Journal of Mathematical Extension
Vol. 10, No. 3, (2016), 11-33
ISSN: 1735-8299
URL: http://www.ijmex.com

# A Smooth Method for Solving Non-Smooth Unconstrained Optimization Problems

**F. Rahmanpour***
Yazd University

**M. M. Hosseini**
Yazd University

**Abstract.** We consider unconstrained optimization problems using the expensive objective function in which the derivatives are not available. This property of problems can often impede the performance of optimization algorithms. Most algorithms usually determine a Quasi-Newton direction and then use line search technique. We propose a smoothing algorithm which is developed to modify trust region and to handle the objective function based on radial basis functions (RBFs). The value of objective function is reduced according to the relation with the predicted reduction of surrogate model. At each iteration we construct the quadratic model based on RBFs. The global convergence of the proposed method is studied. The numerical results are presented for some standard test problem to validate the theoretical results.

**AMS Subject Classification:** 65D05; 65K05; 90C30; 90C56
**Keywords and Phrases:** Derivative-free method, trust region method, non-smooth optimization, unconstrained optimization problems, radial basis functions

## 1. Introduction

Consider the nonlinear unconstrained optimization problem:

Received: August 2015; Accepted: February 2016
*Corresponding author

$$\min_{x} f(x) \tag{1}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$. It is assumed that some of the functions $f(x)$, may be non smooth.

Most of these approaches require derivatives of the objective function. We also assume that some derivatives of the objective is either unavailable or are computationally too expensive to obtain.

Recently, derivative-free trust region algorithms have been used increasingly [7, 13, 17, 23]. A common approach is to combine conventional algorithm such as genetic algorithms or pattern search with surrogate models to solve expensive problems. For instance, Booker et al. [4] and Jones et al. [12] proposed a method based on Kriging basis functions. In recent years, nonlinear optimization is perhaps one of the most common reasons for using derivative-free methods. Forming surrogate models by interpolation has been proposed by Winfield [22] and reviewed by Powel [17] and Conn [7]. Wild, Regis, and Shoemaker [19] constructed a surrogate model based on RBFs.

The present paper gives a new derivative-free method for the solution of (1), which belongs to the class of trust region methods for optimization. Our aim in this paper is to find an efficient algorithm for the global solution of optimization problems. At each iteration a quadratic surrogate model is assumed to approximate the objective function $f(x)$ based on RBFs. we have chosen the position of interpolation points within a sphere of radius $\Delta > 0$ around the trial point. The important idea is the assumption that the interpolation points exist and can good approximate $f(x)$ in small spheres. When the current trial point is not enough close to a local minimum, we update the interpolation points and construct a new model by RBFs.

In the previous methods, whenever a trial point did not decrease the objective function as expected, one of the interpolation points was replaced by another evaluated point. In our approach, all the interpolation points are changed at each iteration, if necessary. Since evaluation of objective function is computationally expensive, we stress the importance of having complete knowledge of all points previously evaluated by the algorithm. There is a fundamental difference between our method and previous algorithms, where, in order to reduce linear algebraic computational costs, with small or large a trust region we can still obtain a good approximation. The proposed method will guarantee global convergence. Also, models based on RBFs have been shown to be of interest for global optimization.

This paper is organized as follows: In Section 2, the surrogate model is introduced. In Section 3, the RBFs are described. In Section 4, we present derivative–free optimization. Section 5, gives a summary of the surrogate model based on RBFs. In Section 6, the algorithm is introduced and its convergence properties are established. Numerical results for some examples are reported in the last Section.

Throughout the paper $||\ .\ ||$ denotes for Euclidean norm and for simplicity we also use subscripts to denote functions evaluated at iterates, for example, $f_k = f(x_k)$, $g_k = g(x_k)$ and $H_k = H(x_k)$.

## 2. Surrogate Method

We consider the following unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a function which is not necessarily differentiable. In this paper, we propose a surrogate model that is smooth, easy to minimize and that approximate well the objective $f(x)$.

We remark that surrogate model is referred to as a technique that uses the sample points to build a surrogate function, which is sufficient to predict the behavior of the objective function.

### 2.1 Quadratic surrogate model

Powell [17] and Conn et al. [7, 8] proposed a quadratic surrogate model as follows:

$$Sm(x_k + s) = f_k + g_k^T s + \frac{1}{2} s^T H_k s, \tag{2}$$

where $g_k = \nabla f(x_k)$ and $H_k = \nabla^2 f(x_k)$. When $f(x)$ is twice differentiable and admits a hessian matrix, $H(x)$ which will always be positive definite.

The goal is to construct the surrogate model $Sm(x_k)$ instead of the objective function $f(x)$, which is computationally simple and inexpensive with good analytical properties. It could be used in optimization because of its simplicity and a suitable algebraic form.

To build a quadratic model, we define the trust region $B_k := \{x \in \mathbb{R}^n : ||x - x_k|| < \Delta_k\}$. At each iteration of the surrogate method, the solution of optimization problem inside $B_k$ [9, 16], as

$$\min_s Sm(x_k + s) \quad s.t. \quad ||s|| < \Delta_k, \tag{3}$$

is needed, for some trust region with radius $\Delta_k > 0$. Ratio of the actual $f(x)$ over the predicted $Sm(x)$ is as follows:

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{Sm_k(x_k) - Sm_k(x_k + s_k)}. \tag{4}$$

Given the standard trust region $0 \leqslant \eta_0 \leqslant \eta_1 < 1$, $0 < \gamma_0 < 1 < \gamma_1$, $0 < \Delta_k \leqslant \Delta_{\max}$ and $x_k \in \mathbb{R}^n$, we define a model $Sm_k$ on $B_k$, and compute a step $s_k$ such that $x_k + s_k \in B_k$, in order to sufficiently reduces the model $Sm(x_k)$.

By accepting the trial point $x_k$, we compute $f(x_k + s_k)$ and $\rho_k$ using (4), then update the surrogate model parameters as follows,

$$x_{k+1} = \begin{cases} x_k + s_k & if \rho_k \geqslant \eta_0 \\ x_k & o.w, \end{cases}$$

and

$$\Delta_{k+1} = \begin{cases} \Delta_k & \eta_0 \leqslant \rho_k < \eta_1, \\ \min\{\gamma_1 \Delta_k, \Delta_{\max}\} & \rho_k \geqslant \eta_1, \\ \gamma_0 \Delta_k & \rho_k < \eta_0. \end{cases}$$

The following assumptions are considered in this section:

1. $f(x)$ is a two times differentiable function.

2. $\{x_k\}$ is a bounded sequence.

Suppose that these assumption holds. Let $s_k$ be a solution of subproblem (3), the following lemma, which can be obtained from the well-known result (Powel) is needed [17].

**Lemma 2.1.1.** *Subproblem (3) satisfies a sufficient decrease condition of the form:*

$$Sm_k(x_k) - Sm_k(x_k + s_k) \geqslant \frac{c}{2}||g_k|| \min(\Delta_k, \frac{||g_k||}{||H_k||}),$$

*for some constant* $c \in (0, 1)$. *We also assume that* $\frac{||g_k||}{||H_k||} = +\infty$ *when* $H_k = 0$.

Now the main questions are as follows: How to build surrogate models, and how to evaluate the accuracy of surrogate models?

## 2.2 Trust region based on the Cauchy point

The line search methods can be globally convergent. We seek the optimal solution of the subproblem (2). It is enough for purposes of global convergence to find an approximate solution $s_k$ that lies within the trust region and gives a sufficient reduction in the model. The sufficient reduction can be quantified in terms of the Cauchy point.

**Definition 2.2.1.** The Cauchy point, given current values $x_k$, $f_k$, $\nabla f_k$, $\Delta_k$, is a point that solves the quadratic model (3) along the direction the minimizers the linear model

$$\min_s l(x_k + s) = f_k + g_k^T s. \tag{5}$$

To calculate the Cauchy point, which we denote by $s_k^c$. We find the vector $s_k^p$ that solves a linear model (5), that is

$$s_k^p = -\Delta_k \frac{\nabla f_k}{||\nabla f_k||}.$$

So, the Cauchy is

$$s_k^c = \tau_k s_k^p,$$

where

$$\tau_k = \begin{cases} 1 & g_k^T H_k g_k \leqslant 0 \\ \min\{\frac{||g_k||^3}{\Delta_k g_k^T H_k g_k}, 1\} & o.w, \end{cases}$$

where $g_k = \nabla f(x_k)$ and $H_k = \nabla^2 f(x_k)$.

## 3. Radial Basis Functions Interpolation

RBFs are widely used for scattered data interpolation. A multivariate interpolation can be stated as follows: Given data $(x_i, f_i)$, $i = 1, \ldots, N$, with $x_i \in \mathbb{R}^n$, $f_i \in \mathbb{R}$, we find a continuous function $Sm(x)$ such that $Sm(x_i) = f_i$, $i = 1, \ldots, N$.

The function $Sm(x)$ is assumed to be given by a linear combination of RBFs, that is,

$$Sm_k(x_k + s) = \sum_{i=1}^{N} \lambda_i \varphi(||s - y^i||) + V(s), \tag{6}$$

where RBF $\varphi(||s - y^i||)$ is centered at point $y^i, i = 1, \ldots, N$. Note that we have $V(s) = \sum_{j=1}^{M} \gamma_j \nu_j(s)$, where $\nu = \{\nu_1(s), \ldots, \nu_M(s)\}$ is an ordered basis for the linear space $\Pi_{M-1}^n$, the space of polynomials of total degree less than $M - 1$ with $n$ variables and $\{\lambda_j\}_{j=1}^N$ are the unknown RBFs cofficients. $Sm(x)$ as defined by (6) has $M$ degrees of freedom. To overcome additional degrees of freedom two constraints are imposed as follows,

$$Sm(x_i) = f_i, \qquad i = 1, \ldots, N, \tag{7}$$

$$\sum_{i=1}^{N} \lambda_i \nu_k(s_i) = 0, \qquad k = 1, \ldots, M. \tag{8}$$

Conditions (7) and (8) can be written in matrix form:

$$\begin{bmatrix} \Phi & V \\ V^T & \emptyset \end{bmatrix} \begin{bmatrix} \lambda \\ \nu \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}. \tag{9}$$

Where $\lambda \in \mathbb{R}^n$ is the undetermined coefficient vector. For the sake of clarity, the matrix $\Phi$ is in the form:

$$\Phi = \begin{bmatrix} \varphi(||x_1 - x_1||) & \cdots & \varphi(||x_1 - x_N||) \\ \vdots & & \vdots \\ \varphi(||x_N - x_1||) & \cdots & \varphi(||x_N - x_N||) \end{bmatrix}_{N\times N}.$$

It can be seen that(9) is well-posed if the coefficient matrix is non-singular [3]. Micchell [15] proved that the interpolation problem in equation (9) is solvable when the following two conditions are met:

1. The points $\{x_j\}_{j=1}^N$ are distinct.
2. The RBFs are used are strictly conditionally positive definite.

**Definition 3.1.** [6, 20] *Let $\nu$ be a basis for $\pi_{M-1}^n$, with the convention that $\pi = \emptyset$ if $M = 0$. A function $\varphi$ is said to be conditionally positive definite (CPD) of order $M$ if for all distinct points $Y \subset \mathbb{R}^n$ and all $\lambda \neq 0$, satisfying $\sum_{i=1}^N \lambda_i \pi(x_i) = 0$, the quadratic form $\sum_{i,j=1}^N \lambda_i \varphi(||x_j - x_i||)\lambda_i$ is positive [6, 20].*

Some of the most popular (twice continuously differentiable) RBFs are shown in Table 1.

**Table 1:** Some examples of popular RBFs and their orders of conditional positive definiteness

| $\phi(r)$ | Order | Parameters | Example |
|---|---|---|---|
| $r^\beta$ | 2 | $\beta \in (2,4)$ | $Cubic, r^3$ |
| $(c^2+r^2)^\beta$ | 2 | $c>0, \beta \in (1,2)$ | MqI, $(c^2+r^2)^{\frac{3}{2}}$ |
| $-(c^2+r^2)^\beta$ | 1 | $c>0, \beta \in (0,1)$ | MqII, $-(c^2+r^2)^{\frac{1}{2}}$ |
| $(c^2+r^2)^{-\beta}$ | 0 | $c>0, \beta>0$ | Inv.Mq, $(c^2+r^2)^{-\frac{1}{2}}$ |
| $Exp(-c^2r^2)$ | 0 | $c>0$ | Gaussian, $Exp(-c^2r^2)$ |

## 4. Derivative Free Optimization

In this section, we suppose that $f(x)$ is a function from $\mathbb{R}^n$ into $\mathbb{R}$ which is not necessarily smooth. The algorithm is based on approximating the function (1) by a positive definite quadratic model. The main idea is to use the available values of $f(x)$ and building a quadratic model by interpolating within a trust region.

Suppose that in the current $x_k$, we have the sample points $Y = \{y^1 = 0, y^2, \ldots, y^N\}$, with $y^i \in \mathbb{R}^n$, $i = 1, \ldots, N$, which contains the points closest to $x_k$ in current iterate. We wish to construct a quadratic model of the form as:

$$Sm_k(x_k + s) = f_k + g_k^T s + \frac{1}{2} s^T H_k s, \tag{10}$$

where the vector $g \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times n}$ is a symmetric matrix. By imposing the interpolation condition in what follows:

$$Sm_k(x + y^j) = f(x + y^j), \qquad j = 1, \ldots, N, \tag{11}$$

it is now needed to evaluate $Sm(x + s)$ on $N = \frac{1}{2}(n + 1)(n + 2)$ points to find an approximating quadratic form, where $n$ is the number of variables [3, 7, 11].

We consider $\{\varphi_i(.)\}_{i=1}^N$ as a basis for the linear space of $n$-dimensional quadratic function. The quadratic function (10) can be expressed as

$$Sm_k(x + y^j) = \sum_{i=1}^{N} \lambda_i \varphi_i(y^j), \qquad j = 1, \ldots, N.$$

For some coefficients $\lambda_i$, which could be determined from the interpolation condition (11),

$$\sum_{i=1}^{N} \lambda_i \varphi_i(y^j) = f(x_k + y^j), \qquad j = 1, \ldots, N.$$

$\lambda_i$, $i = 1 \ldots, N$, are unique if the determinant of the matrix

$$\begin{bmatrix} \varphi_1(y^1) & \cdots & \varphi_1(y^N) \\ \vdots & & \vdots \\ \varphi_N(y^1) & \cdots & \varphi_N(y^N) \end{bmatrix},$$

then, iteratively we optimize and update the surrogate model $Sm_k$ to reach a satisfactory solution.

## 5. Surrogate Methods Based on Radial Basis Functions

In this section, the relevance of the surrogate methods and RBFs is considered. Suppose that

$$Sm(x+s) = \sum_{i=1}^{N} \lambda_i \varphi_i(s) + \sum_{k=1}^{M} \gamma_k \nu_k(s) = \Lambda^T \Phi(s) + \Gamma^T V(s).$$

This model is twice dfferentiable and is important for the convergence part of our method [11, 18]. This study considers interpolation condition at the points of $Y$:

$$Sm_k(x_k + y^i) = f(x_k + y^i), \qquad \forall y^i \in Y.$$

Let $\Phi \in \mathbb{R}^{N\times N}$, $V \in \mathbb{R}^{N\times M}$ be the matrices defined by $\Phi_{ij} = \varphi(||y^i - y^j||)$ and $v_{ij} = \nu_i(y^j)$. Then the interpolation condition can be expressed as $\Phi\Lambda + V\Gamma = f$. By using RBFs, we get the following linear system of equations

$$\begin{bmatrix} \Phi & V \\ V^T & \emptyset \end{bmatrix} \begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}$$

and then,

$$\Rightarrow \begin{bmatrix} \Phi & V \\ \emptyset & -V^T\Phi^{-1}V \end{bmatrix} \begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix} = \begin{bmatrix} f \\ -V^T\Phi^{-1}f \end{bmatrix}, \qquad (12)$$

with the solution $\Gamma = (V^T\Phi^{-1}V)^{-1}V^T\Phi^{-1}f$, $\Lambda = \Phi^{-1}(f - V\Gamma)$.

Sufficient condition for the solvability of system (12) is that the points in $Y$ to be distinct and yield a $V^T$ of full column rank.

Suppose that $V^T = QR$ , where $R \in \mathbb{R}^{(n+1)\times(n+1)}$. If $Z$ is an orthonormal basis for the null space of $V$ [2], using the condition (8), follows that $\Lambda \in \aleph(V)$. Therefore, $\Lambda = Zw$. According to (12), $\Phi\Lambda + V^T\Gamma = f$. Multiplying by $Z^T$ from left gives, $Z^T\Phi\Lambda + Z^TV^T\Gamma = Z^Tf$. Keeping in mind that Z is an orthonormal basis for the null space $V$, we obtain that $Z^TV^T\Gamma = 0$. Hence

$$Z^T\Phi Zw = Z^Tf. \tag{13}$$

Now, we can obtain $w$ from (13) and thus we can compute the vector $\Lambda$. By introducing the RBFs based on cubic spline [1, 6] which is the smoothest functions interpolation and conditional positive definiteness, then $Z^T\Phi Z$ is also positive definite, using Cholesky factorization: $Z^T\Phi Z = LL^T$, for a nonsingular lower triangular $L$ and replacing in (13), $LL^Tw = Z^TC \Rightarrow w = (LL^T)^{-1}Z^TC$, so that

$$||\Lambda|| = ||Zw|| = ||ZL^{T^{-1}}L^{-1}Z^TC|| \leqslant ||L^{-1}||^2|C|.$$

For procure $\Gamma$, we have $\Phi\Lambda + V^T\Gamma = C$ and using the QR factorization, $\Phi\Lambda + QR\Gamma = C$, premultiplying this equation by $Q^T$, results, $R\Gamma = Q^T(C - \Phi\Lambda)$, and because $\Lambda = Zw$ concludes to

$$R\Gamma = Q^T(C - \Phi Zw). \tag{14}$$

In this section, we discuss a method of creating surrogate models. For this purpose $\Phi$ must be conditionally positive definite of order at least 2 (Table 1), and $V \in \Pi_2^n$ be linear. The RBFs interpolation is defined such that at all sample points are established.

The RBFs coefficients $\lambda_i$ and $\nu_i$ must be bounded in magnitude. Define $y^i$ to be $i$th point in $Y$, that is in the vicinity of the trust region. However, for $n \geqslant 1$ condition (11) is not sufficient for the existence and uniqueness of the interpolation, and to guarantee the good quality of the model. Geometric conditions on the set $Y$ are required to ensure the existence and uniqueness of the interpolation [10, 21].

The process can be summarized as follows:
The study chooses the $n+1$ offinely independent points and then generates the other interpolation points.

The cubic spline $\varphi(r) = r^3$ in dimension $n$ is unisolvent (as defined below) on points $Y = \{y^1, \dots, y^N\}$ if the matrix,

$$\left[\varphi(||y^i - y^j||)\right] \qquad 1 \leqslant i, j \leqslant N,$$

is invertible for any choice of $N$ distinct points $y^1, \dots, y^N \in Y$.

**Definition 5.1.** *$Y$ is unisolvent for $\Pi^n_M$ if there exists a unique polynomial in $\Pi^n_M$ of lowest possible degree with interpolation points of $Y$.*

Unisolvent systems of RBFs are widely used in interpolation because they guarantee a unique solution to the interpolation problem. This is equivalent to the interpolation system (9) which is non-singular if the interpolation points set $Y$ is unisolvent.

The collection of $n+1$ distinct points will uniquely determine a polynomial of lowest possible degree in $\Pi^n$. In this section, we describe an algorithm to find $n+1$ of interpolation points which are offinelly independent points. We denote $D := \{d_i \in \mathbb{R}^n \mid f(x_k + d_i)\ is\ known\}$ and $d_i \in \Delta_k$. Algorithm 5.1 shows how to obtain $n+1$ offinelly interpolation points.

**Algorithm 5.2.** For finding n+1 offinely independent points:

**Step 0.** Input $D$, constants $0 < \gamma_0 \leqslant \gamma_1$, $\Delta_k \in (0, \Delta_{max}]$.

**Step 1.** Choose $D = \{d_1, d_2, \dots, d_{|D|}\} \in \mathbb{R}^n$ such that $x_i = x_k + d_i$ are close to $x_k$.

**Step 2.** Let $Z = I_n$.

**Step 3.** While $i, j \geqslant 1$
if $||d_i|| \leqslant \gamma_1.\Delta_k$, define $u = \frac{d_i}{\gamma_0 \Delta_k}$,

if $||proj^u_Z|| \geqslant \gamma_0$, then $y_j = d_i$,
Using the Gram-Schmidt, we obtain orthogonal basis for $Y$ as $\bar{Z}$, update $Z = \bar{Z}$

**Step 4.** If $|Y| < n+1$

if $||d_i|| \leqslant 2\Delta_{max}$, define $u = \frac{d_i}{\gamma_0 \Delta_k}$

if $||proj_Z^u|| \geqslant \gamma_0$, then $y_j = d_i$

Using the Gram-Schmidt process, we obtain an orthonormal basis for $Y$ as $\bar{Z}$. Update $Z = \bar{Z}$

However, with $n+1$ points, the solution of the system (9) is just interpolation obtained for linear function and coefficient $\Lambda = \emptyset$. To build the surrogate model for nonlinear functions, we must add some new points. Algorithm 5.2 shows how we can obtain "well independent" additional sample points in the trust region.

**Algorithm 5.3.** Finding additional independent points:

**Step 0.** Input $Y$ (obtained from algorithm 5.1), $p_{max} = \frac{(n+1)(n+2)}{2}$, $D = \{d_1, d_2, \ldots, d_{|D|}\}$, $\theta > 1$.

While $i \geqslant 1$

**Step 1.** If $|Y| < p_{max}$,

$$\Pi^T = \begin{bmatrix} y^1 = 0 & y^2 & \ldots & y^{|Y|} & d_i \\ 1 & 1 & \ldots & 1 & 1 \end{bmatrix}.$$

**Step 2.** Find the orthogonal basis $Z$ for null space $\Pi$.

**Step 3.** Build the interpolation matrix by using the cubic spline function at sample points $Y$,

$$\Phi_{new} = \begin{bmatrix} \Phi & \Phi_{d_j} \\ \Phi_{d_j}^T & 0 \end{bmatrix}.$$

**Step 4.** Obtain P=$Z^T \Phi_{new} Z$:

$$P = Z^T \Phi_{new} Z = \begin{bmatrix} Z^T \Phi Z & Z^T \Phi_{d_j} Z \\ Z^T \Phi_{d_j}^T Z & 0 \end{bmatrix}.$$

**Step 5.** $P$ is be positive definite for cubic spline function $\varphi(r) = r^3$, note that for $P$ to be is positive definite, the points $Y$ must be distinct.

**Step 6.** Let $P = LL^T$, if all diagonal entries of $L$ are positive, then add $d_j$ to the set of sample points $Y$.

This procedure continues until $|Y| = \frac{(n+1)(n+2)}{2}$. Note that the points $D = \{d_1, \ldots, d_{|D|}\}$ are smartly chosen around the trial point $x_k$ by a

random process.
In a derivative-free algorithm, it is essential to guarantee that whenever necessary a model of the objective function with uniformly good local accuracy can be constructed. The function $f$ is no longer guarantee that the model $Sm(x_k)$ approximates the function locally. Therefore, it is required that the derivative free method similar to what is observed by derivative based models.
The main difference between interpolation models and gradient-based models is that the former are considered as a suitable approximation of the objective function only under some conditions. These conditions depend mainly on the geometry of the points. If they are satisfied, we say that the model is valid in the trust region. If not, new points are generated to improve the accuracy of the model. The class of algorithms based on interpolation models are called conditional trust region method. The term conditional just means that the model is a convenient approximation of $f$ only if some conditions are satisfied. The general framework of trust region methods guarantees the convergence to a first or second-order critical point depending on the assumptions on the model and on the objective function. A full analysis of trust region methods can be found in [9, 20, 21].

Therefore, we have relatively simple analytic expressions for the gradient:

$$\nabla Sm_k(x_k + s) = \sum_{i=1}^{N} \lambda_i \varphi'(||s - y^i||) \frac{s - y^i}{||s - y^i||} + \nabla V(s),$$

and similar hessian $(\nabla^2 Sm_k)$.

## 6. Optimization Surrogate on Radial Basis Functions (OSRB)

This section discusses the details of the derivative-free algorithm for finding a global solution of problem (1). As pointed in the introduction, since the objective function is not necessarily smooth, the traditional methods are not sufficient to search good directions. We give the algorithm in which a surrogate model of problem (1) is solved. The algorithm

proceeds until the magnitude of the objective function become less than a natural stopping criterion.
Here, we propose a derivative–free algorithm. In this algorithm we solve the subproblem (3) which is approximated by using of RBFs (as in Section 5) and obtain search direction $s_k$.
Given the current iterate $x_k$ at step $k$, then we probe the behavior of the objective function $f(x)$ along direction $s_k$. In case sufficient reduction of the function value is obtained, a suitable optimal is computed and is used for the next iteration, i.e. $x_{k+1} = \tilde{x_k} + s_k$. If we do not obtain a sufficient reduction, then the trust region radius $\Delta_k$ is updated and interpolation points are chosen again (Algorithms 5.1 and 5.2). By solving the subproblem (3), we obtain another direction $s_k$ at the next iteration which suitably reduces the objective function.
Given $N$, a set of distinct interpolation points $Y = \{y^1 = 0, y^2, \ldots, y^N\} \in \mathbb{R}^n$ and the function values $\{f(x_k + y^i)\}$, we obtain the surrogate model for $f$ on $Y$. The Algorithm 6.1 is described as follows.

**Algorithm 6.1.** Iteration $k$ of a derivative-free surrogate model:

**Step 0.** Input $\epsilon > 0$, $0 \leqslant \gamma_0 < \gamma_1 \leqslant 1$, $0 < \eta < 1$ and $0 < \Delta_1 \leqslant \Delta_{max}$.
We assume that trial point $x_k$ is given.
While $k \geqslant 1$

**Step 1.** From Algorithms 5.1 and 5.2 find independent points that is denoted by $Y$.

**Step 2.** Obtain surrogate model $Sm(x_k + s)$ by using the RBF's described in Section 5.

**Step 3.** While $||\nabla Sm(x_k)|| > \epsilon$
If $Sm(x_k) - Sm(x_k + s_k) \leqslant \frac{\eta}{2} ||\nabla Sm(x_k)|| \min(\Delta_k, \frac{||\nabla Sm(x_k)||}{||\nabla^2 Sm(x_k)||})$
Obtain a step $s_k$ by solving: min $\{Sm(x_k + s)\ ;\ x_k + s \in B(x_k, \Delta_k)\}$.
Evaluate $f(x_k + s_k)$ and update the trial point according to the ratio $\rho_k$ (4),

$$x_{k+1} = \begin{cases} x_k + s_k & f(x_k + s_k) \leqslant f(x_k), \\ x_k & o.w, \end{cases}$$

$$\Delta_{k+1} = \begin{cases} min\{\gamma_1 \Delta_k, \Delta_{max}\} & f(x_k + s_k) \leqslant f(x_k), \\ \gamma_0 \Delta_k & f(x_k + s_k) > f(x_k). \end{cases}$$

If there is no adaptable direction $s_k$ to minimize $Sm(x_k)$, the trust region will be enlarged, i.e. $\Delta_{k+1} = \theta.\Delta_k$, incremment $k$ by 1 and go to Step 1.

In step 1, the interpolation points set $Y = \{y^1 = 0, \ldots, y^{p_{max}}\}$ are determined which are linearly independent. In step 2, we consider how to construct a model and to obtain parameters of RBFs model from (13), and (14). In step 3, the algorithm uses criteria for model $Sm_k(x_k + s)$, the aim is that $\nabla Sm(x_k)$ is not too different from the gradient of the objective function and updates parameters trust region method. We finds the candidate step $s_k$ by approximately solving the subproblem (4). In this paper, we solve subproblem (4) by using the Fmincon function in Matlab software.

## 6.1 Convergence properties of OSRB algorithm

In this study, the trust region algorithm ensures that $f(x)$ is sampled only within the relaxed level set,
$L(x_0) := \{y \in \mathbb{R}^n |\ ||x - y|| \leqslant \Delta_{\max},\ \forall x; f(x) \leqslant f(x_0)\}$.

**Theorem 6.1.1.** *Let $\{\Delta_k\}$ and $\{x_k\}$ be sequences guaranteed by OSRB Algorithm. Then,* $\lim_{k\to\infty} \Delta_k = 0$ *and* $\lim_{k\to\infty} \nabla f(x_k) = 0$.

**Proof**. After the last successful iteration, there is an infinite number of iterations that are not either acceptable or successful, therefore the trust region is reduced. If $x_{k+1} = x_k + s_k$ is obtained so that $f(x_{k+1}) \leqslant f(x_k)$, then $\Delta_k$ is never increased for sufficiently large $k$, so $\Delta_k$ is decreased at least once every $n$ iterations by a factor of $0 < \gamma < 1$, thus $\Delta_k$ convergence to zero. Secondly, for each $k$, after the $j$th iteration we have $|x_k - x_j|| \leqslant \lim_{k\to\infty} n\Delta_k \to 0$, now,

$||\nabla f(x_k)|| \leqslant ||\nabla f(x_k) - \nabla Sm(x_k) + \nabla Sm(x_k)|| \leqslant ||\nabla f(x_k) - \nabla Sm(x_k)|| + ||\nabla Sm(x_k)||$.

All terms of right hand side are equal to zero. $\square$

The statement of Theorem 6.1.1 gives a natural stopping criterion for OSRB algorithm. It results from the updating of the trust region at the $k$-th iteration.

Surrogate model $Sm_k$ is made such that,
$Sm(s_k) - Sm(0) = \mathcal{G}(0)^T s_k + \frac{1}{2} {s_k}^T \mathcal{H}_k(0) s_k$,
where $\mathcal{G}_k = \nabla f(x_k)$ and $\mathcal{H}_k = \nabla^2 f(x_k)$.

**Assumption 6.1.2.** The subproblem $C(x)$ is bounded below on $L(x_0)$ and $Sm(x)$ is twice continuously differentiable.

**Lemma 6.1.3.** *Suppose that assumption 6.1.2 holds. Then,*

$$Sm(s_k) - Sm(0) \geqslant \frac{1}{2} ||\mathcal{G}_k|| \min\{\Delta_k, \frac{||\mathcal{G}_k||}{\mathcal{H}_k}\}.$$

**Proof**. If $s_k = || - \frac{\mathcal{G}_k}{\mathcal{H}_k}|| \leqslant \Delta_k$, then the quadratic subproblem (3) can be resolved,

$$Sm(s_k) = Sm(0) - \frac{\mathcal{G}_k}{\mathcal{H}_k} \mathcal{G}_k + \frac{1}{2}(-\frac{\mathcal{G}_k}{\mathcal{H}_k})^T \mathcal{H}_k (-\frac{\mathcal{G}_k}{\mathcal{H}_k}),$$

knowing the cubic spline is twice continuously differentiable, $\mathcal{G}_k^T \mathcal{H}_k \mathcal{G}_k$ is positive definite. We know the model is convex along direction $s_k$. Next,

$$Sm(0) - Sm(s_k) = \frac{||\mathcal{G}_k||^2}{\mathcal{H}_k} - \frac{1}{2}\frac{||\mathcal{G}_k||^2}{\mathcal{H}_k} \geqslant \frac{1}{2} ||\mathcal{G}_k|| \min\{\Delta_k, \frac{||\mathcal{G}_k||}{\mathcal{H}_k}\}.$$

Lemma 6.1.3 guarantees that the OSRB Algorithm will sufficient decrease at iteration $k$. □

## 7. Numerical Results

In this section, we present a set of unconstrained problems from [14] which are solved by the OSRB algorithm to accommodate practical experiment to show the efficiency of the proposed method. Notice that the interpolation points are chosen so that interpolation matrix (9) always is invertible even the trust region is very small.

We have employed the Fmincon routine from Matlab which is corresponding to surrogate model. The starting points are randomly chosen in our algorithm. We solve unconstrained problems (1) which are not necessarily smooth to show the efficiency of the proposed method.