook of Global Optimization*. Kluwer, Dordrecht, 1994.

[5] W. Huyer and A. Neumaier. Global optimization by multilevel coordinate search. *Journal of Global Optimization*, 14(4):331–355, 1999.

[6] T. Ishikawa and M. Matsunami. An Optimization Method Based on Radial Basis Functions. *IEEE Transactions on Magnetics*, 33(2):1868–1871, 1997.

[7] T. Ishikawa, Y. Tsukui, and M. Matsunami. A Combines Method for the Global Optimization Using Radial Basis Function and Deterministic Approach. *IEEE Transactions on Magnetics*, 35(3):1730–1733, 1999.

[8] D.R. Jones. Global optimization with response surfaces. Presented at the Fifth SIAM Conference on Optimization, Victoria, Canada, 1996.

[9] D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitz Optimization Without the Lipschitz Constant. *Journal of Optimization Theory and Applications*, 78(1):157–181, 1993.

[10] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

[11] H.J. Kushner. A Versatile Model of a Function of Unknown and Time Varying Form. *Journal of Mathematical Analysis and Applications*, 5:150–167, 1962.

[12] H.J. Kushner. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. *Journal of Basic Engineering*, 86:97–106, 1964.

[13] A.V. Levy and A. Montalvo. The Tunneling Algorithm for the Global Minimization of Functions. *SIAM Journal on Scientific and Statistical Computing*, 6(1):15–29, 1985.

[14] M.J.D. Powell. *Approximation Theory and Methods.* Cambridge University Press, 1981.

[15] M.J.D. Powell. The Theory of Radial Basis Function Approximation in 1990. In W.A. Light, editor, *Advances in Numerical Analysis, Volume 2: Wavelets, Subdivision Algorithms and Radial Basis Functions*, pages 105–210. Oxford University Press, 1992.

[16] M.J.D. Powell. Recent research at Cambridge on radial basis functions. In M.W. Müller, M.D. Buhmann, D.H. Mache, and M. Felten, editors, *New Developments in Approximation Theory, International Series of Numerical Mathematics, Vol. 132*, pages 215–232. Birkhauser Verlag, Basel, 1999.

[17] R. Schaback. Comparison of radial basis function interpolants. In K. Jetter and F. Utreras, editors, *Multivariate Approximations: From CAGD to Wavelets*, pages 293–305. World Scientific, Singapore, 1993.

[18] R. Storn and K. Price. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[19] A. Törn and A. Žilinskas. *Global Optimization.* Springer, Berlin, 1987.

[20] H. Whitney. Analytic extension of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36:63–89, 1934.

[21] A. Žilinskas. Axiomatic Approach to Statistical Models and their Use in Multimodal Optimization Theory. *Mathematical Programming*, 22(1):104–116, 1982.

[22] A. Žilinskas. Axiomatic Characterization of a Global Optimization Algorithm and Investigation of its Search Strategy. *Operations Research Letters*, 4(1):35–39, 1985.