J Glob Optim (2010) 48:173–182
DOI 10.1007/s10898-009-9517-9

On similarities between two models of global optimization: statistical models and radial basis functions

Antanas Žilinskas

Received: 14 December 2009 / Accepted: 16 December 2009 / Published online: 3 January 2010

Abstract Construction of global optimization algorithms using statistical models and radial basis function models is discussed. A new method of data smoothing using radial basis function and least squares approach is presented. It is shown that the P-algorithm for global optimization in the presence of noise based on a statistical model coincides with the corresponding radial basis algorithm.

Keywords Global optimization · Optimization in the presence of noise · Statistical models · Radial basis function

1 Introduction

One of the approaches to tackle global optimization (GO) problems is based on statistical models of objective functions. The idea to construct a GO method using a stochastic function for a model of objective functions was originally proposed by Kushner [8,9]. The method by Kushner has later been axiomatically justified in [18,19] and named the P-algorithm. In the seventies of the last century active research of statistical model-based GO methods was started in Russia and in Lithuania. For the results of this period we refer to [13,20]. In Russia the research of statistical models-based GO was initiated by Strongin.

The statistical models-based approach to GO has been proven successful for optimization of expensive objective functions; see e.g. monographs [10,14–16]. Similar algorithms have been developed recently using different, a so-called radial basis function (RBF), model [6,7,11]. Formal similarity between the algorithms based on statistical and RBF models in the case of optimization without noise has been shown in [6]. In the present paper we continue analysis of such similarity.

GO in the presence of noise is an especially difficult problem attracting considerably less attention from researchers than GO without noise. For the early algorithms based on

A. Žilinskas (✉)
Institute of Mathematics and Informatics, Vilnius, Lithuania
e-mail: antanasz@ktl.mii.lt

statistical models we refer to [8,9,12,17]. Recent publications [3,4] indicate some renewal of interest in such algorithms. Noisy optimization problems are tackled also by means of other stochastic approaches; see e.g. [1]. For stimulation of research in this subject it can be helpful to establish common aspects of different approaches.

In the present paper we present a simplified proof of similarity between the P-algorithm (where a Gaussian stochastic function is used for the statistical model) and the RBF interpolation based algorithm (where basis function coincides with covariance function of the Gaussian stochastic function used for a model in the P-algorithm). Further we show that the similarity between statistical and RBF models-based algorithms extends also for the case of optimization in the presence of noise. The proven similarity can be helpful to integrate theoretical results and experience of both approaches for development of efficient algorithms of global optimization especially in the presence of noise.

2 P-algorithm and radial basis algorithm for optimization without noise

In this section we show equivalence of the P-algorithm and of the radial basis algorithm [6] in the case of optimization without noise. Although this question has been investigated in [6], we proceed with slightly different assumptions enabling to shorten the proof of the equivalence, and to make the proof more transparent. The rather straight forward generalization of our proof to the case of noisy optimization is presented in Sect. 4.

Let us consider the minimization problem

$$\min_{x\in A} f(x),\ A \subseteq \mathbb{R}^d, \tag{1}$$

at the $(n+1)$-th minimization step where function values at the points x_i are known (computed or observed): $y_i = f(x_i),\ i = 1,\dots,n$. An optimization algorithm should define the next observation point x_{n+1}, and in the most general case x_{n+1} can depend on all available information on $f(x)$ including $x_i,\ y_i,\ i = 1,\dots,n$.

The P-algorithm is defined using a stochastic function $\xi(x)$ for a statistical model of objective functions, and the idea is to maximize the probability of improvement at the current minimization step defined with respect to the model [15,19]:

$$x_{n+1} = \arg\max_{x\in A} \mathbf{P}\{\xi(x) \le \tilde{y}_{on} | \xi(x_i) = y_i,\ i = 1,\dots,n\}, \tag{2}$$

where $\tilde{y}_{on}$ is an aspiration level at the $(n+1)$-st minimization step, e.g. $\tilde{y}_{on} = y_{on} - \varepsilon_n,\ y_{on} = \min_{i=1,\dots,n} y_i,\ \varepsilon_n > 0$. Assuming $\xi(x)$ is a Gaussian stochastic function the maximization in (2) can be reduced to the maximization of

$$\frac{\tilde{y}_{on} - m_n(x|\xi(x_i) = y_i,\ i = 1,\dots,n)}{s_n(x|\xi(x_i) = y_i,\ i = 1,\dots,n)}, \tag{3}$$

where $m_n(x|x_i, y_i, i = 1,\dots,n)$ and $s_n^2(x|\xi(x_i) = y_i,\ i = 1,\dots,n)$ denote the conditional mean and the conditional variance of $\xi(x)$ with respect to $\xi(x_i) = y_i,\ i = 1,\dots,n$. Such a characterization of the P-algorithm is sufficient for further analysis; for implementation aspects we refer to [15,16].

Another global optimization method is based on a deterministic RBF model; see [6]. Similarly as before it uses the values of the objective function $y_i = f(x_i),\ i = 1,\dots,n$, at the already evaluated points. A value of an objective function $f(\cdot)$ at an arbitrary point $x \in \mathbb{R}^d$ can be interpolated by the radial function

$$\mu_n(x|x_i, y_i, i = 1, \ldots, n) = \sum_{i=1}^{n} \lambda_i \phi(||x - x_i||), \tag{4}$$

where $\phi(\cdot)$ is a chosen positive definite function, and the coefficients λ_i are obtained from the condition of interpolating the data $(x_i, y_i = f(x_i))$, $i = 1, \ldots, n$, by means of (4). A naive idea to perform the next observation at the minimum point of the response surface defined by (4) should be rejected because of strong locality of such strategy: the minimum point of (4) normally is in close vicinity of (even can coincide with) the best found point. Disadvantages of the naive strategy are discussed in more detail in [15]. An elaboration of the algorithm proposed by H.-M.Gutmann is discussed below, after a few remarks concerning the interpolating function (4).

In this paper the standard RBFs [2] are considered without the extra polynomial summands used in [6]; the latter enable the involvement of non positive definite $\phi(\cdot)$. We focus on the standard RBFs since in this case $\phi(\cdot)$ can be interpreted as a covariance function of a stochastic function. Many such basis functions are available, e.g. the Gaussian function $\phi(r) = \exp(-\gamma r^2)$, $r \geq 0$, where $\gamma > 0$. The coefficients λ_i are defined by the system of linear equations $\mu_n(x_i|\cdot) = y_i, i = 1, \ldots, n$, whose solution is guaranteed by the positive definiteness of the matrix $\Psi = (\phi(||x_i - x_j||))$.

Despite very different theoretical concepts of statistical models and RBF models, the heuristic ideas of the algorithms based on both concepts are similar. The P-algorithm for current observation chooses the point where the probability to observe a function value less than $\tilde{y}_{on}$ is maximal. The radial basis algorithm performs the current observation of $f(\cdot)$ at the 'most likely' site for the value of $f(\cdot)$ to be equal to the target value $\tilde{y}_{on}$, where the term 'most likely' is interpreted in the sense of the concept of 'bumpiness' introduced in [6] and briefly explained below.

A natural criterion to evaluate the interpolation is a seminorm of the considered interpolating function

$$||\mu_n(\cdot|x_i, y_i, i = 1, \ldots, n)||^2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \phi(||x_i - x_j||) \right); \tag{5}$$

for the reasoning of this seminorm we refer to [5].

Let the known function values $y_i = f(x_i)$, $i = 1, \ldots, n$, be interpolated by means of RBF. The point of the next observation is chosen aiming to get the target value of the objective function equal to $\tilde{y}_{on}$, and minimally increasing the norm of the interpolant implied by the augmentation of x_i, y_i, $i = 1, \ldots, n$, with $x_{n+1}, \tilde{y}_{on}$. Let us find a point x_{n+1} such that the norm of $\mu_{n+1}(x|x_i, y_i, i = 1, \ldots, n + 1)$ is minimal, where $y_{n+1} = \tilde{y}_{on}$. Such a point for the value $\tilde{y}_{on}$ seems 'most likely' assuming that the behavior of the minimal norm interpolating function is most natural of all interpolators. According to the terminology of [6] such choice of x_{n+1} minimizes 'bumpiness' of the response surface. Formally, the algorithm is constructed sequentially tuning the radial function interpolant by means of minimization of the seminorm with respect to the forecasted global minimum value $\tilde{y}_{on}$.

In the formulas below we use the shorthand $\mu_n(x) = \mu_n(x|\cdot)$, and the notations

$$\Lambda = (\lambda_1, \ldots, \lambda_n)^T, \ \Phi(x) = (\phi(||x - x_1||), \ldots, \phi(||x - x_n||))^T. \tag{6}$$

Formula (4) using these notations can be rewritten in the form $\mu_n(x) = \Lambda^T \Phi(x)$.

Similarly the RBF interpolator using an extended set of data (x_i, y_i), $i = 1, \ldots, n+1$, is defined by the formulas

$$\mu_{n+1}(x) = \sum_{i=1}^{n+1} \omega_i \phi(||x - x_i||) = \Omega^T \cdot \begin{pmatrix} \Phi(x) \\ \phi(||x - x_{n+1}||) \end{pmatrix}. \tag{7}$$

The vectors of coefficients Λ^T and Ω^T can be calculated as solutions of systems of linear equations corresponding to the condition of interpolation

$$\Lambda = \Psi^{-1} \cdot Y, \; Y = \begin{pmatrix} y_1 \\ \ldots \\ y_n \end{pmatrix},$$

$$\Psi = \begin{pmatrix} \phi(0) & \ldots & \phi(||x_1 - x_n||) \\ \ldots & \ldots & \ldots \\ \phi(||x_n - x_1||) & \ldots & \phi(0) \end{pmatrix},$$

$$\Omega = \mathbf{\Psi}^{-1} \cdot \begin{pmatrix} Y \\ \tilde{y}_{on} \end{pmatrix}, \tag{8}$$

$$\mathbf{\Psi} = \begin{pmatrix} \Psi & \Phi(x_{n+1}) \\ \Phi(x_{n+1})^T & \phi(0) \end{pmatrix}. \tag{9}$$

The squared seminorm of $\mu_{n+1}(\cdot)$ is equal to

$$||\mu_{n+1}(\cdot)||^2 = \Omega^T \mathbf{\Psi} \Omega = (Y^T, \tilde{y}_{on}) \mathbf{\Psi}^{-1} \begin{pmatrix} Y \\ \tilde{y}_{on} \end{pmatrix}, \tag{10}$$

where the expression of Ω from (8) is taken into account.

To invert the matrix $\mathbf{\Psi}$ presented as a block matrix in (9) the formula of Frobenius can be applied

$$\mathbf{\Psi}^{-1} = \begin{pmatrix} \Psi^{-1} + \frac{1}{h}\Psi^{-1}\Phi(x_{n+1})\Phi(x_{n+1})^T\Psi^{-1} & -\frac{1}{h}\Psi^{-1}\Phi(x_{n+1}) \\ -\frac{1}{h}\Phi(x_{n+1})^T\Psi^{-1} & \frac{1}{h} \end{pmatrix}, \tag{11}$$

where

$$h = \phi(0) - \Phi(x_{n+1})^T \Psi^{-1} \Phi(x_{n+1}). \tag{12}$$

Calculation of the norm (10) using the latter expression of $\mathbf{\Psi}^{-1}$ gives the following result

$$\begin{aligned} ||\mu_{n+1}(\cdot)||^2 &= (Y^T, \tilde{y}_{on}) \mathbf{\Psi}^{-1} \begin{pmatrix} Y \\ \tilde{y}_{on} \end{pmatrix} \\ &= (Y^T, \tilde{y}_{on}) \begin{pmatrix} \Psi^{-1}Y + \frac{1}{h}\Psi^{-1}\Phi(x_{n+1})\Phi(x_{n+1})^T\Psi^{-1}Y - \frac{\tilde{y}_{on}}{h}\Psi^{-1}\Phi(x_{n+1}) \\ -\frac{1}{h}\Phi(x_{n+1})^T\Psi^{-1}Y + \frac{\tilde{y}_{on}}{h} \end{pmatrix} \\ &= \Lambda\Psi\Lambda + \frac{(\tilde{y}_{on} - \Phi(x_{n+1})^T\Psi^{-1}Y)^2}{\phi(0) - \Phi(x_{n+1})^T\Psi^{-1}\Phi(x_{n+1})}. \end{aligned} \tag{13}$$

From (13) the subsequent equality follows

$$||\mu_{n+1}(\cdot)||^2 = ||\mu_n(\cdot)||^2 + \frac{(\tilde{y}_{on} - \Phi(x_{n+1})^T\Psi^{-1}Y)^2}{\phi(0) - \Phi(x_{n+1})^T\Psi^{-1}\Phi(x_{n+1})},$$

where the first summand does not depend on x_{n+1}. Therefore for the next observation a minimum point of the following function

$$\arg\min_{x_{n+1}} \frac{(\tilde{y}_{on} - \Phi(x_{n+1})^T \Psi^{-1} Y)^2}{\phi(0) - \Phi(x_{n+1})^T \Psi^{-1} \Phi(x_{n+1})}, \tag{14}$$

should be chosen.

Let us return to (3) and specify $\xi(x)$, $x \in \mathbb{R}^n$ as a homogeneous isotropic Gaussian random field with zero mean and covariance function $\phi(\cdot)$. The conditional mean and the conditional variance of $\xi(x)$ with respect to $\xi(x_i) = y_i$, $i = 1, \ldots, n$, is equal to

$$m_n(x|\xi(x_i) = y_i,\ i = 1, \ldots, n) = \Phi(x)^T \Psi^{-1} Y,$$
$$s_n^2(x|\xi(x_i) = y_i,\ i = 1, \ldots, n) = \phi(0) - \Phi(x)^T \Psi^{-1} \Phi(x),$$

correspondingly. Therefore the maximization of (3) is reduced to the maximization of

$$\frac{\tilde{y}_{on} - \Phi(x_{n+1})^T \Psi^{-1} Y}{\sqrt{\phi(0) - \Phi(x_{n+1})^T \Psi^{-1} \Phi(x_{n+1})}}. \tag{15}$$

The minimization of (14) is equivalent to the maximization of (15), since expression (14) is equal to the squared expression of (15), and the target level $\tilde{y}_{on}$ is naturally chosen less than $\min m_n(x|\xi(x_i) = y_i,\ i = 1, \ldots, n)$. The latter conclusion means that the statistical model-based P-algorithm and the RBF model-based algorithm are identical.

3 A new version of least squares radial basis approximation

RBFs are widely used for interpolation. Besides of interpolation they can be used also to smooth noisy observations. However, the standard RBF approximation obtained by the least squares method (see, e.g. [2]) is not appropriate for the generalization of the H.-M. Gutmann algorithm for optimization of noisy functions. In this section a new method for RBF approximation of noisy functions is proposed.

Before constructing an algorithm for minimization in the presence of noise we will consider the problem of evaluation of a function value $f(x)$ using observations corrupted by noise $z_i = f(x_i) + \zeta_i$, $i = 1, \ldots, n$ where ζ_i are independent random variables. The standard RBF model is designed for interpolation, and its application to data smoothing requires some modification. The least squares technique can be applied to develop such a modification [2]. Since the interpolation of the noisy function values is not reasonable the surrogate function values $\tilde{y}_j$, $j = 1, \ldots, m$, at a smaller number ($m < n$) of the selected centers $\tilde{x}_j$, $j = 1, \ldots, m$, are interpolated. The surrogate function values $\tilde{y}_j$, $j = 1, \ldots, m$ are defined by means of the least squares method minimizing the mean square error of approximation of the data x_i, z_i, $i = 1, \ldots, n$, by means of the RBF interpolating $\tilde{x}_j, \tilde{y}_j$, $j = 1, \ldots, m$. The parameter m defines smoothness of the approximating RBF. Application of such a least squares based quasi interpolation in constructing of an optimization algorithm is difficult already because of ambiguity in the choice of m; the choice of centers $\tilde{x}_j$, $j = 1, \ldots, m$, is even more problematic.

We propose a new version of the least squares approximation based on general concept of minimum norm approximating functions, and on the supposed knowledge of variance of ζ_i. Strictly speaking we propose to construct a minimum norm radial function

$$\mu_n(x|x_i, z_i, i = 1, \ldots, n) = \sum_{i=1}^{n} \lambda_i \phi(||x - x_i||), \tag{16}$$

such that

$$\sum_{j=1}^{n} (\mu_n(x_j|x_i, z_i, i = 1, \ldots, n) - z_j)^2 = n\sigma^2, \tag{17}$$

where σ^2 denotes the variance of ζ_i.

Necessary first order conditions of minimum of $||\mu_n(\cdot)||^2$ subject to constraint (17) can be written as

$$\nabla \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \phi(||x_i - x_j||) + \nu \sum_{j=1}^{n} \left(z_j - \sum_{i=1}^{n} \lambda_i \phi(||x_j - x_i||) \right)^2 \right)$$

$$= \nabla \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \phi_{ij} - 2\nu \sum_{j=1}^{n} \left(z_j - \sum_{i=1}^{n} \lambda_i \phi_{ij} \right) \nabla \sum_{i=1}^{n} \lambda_i \phi_{ij} = 0, \tag{18}$$

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n} \lambda_i \phi_{ij} - z_j \right)^2 = n\sigma^2, \tag{19}$$

where ∇ denotes the gradient operator with respect to Λ, $\phi_{ij} = \phi(||x_i - x_j||)$, and ν is the Lagrange multiplier corresponding to the constraint (17).

Taking symmetry and positive definiteness of Ψ into account (18) can be reduced to the following vector equation

$$\Psi\Lambda - \nu\Psi Z + \nu\Psi\Psi\Lambda = 0,$$

which can be simplified to

$$Z - \Psi\Lambda = \frac{1}{\nu}\Lambda. \tag{20}$$

Substitution of Z from (20) into the equation (19) gives

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n} \lambda_i \phi_{ij} - z_j \right)^2 = (Z - \Psi\Lambda)^T (Z - \Psi\Lambda) = \frac{1}{\nu^2} \Lambda^T \Lambda = n\sigma^2. \tag{21}$$

Solution (with respect to ν and Λ) of the system of equation defined by (20) and (21) gives the vector of coefficients Λ necessary for definition of the radial function (16) smoothing the considered data x_i, z_i, $i = 1, \ldots, n$. The approximating function can be written in the following form

$$\mu_n(x|x_i, z_i, i = 1, \ldots, n) = \sum_{i=1}^{n} \lambda_i \phi(||x - x_i||) = \Phi(x)^T (\Psi + \frac{1}{\nu} I)^{-1} Z, \tag{22}$$

where I denotes the unit matrix. Formula (22) coincides with the formula of the conditional mean of a homogenous isotropic Gaussian random function (with covariance function $\phi(\cdot)$ and a priori mean equal to zero) up to summands in the matrix diagonal. In (22) the diagonal element is equal to $\phi_{ii} + \frac{1}{\nu}$ but in the formula of conditional mean of the Gaussian random function the diagonal element is equal to $\phi_{ii} + \sigma^2$.

For the construction of an optimization algorithm we consider in Sect. 4 the case when the available information is $z_i, i = 1, \ldots, n$, (i.e. n observations corrupted by noise) and $y_{n+1} = f(x_{n+1})$ (i.e. one observation without noise). In such a case a minimum norm radial function

$$\mu_{n+1}(x|x_i, z_i, i = 1, \ldots, n,\ x_{n+1}, y_{n+1}) = \sum_{i=1}^{n+1} \omega_i \phi(||x - x_i||) = \Omega^T \cdot \begin{pmatrix} \Phi(x) \\ \phi(||x - x_{n+1})|| \end{pmatrix}, \tag{23}$$

should be found taking two constraints into account

$$\sum_{j=1}^{n} (\mu_{n+1}(x_{n+1}|x_i, z_i, i = 1, \ldots, n,\ x_{n+1}, y_{n+1}) - z_j)^2 = n\sigma^2, \tag{24}$$

$$\mu_{n+1}(x_{n+1}|x_i, z_i, i = 1, \ldots, n,\ x_{n+1}, y_{n+1}) = y_{n+1}. \tag{25}$$

Necessary first order conditions of the minimum of the norm of (23), with the constraints (24), and (25) can be written as

$$\nabla \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \omega_i \omega_j \phi_{ij} - 2\nu_1 \sum_{j=1}^{n} \left(z_j - \sum_{i=1}^{n+1} \omega_i \phi_{ij} \right) \nabla \sum_{i=1}^{n+1} \omega_i \phi_{ij} + \nu_2 \nabla \sum_{i=1}^{n+1} \omega_i \phi_{n+1,i} = 0, \tag{26}$$

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n+1} \omega_i \phi_{ij} - z_j \right)^2 = n\sigma^2, \tag{27}$$

$$\sum_{i=1}^{n+1} \omega_i \phi_{n+1,i} = y_{n+1}, \tag{28}$$

where ν_1 and ν_2 are Lagrange multipliers corresponding to the constraints (27) and (28) correspondingly. After differentiation and simple rearrangements of terms in (26) the necessary conditions can be reduced to the following vector equation

$$\mathbf{\Psi\Omega} - \nu_1 \mathbf{\Psi Z} + \nu_1 \mathbf{\Psi\Psi\Omega} + \nu_2 \begin{pmatrix} \Phi(x_{n+1}) \\ \phi_{n+1,\,n+1} \end{pmatrix} = 0, \tag{29}$$

where

$$\mathbf{\Psi} = \begin{pmatrix} \Psi & \Phi(x_{n+1}) \\ \Phi(x_{n+1})^T & \phi_{n+1,\,n+1} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z \\ y_{n+1} \end{pmatrix}, \quad \mathbf{\Omega} = \begin{pmatrix} \Omega \\ \omega_{n+1} \end{pmatrix}. \tag{30}$$

Multiplication of (29) by $\mathbf{\Psi}^{-1}$ gives the following vector equation

$$\mathbf{\Omega} - \nu_1 \mathbf{Z} + \nu_1 \mathbf{\Psi\Omega} + \nu_2 \begin{pmatrix} 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix} = 0. \tag{31}$$

The last row of (31)

$$\omega_{n+1} - \nu_1 y_{n+1} + \nu_1 y_{n+1} + \nu_2 = 0,$$

shows that

$$\nu_2 = -\omega_{n+1}. \tag{32}$$

Taking (32) into account, equation (31) can be written in the form

$$\mathbf{\Omega} = \mathbf{\Upsilon}^{-1}\mathbf{Z},\ \mathbf{\Upsilon} = \begin{pmatrix} \Psi + \frac{1}{\nu_1} I & \Phi(x_{n+1}) \\ \Phi(x_{n+1})^T & \phi(0) \end{pmatrix}, \tag{33}$$

and the corresponding RBF coincides with the expression of the conditional mean of the corresponding Gaussian random function with respect to $(z_1, \ldots, z_n, y_{n+1})$:

$$\mu_{n+1}(x|x_i, z_i, i = 1, \ldots, n,\ x_{n+1}, y_{n+1}) = \Phi(x)^T \mathbf{\Omega} = \Phi(x)^T \mathbf{\Upsilon}^{-1}\mathbf{Z}.$$

The vector of RBF coefficients $\mathbf{\Omega}$ is defined by the system of two equations: (33) and

$$\nu_1^2 = \frac{\Omega^T \Omega}{n\sigma^2}. \tag{34}$$

4 P-algorithm and radial basis algorithm for optimization in the presence of noise

The P-algorithm for minimization in the presence of noise is a simple generalization of the corresponding algorithm for minimization without noise; it is defined by an expression similar to (2), where the conditional probability is calculated with respect to the noisy data. The noise is modelled by the independent random variables ζ_i, and the available information about the objective function is $(x_i, z_i),\ i = 1, \ldots, n, z_i = f(x_i) + \zeta_i$. Let $\xi(x)$ (a model of objective functions) be a stationary (homogenous and isotropic) Gaussian stochastic function with zero mean, unit variance, and correlation function $\phi(x)$ depending only on $||x||$. Let ζ_i be Gaussian random variables with zero mean and variance σ^2. An unknown value $f(x)$ can be approximated using the statistical model $\xi(\cdot)$ by the conditional mean of $\xi(x)$ with respect to $\xi(x_i) + \zeta_i = z_i,\ i = 1, \ldots, n$:

$$\begin{aligned} &\mathbf{E}(\xi(x)|\xi(x_i) + \zeta_i = z_i,\ i = 1, \ldots, n) \\ &= \Phi(x)^T \cdot \begin{pmatrix} \phi(0) + \sigma^2 & \ldots & \phi(||x_1 - x_n||) \\ \ldots & \ldots & \ldots \\ \phi(||x_n - x_1||) & \ldots & \phi(0) + \sigma^2 \end{pmatrix}^{-1} \cdot Z. \end{aligned} \tag{35}$$

The minimum and corresponding minimum point of (35) present a current approximation to the minimum and minimum point of $f(x)$. The next observation of the P-algorithm is at the point

$$x_{n+1} = \arg \max_{x \in A} \mathbf{P}\{\xi(x) \le \tilde{y}_{on} | \xi(x_i) + \zeta_i = z_i,\ i = 1, \ldots, n\},$$

where $\tilde{y}_{on}$ is an estimate of the global minimum, e.g. obtained by subtraction of a certain value ε_n from the current estimate of the global minimum defined as mentioned above. Sometimes a good estimate of the global minimum can be known before optimization.

Similarly as in the case without noise the essential constituent of the P-algorithm is the maximization of

$$\frac{\tilde{y}_{on} - m_n(x|\cdot)}{s_n(x|\cdot)}, \tag{36}$$

where $m_n(x|\cdot)$ should be substituted by (35) and $s_n(x|\cdot)$ should be substituted by the following expression

$$s_n(x|\xi(x)+\zeta_i=z_i,\ i=1,\ldots,n)$$
$$=\left(\phi(0)-\Phi(x)^T\cdot\begin{pmatrix}\phi(0)+\sigma^2 & \ldots & \phi(||x_1-x_n||)\\ \ldots & \ldots & \ldots\\ \phi(||x_n-x_1||) & \ldots & \phi(0)+\sigma^2\end{pmatrix}^{-1}\cdot\Phi(x)\right)^{\frac{1}{2}}. \quad (37)$$

The construction of a radial basis algorithm for minimization in the presence of noise is based on the idea of minimum norm approximation; as mentioned above this idea was expressively described in [6] as minimization of "bumpiness". To generalize the results concerning minimization without noise to the case of minimization in the presence of noise we proceed similarly as in Sect. 1. Suppose the results of observations corrupted by noise $(x_i,\ z_i,\ i=1,\ldots,n)$ are available. Parameter ν (Lagrange multiplier of the noise related constraint) is estimated by means of solution of the system of equations composed of (20) and (21), and a function value $f(x)$ is approximated using the formula (23) which coincides with the formula of conditional mean of the Gaussian random function (35) up to diagonal elements. A target value to reach at the current minimization step $\tilde{y}_{on}$ normally (if the global minimum is not known in advance) is obtained subtracting a certain value ε_n from the minimum of $\mu_{n+1}(\cdot|\cdot)$. The point of next evaluation of the objective function value is chosen as 'most likely' site for $\tilde{y}_{on}$ which is computed by means of the following minimization:

$$\min_{x\in A}||\mu_{n+1}(\cdot|x_i,\ z_i,\ i=1,\ldots,n,\ (x,\tilde{y}_{on}))||.$$

Since the noise related parameter ν is already evaluated, now the vector $\mathbf{\Omega}$ can be defined by (33,34) where ν_1 is replaced by ν. Therefore the corresponding squared seminorm is defined by

$$||\mu_{n+1}(\cdot)||^2=\mathbf{\Omega^T\Upsilon\Omega}=\mathbf{Z}^T\mathbf{\Upsilon^{-1}\Upsilon\Upsilon^{-1}Z}=\mathbf{Z}^T\mathbf{\Upsilon^{-1}Z}, \quad (38)$$

where the function value y_{n+1} at the point x_{n+1} is supposed equal to $\tilde{y}_{on}$, and $\mathbf{\Upsilon}$ is defined by (33).

Application of the Frobenius formula (11) to invert $\mathbf{\Upsilon}$ in (38) (similarly to the development of the formula (13)) gives the following result

$$||\mu_{n+1}(\cdot)||^2=Z^T(\Psi+\frac{1}{\nu}I)^{-1}Z+\left(\frac{\tilde{y}_{on}-Z^T(\Psi+\frac{1}{\nu}I)^{-1}\Phi(x_{n+1})}{\phi(0)-\Phi(x_{n+1})(\Psi+\frac{1}{\nu}I)^{-1}\Phi(x_{n+1})}\right)^2, \quad (39)$$

where the first summand (equal to the squared seminorm of $\mu_n(\cdot)$) does not depend on x_{n+1}, and the second summand (supposed to minimize with respect to x_{n+1}) coincides up to the noise related parameter with the squared function (36) appearing in the definition of the P-algorithm. Thus the structure of the P-algorithm and that of the radial basis algorithm are analogous, not only when minimizing without noise but also in the case of minimization in the presence of noise.

Such a conclusion facilitates construction of new algorithms combining advantages of statistical models based and RBF based approaches.

5 Conclusions

The similarity between the P-algorithm and the radial basis algorithm when it was shown by H.-M.Gutmann seemed rather surprising because of the difference of basic assumptions of both approaches. As shown in this paper, the similarity extends to minimization in the presence of noise.

Acknowledgments The research is supported by the Agency for International Science and Technology Development Programmes in Lithuania through the COST programme. The author is grateful for the remarks made by two anonymous referees, which helped improve the presentation and contribution of the paper.

References

1. Alexander, D., Bugler, D., Calvin, J., Romeijn, H., Sheriff, R.: Approximate implementation of pure random search in the presence of noise. J. Glob. Optim. **31**, 601–612 (2005)
2. Buhmann, M.: Radial Basis Functions: Theory and Implementations. Cambridge University Press, Cambridge (2003)
3. Calvin, J.: Nonadaptive univariate optimization for observations with noise. In: Törn, A., Žilinskas, J. (eds.) Models and Algorithms for Global Optimization, pp. 185–192. Springer, New York (2007)
4. Calvin, J., Zilinskas, A.: One-dimensional global optimization for observations with noise. Int. J. Comput. Math. Appl. **50**, 157–169 (2005)
5. Gutmann, H.-M.: On the semi norm of radial basis function interpolants. J. Approx. Theory **111**, 315–328 (2001)
6. Gutmann, H.-M.: A radial basis function method for global optimization. J. Glob. Optim. **19**, 201–227 (2001)
7. Holmström, K.: An adaptive radial basis algorithm for black box global optimization. J. Glob. Optim. **41**, 447–464 (2008)
8. Kushner, H.: A versatile stochastic model of a function of unknown and time-varying form. J. Math. Anal. Appl. **5**, 150–167 (1962)
9. Kushner, H.: A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. J. Basic Eng. **86**, 97–106 (1964)
10. Mockus, J.: Bayesian Approach to Global Optimization. Kluwer Academic Publishers, Dordrecht (1988)
11. Regis, R.G., Shoemaker, Ch.A.: Improved strategies for radial basis function methods for global optimization. J. Glob. Optim. **37**, 113–135 (2007)
12. Strongin, R.: Information method for multiextremal minimization with noisy observations (in Russian). Trans. USSR Acad. Sci. Ser. Eng. Cybern. **6**, 116–126 (1969)
13. Strongin, R.: Numerical Methods in Multiextremal Problems (in Russian). Nauka, Moscow (1978)
14. Strongin, R., Sergeyev, Y.: Global Optimization with Non-Convex Constraints. Kluwer Academic Publishers, Dordrecht (2000)
15. Torn, A., Zilinskas, A.: Global optimization. Lect. Notes Comput. Sci. **350**, 1–252 (1989)
16. Zhigljavsky, A., Zilinskas, A.: Stochastic Global Optimization. Springer, New York (2008)
17. Zilinskas, A.: Two algorithms for one-dimensional multimodal minimization. Optimization **12**, 53–63 (1981)
18. Zilinskas, A.: Axiomatic approach to statistical models and their use in multimodal optimization theory. Math. Program. **22**, 104–116 (1982)
19. Zilinskas, A.: Axiomatic characterization of a global optimization algorithm and investigation of its search strategies. Oper. Res. Lett. **4**, 35–39 (1985)
20. Zilinskas, A.: Global Optimization: Axiomatics of Statuistical Models, Algorithms, and Their Applications (in Russian). Mokslas, Vilnius (1986)